TITLE

Clinical prediction rules for prognosis and treatment prescription in neck pain: a systematic review

ABSTRACT
Clinical prediction rules (CPRs) developed to identify sub-groups of people with neck pain for different prognoses (i.e. prognostic) or response to treatments (i.e. prescriptive) have been recommended as a research priority to improve health outcomes for these conditions. A systematic review was undertaken to identify prognostic and prescriptive CPRs relevant to the conservative management of adults with neck pain and to appraise stage of development, quality and readiness for clinical application. Six databases were systematically searched from inception until 4th July 2016. Two independent reviewers assessed eligibility, risk of bias (PEDro and QUIPS), methodological quality and stage of development. 9,840 records were retrieved and screened for eligibility. Thirty-two studies reporting on 26 CPRs were included in this review. Methodological quality of included studies varied considerably. Most prognostic CPR development studies employed appropriate designs. However, many prescriptive CPR studies (n = 12/13) used single group designs and/or analysed controlled trials using methods that were inadequate for identifying treatment effect moderators. Most prognostic (n = 11/15) and all prescriptive (n = 11) CPRs have not progressed beyond the derivation stage of development. Four prognostic CPRs relating to acute whiplash (n = 3) or non-traumatic neck pain (n = 1) have undergone preliminary validation. No CPRs have undergone impact analysis. Most prognostic and prescriptive CPRs for neck pain are at the initial stage of development and therefore routine clinical use is not yet supported. Further validation and impact analyses of all CPRs are required before confident conclusions can be made regarding clinical utility.

Keywords
Clinical prediction rule, neck pain, prognosis, treatment

INTRODUCTION

Neck pain is the fourth leading cause of global disability and has an annual prevalence rate exceeding 30%[1,2]. Prolonged disability is common and poses considerable physical, psychological and economic consequences to individuals and society[3,4]. Health professionals face uncertainty in decision-making when managing neck pain due to conflicting reports of treatment effectiveness and difficulty in predicting prognosis[3]. As such, the identification of sub-groups within neck pain populations has been recognised as a research priority for improving management strategies[3,5].

Clinical prediction rules (CPRs) can be used to guide clinical decision-making in the assessment and treatment of individuals by enabling categorisation of those who have meaningful differences in symptomology[6,7]. CPRs are mathematically derived tools that quantify the contribution of various patient characteristics to create a set of variables that can be used to make predictions about an individual's diagnosis, prognosis or response to a specific intervention[7]. Diagnostic CPRs aim to enhance the detection of a specific condition and are developed using cross-sectional study designs to compare CPR findings to an established 'gold standard' test[7]. Prognostic CPRs enable estimation of the probability that a state of health such as change in pain or disability will occur in the future[8], and are ideal for educating patients regarding anticipated outcome as well as prioritising individuals for intervention[6,8]. Longitudinal study designs, such as prospective cohorts, are optimal because CPR findings are compared to changes in patient status over time[7]. Prescriptive CPRs guide decision-making by estimating the likelihood of successful response to a specific intervention[9]. Study designs that include a control group, such as randomised controlled trials (RCTs) are critical to the development of this type of CPR to ensure that treatment effect modifiers are discriminated from non-specific prognostic predictors[9,10].

The development of all types of CPRs broadly involves three stages[6]. First, CPRs are derived using statistical analyses to determine a set of variables with the greatest predictive power[6]. Derived CPRs are not recommended for clinical use because they may reflect chance statistical associations or be specific to the study population[6]. Second, CPRs are validated by prospective application in a new patient cohort[6]. Narrow validation involves testing the tool in a setting and population that is similar to the derivation study, whereas broad validation comprises application to a wider spectrum of patients and clinicians[6]. A successfully validated CPR may be used by clinicians with some level of confidence in its predictive accuracy[6,11]. The final stage of development, impact analysis, involves testing to see if CPR application results in changed clinician behavior and improved patient outcomes[6]. Only after impact analysis can a clinician be fully confident that CPR use may improve outcomes[6].

Recently, there has been an increase in the number of CPRs that target musculoskeletal conditions[12]. Whilst systematic reviews have examined CPRs for low back[13-15] and musculoskeletal pain[16-18], to our knowledge, CPR studies pertaining specifically to neck pain have not been reviewed systematically. This is important as there is evidence that prognostic indictors, treatment responses and recovery pathways differ between neck and other musculoskeletal pain conditions[19-23]. Clinician knowledge of CPRs appears variable and adoption in practice is often poor[24]. It has been suggested that low uptake may result from difficulty interpreting CPRs and appraising their quality[6,25]. The aim of this systematic review was to identify prognostic and prescriptive CPRs relevant to the conservative management of adults with neck pain and appraise stage of development, quality and readiness for application in clinical practice.

METHOD

Registration

The protocol for this review was registered with the International Prospective Register of Systematic Reviews (PROSPERO) and is available at http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015023362 (record number CRD42015023362).

Data sources and searches

A systematic literature search was conducted in PubMed, Embase, CINAHL Plus, AMED, PEDro and Cochrane Library databases from inception until 4th July 2016. A validated search strategy with high sensitivity in identifying CPRs[26] was used in combination with neck pain-specific strings suggested by the Cochrane Back Review Group[27] and, in consultation with a medical librarian, adapted for neck pain (see **Appendix A**). Supplementary strategies comprised citation tracking in Scopus and reference list screening of included studies.

Study selection

A CPR was operationally defined as a mathematically derived clinical tool designed to calculate the contribution of patient characteristics to create a set of variables with specific cut points that could be applied to make predictions about an individual's prognosis or response to a specific intervention[6]. Study eligibility criteria are outlined in **Table 1**. Briefly, studies were included that reported on prognostic and prescriptive CPRs relating to the conservative management of adults with non-specific, idiopathic, mechanical, traumatic, postural, cervical radiculopathy or whiplash associated neck pain. Identified records were downloaded to an electronic reference management system and duplicates removed. Two independent reviewers screened titles and abstracts of all records. The full texts of potentially eligible studies were screened by both reviewers against eligibility criteria to determine ultimate inclusion in the review. Disagreement on study eligibility was resolved by consensus or when unable, by consultation with a third reviewer.

Classification of CPR type

Inconsistencies exist in the classification of prognostic and prescriptive CPRs within review studies. Some reviews have classified CPRs based on the stated aims of each study[17,28], whilst others have distinguished CPR type by study design[15,16]. The latter recognises that using single group cohorts to develop prescriptive CPRs results in models of only prognostic value[10]. However, disregarding the original purpose of the CPR derivation will likely result in a lack of distinction between prognostic CPRs that were developed in response to a specific clinical need and those that were not. It also creates inconsistencies in nomenclature from original research reports. Therefore, in order to avoid confusion, we classified CPR type based on the original aims of each study. The appropriateness of research design for the purpose of each CPR's development was evaluated.

Data extraction and quality assessment

One reviewer extracted data including: CPR type, function, study design, patient population, potential predictor variables, outcome criteria, number of events per outcome, method of analysis and final CPR performance from each study. A second reviewer independently checked these data. The internal validity of included studies was appraised using two standardised tools. The Physiotherapy Evidence Database (PEDro) Scale[29] was used for studies with an RCT design. The PEDro scale is an 11-item scale that is valid and reliable in rating the methodological quality of RCTs[29,30]. All other studies, including those that pooled data from multiple RCT treatment groups, were evaluated using the Quality in Prognosis Studies (QUIPS) tool[31]. QUIPS is a six-item tool designed for use in observational prognostic studies[31]. Items that were unclear or not mentioned within the study were considered to be unmet. Two reviewers independently completed each appraisal tool for all included studies. Disagreement was resolved by consensus or consultation with a third reviewer. High risk of bias ratings informed, but did not exclude studies from inclusion in the synthesis of results.

Data synthesis and analysis

A qualitative synthesis of studies was performed by appraising the methodological quality, stage of development and readiness for clinical use of CPRs within each study. Methodological quality was evaluated by two independent reviewers using criteria employed in a recent review of CPRs applicable to low back pain[15]. These criteria are comprehensive, specific to CPR stage of development, and were selected because a validated tool does not exist for this purpose. CPR stage of development and readiness for clinical use were classified using frameworks from McGinn and colleagues[6]. Meta-analysis was not attempted due to heterogeneity in population, CPR function and outcome variables of included studies. Inter-rater agreement for each stage of the study selection process, risk of bias assessment and quality appraisal was calculated using absolute and chance-corrected degrees of

agreement (Kappa statistic) with 95% confidence intervals (CI) using predetermined strength of agreement labels[32]. Statistical analyses were conducted using IBM SPSS Statistics (version 22).

Reporting

This manuscript was prepared in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement[33].

RESULTS

Search results and study selection

Electronic searches yielded 12,714 records. After the removal of duplicates, 9,840 titles and abstracts were screened and 188 full text articles were assessed against eligibility criteria (**Figure 1**). Thirty-two studies were deemed appropriate for inclusion. The third reviewer was consulted on four occasions to clarify the eligibility of studies attempting validation of non-mathematically derived models. The most common reason for exclusion was not meeting this review's operational definition of a CPR. Absolute inter-rater agreement was 98.1% for title and abstract screening and 90.3% for full text eligibility. Chance-corrected agreement for screening of full text was substantial ($\kappa = 0.78$, 95% CI 0.67 to 0.89).

Study characteristics

Prognostic CPRs: Nineteen studies reported on the development of 15 prognostic CPRs (**Supplementary Table 1**). The majority of these studies were of a prospective cohort design (n = 12). Thirteen studies concerned CPRs for use in people with acute whiplash; five related to non-traumatic neck pain; and one targeted cervical radiculopathy. Outcome measures included neck disability (n = 9), work disability (n = 5), pain (n = 4) and perceived recovery (n = 2). Most studies concerned CPR derivation only (n = 12). Five studies performed CPR validation and two comprised both derivation and validation within the same report. No prognostic CPR impact analysis studies were identified.

Prescriptive CPRs: Thirteen studies reported on the development of 11 prescriptive CPRs (**Supplementary Table 2**). The majority of these studies were of single group design (n = 10). Only one derivation study appropriately analysed an RCT for the development of this type of CPR. Interventions included spinal manipulation (n = 5), cervical traction (n = 3), exercise program (n = 2), and a combination of multiple techniques (n = 3). Treatment success was determined using self-perceived improvement (n = 7), neck disability (n = 2) and a combination of outcome measures (n = 4). Most studies reported the derivation stage of development only (n = 9). One study combined derivation and validation within the same report, two attempted validation, and one was not classified as it concerned the validation of a CPR that was not mathematically derived[34]. No validation studies supported the predictive ability of the derived models[35-37]. No reports of prescriptive CPR impact analysis were identified.

Risk of bias assessment

Prognostic CPRs: All nineteen studies were evaluated using the QUIPS tool (**Table 2**). Whilst many studies sufficiently reported study attrition (n = 10) and outcome measurement (n =9), a moderate or high risk of bias was evident in most studies for statistical analysis and reporting (n = 16) and study participation (n = 14) criteria.

Prescriptive CPRs: The QUIPS tool was used to appraise risk of bias in the prescriptive derivation studies due to use of single group (n = 10) and pooled RCT (n = 1) study designs (**Table 2**). Many studies adequately reported outcome measurement (n =7) and study attrition (n = 6). However, a moderate or high risk of bias was evident in the majority of studies for statistical analysis and reporting (n = 10) and study participation (n = 7) criteria. Potential risk of bias in two prescriptive validation CPR studies that used an RCT design (n = 2) were evaluated against the PEDro scale (**Table 3**), with scores of 6 and 7 out of 10 respectively, indicating moderate quality[38]. The most frequent source of bias was lack of patient and therapist blinding and inadequate baseline comparability of treatment groups.

Overall inter-rater agreement for PEDro and QUIPS were substantial with κ = 0.68 (95% CI 0.36 to 1.0, absolute agreement 86.4%) and κ = 0.79 (95% CI 0.71 to 0.86, absolute agreement 84.4%) respectively. All disagreements concerning risk of bias were resolved by consensus.

Qualitative appraisal

Results of methodological quality specific to derivation and validation are outlined in **Table 4** and **Table 5** respectively. Studies that included both stages within one report were evaluated against criteria for both.

Prognostic CPRs: All 14 prognostic CPR derivation studies blinded the assessment of predictor variables, and most justified variable selection (n = 13). Many studies, however, did not justify participant numbers (n = 12), use predictor variables with demonstrated reliability (n = 11), test for co-linearity in predictor variables (n = 10), or include at least 10 outcome events per predictor variable (n = 10). For validation studies (n = 7), accurate application of the rule was present in the majority of cases (n = 5). Failure to report missing data (n = 6) and application in a different clinical setting (n = 5) were frequently not met.

Prescriptive CPRs: All prescriptive CPR derivation studies (n = 10) blinded the assessment of predictor variables and outcome measures, and described the mathematical techniques used. Most of these studies, however, did not justify participant numbers (n = 9), test for co-linearity in predictor variables (n = 9), or include at least 10 outcome events per predictor variable (n = 9). Of four validation studies, most performed a prospective validation in a new patient population (n = 3), used a representative sample, and accurately applied and described the rule (n = 3). Description of uncertainty in post-test probability (n = 4), and description of uncertainty in CPR accuracy (n = 3) were most frequently not met.

Interrater agreement for derivation criteria was almost perfect (κ = 0.83 (95% CI 0.78 to 0.88, absolute agreement 91.2%) and substantial for validation criteria (κ = 0.75 (95% CI 0.63 to 0.87, absolute agreement 87.3%).

DISCUSSION

Thirty two studies were identified that reported on the development of 15 prognostic and 11 prescriptive CPRs relating to the conservative management of adults with neck pain. Whilst the majority of prognostic CPRs remain at the derivation stage of development, preliminary investigations of validity have been successfully performed on four models. Ten out of 11 prescriptive CPRs were derived using study designs that were inappropriate for the purpose of developing a prescriptive tool. The study design and/or analyses used in three subsequent validation studies were not appropriate to replicate the results of these models.No CPRs of either type evaluated in this review have undergone impact analyses[6]. Future CPR studies should consider identified methodological shortcomings including inappropriate study design, insufficient sample size, and incomplete reporting of statistical analyses and model performance.

Prognostic CPR readiness for use

Most prognostic CPRs were at derivation stage[39-48] and therefore are not yet at a stage of development supporting routine clinical use[6]. At this stage, clinicians may consider using individual predictor variables contained within these models[6]. For example, higher neck disability (e.g. Neck Disability Index) was identified by numerous CPRs as a predictor of non-recovery for pain and disability outcomes[39,41,43,49,50]. Assessment of neck disability may be useful in informing judgements on prognosis where the population is similar to that used in the CPR's derivation[6]. The prominence of psychological and social predictor variables within prognostic CPRs confirms the likely importance of these factors for prognosis. The assessment of psychosocial factors has been recognised as necessary for improving outcomes in people with spinal pain and as such, would seem important to examine in the clinical setting[51]. However, given the considerable breadth of biopsychosocial variables proposed

within the reviewed CPRs, consideration of specific predictor variables should be used to compliment and not replace usual clinical decision-making strategies[6].

A number of studies progressed the development of prognostic CPRs to either narrow or broad validation[19,49,52-56]. Application of narrowly validated CPRs requires caution because evidence for model generalisability is not strong[6]. Three CPRs relating to acute whiplash prognosis were identified as having undergone narrow validation. Hence, these may be considered for use with populations and clinical settings similar to those of the development studies[6]. First, Ritchie and colleagues (2015) validated a CPR comprising initial disability levels, hyper-arousal symptoms and patient age to predict recovery or ongoing disability at six to 12 months post whiplash injury in Queensland, Australia[49]. This model is promising in that both the derivation and validation studies satisfied the majority of methodological quality criteria used in this review, and the accuracy and post-test probability remained strong on validation[49]. A preliminary attempt at further validation of this was tested in a different geographical (USA) setting to that of the initial development studies, suggesting that the CPR is being considered more widely within the research community, and the results were supportive of the models validity[57]. Second, Kasch and colleagues (2011 and 2013) developed a CPR comprising cervical range of motion, number of non-painful complaints and baseline pain intensity to predict 12 month work disability that displayed very good discriminative ability (area under the curve = 0.90, 95% CI 0.74 to 1.0) and a positive likelihood ratio of 7.8 for the highest of seven risk strata, which is likely to cause moderate shifts in pre- to post-test probability[52,56,58]. The CPR's included variables, predictor variable measurement and scoring algorithm were altered between development studies for reasons that are not clear[53,56,59]. Whilst updating a CPR is not unusual, modifications are usually made in response to disappointing validation accuracy, which in this instance does not appear to have been determined[60]. These changes could impact methodological quality and risk of bias. Finally, a whiplash prognostic equation including impaired neck movement, headache, head trauma, age, neck pain intensity, headache intensity, nervousness, neuroticism, and focused attention variables developed by Radanov and Sturzenegger (1996) also demonstrated good post-test probability (88%) on validation[55]. Interpretation of the results is difficult due to incomplete reporting of CPR accuracy, omission of outcome measure definition and use of a small validation sample size (n = 16). Broad validation is optimal because it enables confirmation of CPR accuracy across a diverse spectrum of patients and settings[6]. Schellingerhout and colleagues (2010)[54] reported on the only CPR identified by this review to have maintained accuracy on broad validation. The rule's predictive utility was modest with a positive predictive value of 51% (95% CI 43 to 59) and a positive likelihood ratio of 1.6, which is unlikely to cause a significant shift in pre- to post-test probability[58]. However, given reasonable methodological quality, maintenance of predictive ability and prospective validation in a new and geographically different population, this CPR may be considered for use in predicting prognosis based on global perceived recovery for individuals with non-traumatic neck pain[6].
Impact analysis is ultimately required to ensure that application of statistically accurate predictive models results in clinically beneficial consequences[6].

Prescriptive CPR readiness for use

The derived stage and lack of sufficient validity of prescriptive CPRs identified in this review[34,61-66] means that clinical use of any model is not yet advised[6]. Unlike predictor variables identified in the reviewed prognostic CPRs, attention to components included in prescriptive CPRs is not recommended. The use of inappropriate (single group) designs in all but one identified prescriptive CPR derivation studies means that predictors of response to treatment may not have been differentiated from predictors of outcome regardless of treatment[9,10]. Researchers employing these methods risk creating a CPR that is merely prognostic given a specific treatment[7,10]. Additionally, it has been advocated that subsequent validation of such CPRs should not be considered as adequate replication (even if an appropriate RCT design has been used), because the initial model was inappropriately derived[10]. Failure to identify treatment effect modifiers during derivation also heightens the possibility of reduced CPR accuracy when validation is attempted using an appropriate study design[10]. Two identified prescriptive CPR validation studies employed RCTs to validate prescriptive models that were derived using single group designs and found the results did not support the prescriptive validity of either rule[35,36].

General methodological considerations for CPR development

Several potential sources of bias relevant to both CPR types were identified and should be addressed in future studies. Only three derivation studies performed a priori sample size calculations[41,50,64]. These calculations are necessary to avoid the risk of model over fitting or optimistic predictive performance, as well as to reduce the likelihood of disappointing validation accuracy[60,67]. It has been suggested that 10 to 15 events per outcome be a guide for adequate sample size[6]. Second, many studies reported incomplete statistical analyses. CPR accuracy, post-test probabilities and associated uncertainty intervals would enable improved interpretation of a CPR's predictive ability[68,69]. Descriptions of missing values and the use of imputation methods, where applicable should also be included[67].

Strengths and limitations of the review

To our knowledge this is the first review to systematically examine CPRs relevant to the conservative management of neck pain. As such, it provides a comprehensive synthesis of CPRs that may be of benefit to clinicians who treat this population. CPRs were operationally defined in this review for the purpose of transparency, reproducibility and to ensure that the included tools could be reasonably applied by clinicians to individual patients. Our definition was liberal in comparison to other reviews in that it did not require explicit use of the term 'clinical prediction rule'. Additionally, our inclusion criteria did not require specific statistical analyses. Consequently, this review was inclusive and may aid clinicians in interpreting and comparing the multitude of CPRs proposed in the literature. This review is limited in that criteria used to assess the methodological quality of included studies have not yet been validated. However, these criteria were systematically developed using key factors from the literature on methodological standards for CPR development[6,7,67-71], contain components commonly used in other reviews, and were used in combination with validated PEDro and QUIPS scales to improve assessment rigor. The classification of CPR type using study aim, and not design will have led to over-reporting of CPRs that have prescriptive value. However, classification by original study aim ensures consistent nomenclature with original reports, and discriminates CPRs developed for prognostic purposes from those developed from inadequate prescriptive methods

Conclusions and recommendations for future research

The substantial number of CPRs identified in this review confirms that progress is being made towards the identification of sub-groups of patients with neck pain. Clinical use of most prognostic CPRs is not yet recommended because of their early stage of development and moderate methodological quality. However, clinicians may choose to consider four validated prognostic CPRs or assessment of individual predictor variables contained within these models, to inform judgements of outcome. Derivation stage of development was similarly prominent in prescriptive CPRs, however, uncertain prescriptive value resulting from a reliance on single group study designs is a major limitation to the use of these models. To provide stronger evidence to support the clinical use of all CPRs, future research should employ study designs that are appropriate for the type of CPR being developed, and ensure adequate broad validation and impact analyses.

ACKNOWLEDGEMENTS

Thank you to XXX from XXX and XXX from XXX for their review of search strategies.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

1. Fejer R, Kyvik KO, Hartvigsen J. The prevalence of neck pain in the world population: a systematic critical review of the literature. *European Spine Journal.* 2006;15(6):834-848.
2. US Burden of Disease Collaborators. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. *Journal of the American Medical Association.* 2013;310(6):591-606.
3. Cohen SP. Epidemiology, diagnosis, and treatment of neck pain. *Mayo Clinic Proceedings.* 2015;90(2):284-299.
4. Vasseljen O, Woodhouse A, Bjørngaard JH, Leivseth L. Natural course of acute neck and low back pain in the general population: the HUNT study. *Pain.* 2013;154(8):1237-1244.
5. Foster NE, Dziedzic KS, Windt DA, Fritz JM, Hay EM. Research priorities for non-pharmacological therapies for common musculoskeletal problems: nationally and internationally agreed recommendations. *BMC Musculoskeletal Disorders.* 2009;10(1):3.
6. McGinn TG, Guyatt GH, Wyer PC, et al. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. *Journal of the American Medical Association.* 2000;284(1):79-84.
7. Beattie P, Nelson R. Clinical prediction rules: what are they and what do they tell us? *Australian Journal of Physiotherapy.* 2006;52(3):157-163.
8. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *British Medical Journal.* 2009;338:b375.
9. Nee RJ, Coppieters MW. Interpreting research on clinical prediction rules for physiotherapy treatments. *Manual Therapy.* 2011;16(2):105-108.
10. Hancock M, Herbert RD, Maher CG. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. *Physical Therapy.* 2009;89(7):698-704.
11. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Annals of Internal Medicine.* 2006;144(3):201-209.
12. Keogh C, Wallace E, O'Brien KK, et al. Developing an international register of clinical prediction rules for use in primary care: a descriptive analysis. *The Annals of Family Medicine.* 2014;12(4):359-366.
13. Patel S, Friede T, Froud R, Evans DW, Underwood M. Systematic review of randomized controlled trials of clinical prediction rules for physical therapy in low back pain. *Spine.* 2013;38(9):762-769.
14. Haskins R, Osmotherly PG, Rivett DA. Diagnostic clinical prediction rules for specific subtypes of low back pain: a systematic review. *Journal of Orthopaedic & Sports Physical Therapy.* 2015;45(2):1-45.
15. Haskins R, Osmotherly PG, Rivett DA. Validation and impact analysis of prognostic clinical prediction rules for low back pain is needed: a systematic review. *Journal of Clinical Epidemiology.* 2015;68(7):821-832.
16. van Oort L, van den Berg T, Koes BW, et al. Preliminary state of development of prediction models for primary care physical therapy: a systematic review. *Journal of Clinical Epidemiology.* 2012;65(12):1257-1266.
17. Stanton TR, Hancock MJ, Maher CG, Koes BW. Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. *Physical Therapy.* 2010;90(6):843-854.
18. Beneciuk JM, Bishop MD, George SZ. Clinical prediction rules for physical therapy interventions: a systematic review. *Physical Therapy.* 2009;89(2):114-124.
19. Dagfinrud H, Storheim K, Magnussen L, et al. The predictive validity of the Örebro Musculoskeletal Pain Questionnaire and the clinicians' prognostic assessment following manual therapy treatment of patients with LBP and neck pain. *Manual Therapy.* 2013;18(2):124-129.
20. Wertli MM, Held U, Campello M, Weiner SS. Obesity is associated with more disability at presentation and after treatment in low back pain but not in neck pain: findings from the OIOC registry. *BMC Musculoskeletal Disorders.* 2016;17(1):1.

21. Uluğ N, Yakut Y, Alemdaroğlu İ, Yılmaz Ö. Comparison of pain, kinesiophobia and quality of life in patients with low back and neck pain. *Journal of Physical Therapy Science.* 2016;28(2):665.
22. Webb R, Brammah T, Lunt M, Urwin M, Allison T, Symmons D. Prevalence and predictors of intense, chronic, and disabling neck and back pain in the UK general population. *Spine.* 2003;28(11):1195-1202.
23. Sihawong R, Sitthipornvorakul E, Paksaichol A, Janwantanakul P. Predictors for chronic neck and low back pain in office workers: a 1-year prospective cohort study. *Journal of Occupational Health.* 2016;58(1):16-24.
24. Haskins R, Osmotherly PG, Southgate E, Rivett DA. Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain. *Manual Therapy.* 2014;19(2):142-151.
25. Adams ST, Leveson SH. Clinical prediction rules. *British Medical Journal.* 2012;344:d8312.
26. Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PloS ONE.* 2012;7(2):e32844.
27. Cochrane Back Review Group. Updated Search Strategies for CBG January 2013. http://back.cochrane.org/sites/back.cochrane.org/files/uploads/PDF/CBRG%20Search%20Strategies%20Jan%202013.pdf. Accessed 21 May 2015.
28. Haskins R, Rivett DA, Osmotherly PG. Clinical prediction rules in the physiotherapy management of low back pain: a systematic review. *Manual Therapy.* 2012;17(1):9-21.
29. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. *Physical Therapy.* 2003;83(8):713-721.
30. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association.* 1995;273(5):408-412.
31. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. *Annals of Internal Medicine.* 2013;158(4):280-286.
32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977:159-174.
33. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *International Journal of Surgery.* 2010;8(5):336-341.
34. Fritz JM, Brennan GP. Preliminary examination of a proposed treatment-based classification system for patients receiving physical therapy interventions for neck pain. *Physical Therapy.* 2007;87(5):513-524.
35. Fritz JM, Thackeray A, Brennan GP, Childs JD. Exercise only, exercise with mechanical traction, or exercise with over-door traction for patients with cervical radiculopathy, with or without consideration of status on a previously described subgrouping rule: a randomized clinical trial. *Journal of Orthopaedic & Sports Physical Therapy.* 2014;44(2):45-57.
36. Cleland JA, Mintken PE, Carpenter K, et al. Examination of a clinical prediction rule to identify patients with neck pain likely to benefit from thoracic spine thrust manipulation and a general cervical range of motion exercise: multi-center randomized clinical trial. *Physical Therapy.* 2010;90(9):1239-1250.
37. Keating JL, Kent P, Davidson M, Duke R, McKinnon L, De Nardis R. Predicting short-term response and non-response to neck strengthening exercise for chronic neck pain. *Journal of Whiplash & Related Disorders.* 2005;4(1):43-55.
38. PEDro Statistics. 2015; http://www.pedro.org.au/english/downloads/pedro-statistics/. Accessed 18 January, 2016.
39. Atherton K, Wiles NJ, Lecky FE, et al. Predictors of persistent neck pain after whiplash injury. *Emergency Medicine Journal.* 2006;23(3):195-201.
40. Hartling L, Pickett W, Brison RJ. Derivation of a clinical decision rule for whiplash associated disorders among individuals involved in rear-end collisions. *Accident Analysis & Prevention.* 2002;34(4):531-539.

41. Williamson E, Williams MA, Gates S, Lamb SE. Risk factors for chronic disability in a cohort of patients with acute whiplash associated disorders seeking physiotherapy treatment for persisting symptoms. *Physiotherapy.* 2015;101(1):34-43.
42. Gabel CP, Burkett B, Neller A, Yelland M. Can long-term impairment in general practitioner whiplash patients be predicted using screening and patient-reported outcomes? *International Journal of Rehabilitation Research.* 2008;31(1):79-80.
43. Nederhand MJ, Ijzerman MJ, Hermens HJ, Turk DC, Zilvold G. Predictive value of fear avoidance in developing chronic neck pain disability: consequences for clinical decision making. *Archives of Physical Medicine & Rehabilitation.* 2004;85(3):496-501.
44. Carroll LJ, Cassidy JD, Cote P. The role of pain coping strategies in prognosis after whiplash injury: passive coping predicts slowed recovery. *Pain.* 2006;124(1-2):18-26.
45. Grooten WJA, Mulder M, Josephson M, Alfredsson L, Wiktorin C. The influence of work-related exposures on the prognosis of neck/shoulder pain. *European Spine Journal.* 2007;16(12):2083-2091.
46. Landers MR, Creger RV, Baker CV, Stutelberg KS. The use of fear-avoidance beliefs and nonorganic signs in predicting prolonged disability in patients with neck pain. *Manual Therapy.* 2008;13(3):239-248.
47. Vos CJ, Verhagen AP, Koes BW. The ability of the acute low back pain screening questionnaire to predict sick leave in patients with acute neck pain. *Journal of Manipulative & Physiological Therapeutics.* 2009;32(3):178-183.
48. Cleland JA, Fritz JM, Whitman JM, Heath R. Predictors of short-term outcome in people with a clinical diagnosis of cervical radiculopathy. *Physical Therapy.* 2007;87(12):1619-1632.
49. Ritchie C, Hendrikz J, Jull G, Elliott J, Sterling M. External validation of a clinical prediction rule to predict full recovery and ongoing moderate/severe disability following acute whiplash injury. *Journal of Orthopaedic & Sports Physical Therapy.* 2015;45(4):242-250.
50. Ritchie C, Hendrikz J, Kenardy J, Sterling M. Derivation of a clinical prediction rule to identify both chronic moderate/severe disability and full recovery following whiplash injury. *Pain.* 2013;154(10):2198-2206.
51. Weiner BK. Spine update: the biopsychosocial model and spine care. *Spine.* 2008;33(2):219-223.
52. Kasch H, Kongsted A, Qerama E, Bach FW, Bendix T, Jensen TS. A new stratified risk assessment tool for whiplash injuries developed from a prospective observational study. *BMJ Open.* 2013;3(1):e002050.
53. Kasch H, Qerama E, Kongsted A, Bendix T, Jensen TS, Bach FW. Clinical assessment of prognostic factors for long-term pain and handicap after whiplash injury: a 1-year prospective study. *European Journal of Neurology.* 2008;15(11):1222-1230.
54. Schellingerhout JM, Heymans MW, Verhagen AP, Lewis M, de Vet HC, Koes BW. Prognosis of patients with nonspecific neck pain: development and external validation of a prediction rule for persistence of complaints. *Spine.* 2010;35(17):E827-E835.
55. Radanov BP, Sturzenegger M. Predicting recovery from common whiplash. *European Neurology.* 1996;36(1):48-51.
56. Kasch H, Qerama E, Kongsted A, Bach FW, Bendix T, Jensen TS. The risk assessment score in acute whiplash injury predicts outcome and reflects biopsychosocial factors. *Spine.* 2011;36:S263-S267.
57. Elliott JM, Courtney DM, Rademaker A, Pinto D, Sterling MM, Parrish TB. The Rapid and Progressive Degeneration of the Cervical Multifidus in Whiplash: A MRI study of Fatty Infiltration. *Spine.* 2015.
58. Jaeschke R, Guyatt GH, Sackett DL, et al. Users' Guides to the Medical Literature: III. How to use an article about a diagnostic test B. What are the results and will they help me in caring for my patients? *Journal of the American Medical Association.* 1994;271(9):703-707.
59. Kasch H, Bach FW, Jensen TS. Handicap after acute whiplash injury: a 1-year prospective study of risk factors. *Neurology.* 2001;56(12):1637-1643.
60. Toll D, Janssen K, Vergouwe Y, Moons K. Validation, updating and impact of clinical prediction rules: a review. *Journal of Clinical Epidemiology.* 2008;61(11):1085-1094.

61. Tseng YL, Wang WTJ, Chen WY, Hou TJ, Chen TC, Lieu FK. Predictors for the immediate responders to cervical manipulation in patients with neck pain. *Manual Therapy.* 2006;11(4):306-315.
62. Puentedura EJ, Cleland JA, Landers MR, Mintken P, Louw A, Fernandez-de-Las-Penas C. Development of a clinical prediction rule to identify patients with neck, pain likely to benefit from thrust joint manipulation to the cervical spine. *Journal of Orthopaedic & Sports Physical Therapy.* 2012;42(7):577-592.
63. Saavedra-Hernandez M, Castro-Sanchez AM, Fernandez-de-Las-Penas C, Cleland JA, Ortega-Santiago R, Arroyo-Morales M. Predictors for identifying patients with mechanical neck pain who are likely to achieve short-term success with manipulative interventions directed at the cervical and thoracic spine. *Journal of Manipulative & Physiological Therapeutics.* 2011;34(3):144-152.
64. Cai C, Ming G, Ng LY. Development of a clinical prediction rule to identify patients with neck pain who are likely to benefit from home-based mechanical cervical traction. *European Spine Journal.* 2011;20(6):912-922.
65. Hanney WJ, Kolber MJ, George SZ, Young I, Patel CK, Cleland JA. Development of a preliminary clinical prediction rule to identify patients with neck pain that may benefit from a standardized program of stretching and muscle performance exercise: a prospective cohort study. *International Journal of Sports Physical Therapy.* 2013;8(6):756-776.
66. Nee RJ, Vicenzino B, Jull GA, Cleland JA, Coppieters MW. Baseline characteristics of patients with nerve-related neck and arm pain predict the likely response to neural tissue management. *Journal of Orthopaedic & Sports Physical Therapy.* 2013;43(6):379-391.
67. Bouwmeester W, Zuithoff NPA, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Medicine.* 2012;9(5):e1001221.
68. Lubetzky-Vilnai A, Ciol M, McCoy SW. Statistical analysis of clinical prediction rules for rehabilitation interventions: current state of the literature. *Archives of Physical Medicine & Rehabilitation.* 2014;95(1):188-196.
69. Haskins R, Osmotherly PG, Tuyl F, Rivett DA. Uncertainty in clinical prediction rules: the value of credible intervals. *Journal of Orthopaedic & Sports Physical Therapy.* 2014;44(2):85-91.
70. Childs JD, Cleland JA. Development and application of clinical prediction rules to improve decision making in physical therapist practice. *Physical Therapy.* 2006;86(1):122-131.
71. Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. *Annals of Emergency Medicine.* 1999;33(4):437-447.
72. Cleland JA, Childs JD, Fritz JM, Whitman JM, Eberhart SL. Development of a clinical prediction rule for guiding treatment of a subgroup of patients with neck pain: use of thoracic spine manipulation, exercise, and patient education. *Physical Therapy.* 2007;87(1):9-23.
73. Raney NH, Petersen EJ, Smith TA, et al. Development of a clinical prediction rule to identify patients with neck pain likely to benefit from cervical traction and exercise. *European Spine Journal.* 2009;18(3):382-391.
74. Schellingerhout JM, Verhagen AP, Heymans MW, et al. Which subgroups of patients with non-specific neck pain are more likely to benefit from spinal manipulation therapy, physiotherapy, or usual care? *Pain.* 2008;139(3):670-680.
75. Hartling L, Brison RJ, Ardern C, Pickett W. Prognostic value of the Quebec classification of whiplash-associated disorders. *Spine.* 2001;26(1):36-41.
76. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain.* 1992;50(2):133-149.
77. Sobel JB, Sollenberger P, Robinson R, Polatin PB, Gatchel RJ. Cervical nonorganic signs: a new clinical tool to assess abnormal illness behavior in neck pain patients: a pilot study. *Archives of Physical Medicine and Rehabilitation.* 2000;81(2):170-175.

Table 1 Study eligibility criteria

Inclusion criteria
• Aim of derivation, validation and/or impact analysis of one or more CPRs relating to prognosis or conservative management of adults with non-specific, idiopathic, mechanical, traumatic, postural, cervical radiculopathy or whiplash associated neck pain • CPRs developed in the form of a score chart, algorithm or model containing at least two predictor variables drawn from clinical characteristics such as patient history, assessment findings or simple laboratory results • CPRs with required components that are able to be assessed or easily obtained and interpreted by clinicians • Publication in a peer reviewed journal
Exclusion criteria
• Diagnostic CPRs • CPRs concerning prognosis or management of malignancy, infection, fracture, systemic inflammatory disease, headache, spinal cord injury or spinal pain where there is no differentiation between cervical, thoracic and/or lumbar symptom locations • CPRs that target surgical or radiological management of neck pain • Studies that included children <17 years of age • Study protocols or literature published in the form of conference proceedings or abstracts, reviews, editorials, commentaries, letters, dissertations, books, book chapters and practice guidelines • Case study or case series design • Study not reported in English language

Table 2 Potential risk of bias of cohort and combined-RCT studies by CPR type as assessed using QUIPS

Study	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Study confounding	Statistical analysis/report
Prognostic CPRs						
Atherton et al (2006)[39]	Moderate	Moderate	Moderate	High	Not applicable	Moderate
Carroll et al (2006)[44]	Moderate	High	Moderate	Moderate	Moderate	Moderate
Cleland et al (2007)[48]	Low	Low	Low	Low	Not applicable	Moderate
Dagfinrud et al (2013)[19]	High	High	Moderate	Low	Not applicable	Moderate
Gabel et al (2008)[42]	High	High	High	Low	Not applicable	High
Grooten et al (2007)[45]	Low	Low	High	High	Low	Low
Hartling et al (2002)[40]	High	Moderate	High	High	Not applicable	High
Kasch et al (2001)[59]	Moderate	Low	Moderate	High	Not applicable	Moderate
Kasch et al (2008)[53]	Low	Low	Not applicable	High	Not applicable	Moderate
Kasch et al (2011)[56]	High	Low	High	High	Not applicable	High
Kasch et al (2013)[52]	Low	Low	Not applicable	High	Moderate	Moderate
Landers et al (2008)[46]	Moderate	Low	Moderate	Low	Not applicable	High
Nederhand et al (2004)[43]	Low	Low	Low	Low	Not applicable	High
Radanov & Sturzenegger (1996)[55]	Moderate	Low	Moderate	High	Not applicable	High
Ritchie et al (2013)[50]	Moderate	Moderate	Low	Low	Not applicable	Low
Ritchie et al (2015)[49]	Moderate	Low	Low	Low	Not applicable	Low
Schellingerhout et al (2010)[54]	Moderate	High	Low	Low	Not applicable	Moderate
Vos et al (2009)[47]	High	Moderate	High	High	Not applicable	Moderate
Williamson et al (2015)[41]	Moderate	Moderate	Moderate	Low	Moderate	Moderate
Prescriptive CPRs						
Cai et al (2011)[64]	Moderate	Low	Low	Moderate	Not applicable	Moderate
Cleland et al (2007)[72]	Low	Low	Low	Low	Not applicable	Moderate
Fritz & Brennan (2007)[34]	High	High	Not applicable	Low	High	High
Hanney et al (2013)[65]	Moderate	Moderate	Moderate	Moderate	Not applicable	Moderate
Keating et al (2005)[37]	Moderate	High	High	Moderate	Not applicable	High
Nee et al (2013)[66]	Moderate	Low	Moderate	Low	Not applicable	High
Puentedura et al (2012)[62]	Low	Low	Low	Low	Not applicable	Moderate
Raney et al (2009)[73]	Low	Moderate	Moderate	Low	Not applicable	High
Saavedra-Hernandez et al (2011)[63]	Low	Low	Moderate	Low	Not applicable	High
Schellingerhout et al (2008)[74]	Moderate	High	Low	Low	Low	High
Tseng et al (2006)[61]	Moderate	Low	Moderate	High	Not applicable	Low

Table 3 Potential risk of bias of RCT designed studies as assessed using PEDro scale

Study	Eligibility Criteria	Random allocation	Concealed allocation	Groups similar at baseline	Participant blinding	Therapist blinding	Assessor blinding	Adequate follow-up	Intention to treat analysis	Between-group difference reported	Point estimate and variability	Score (/10)
Prescriptive CPRs												
Fritz et al (2014)[35]	Y	Y	Y	N	N	N	Y	Y	Y	Y	Y	7
Cleland et al (2010)[36]	Y	Y	Y	N	N	N	Y	N	Y	Y	Y	6

Y = yes, N = no

Table 4 Methodological quality of CPR derivation studies by CPR type

Study	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Prognostic CPRs																			
Atherton et al (2006)[39]	Y	Y	N	N	Y	Y	Y	N	N	Y	Y	N	Y	N	N	N	N	N	Y
Carroll et al (2006)[44]	Y	Y	N	Y	N	Y	Y	Y	N	N	Y	Y	Y	Y	N	N	Y	Y	Y
Cleland et al (2007)[48]	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	N	Y	N	N	N	Y	N	Y
Gabel et al (2008)[42]	Y	N	N	N	N	N	Y	N	Y	Y	N	N	N	N	N	NA[b]	N	N	Y
Grooten et al (2007)[45]	Y	N	N	N	Y	Y	Y	N	N	Y	Y	N	Y	N	N	NA[c]	Y	Y	Y
Hartling et al (2002)[40]	Y	Y	N	Y	Y	Y	Y	N	N	Y	Y	Y	Y	N	NA	N	Y	Y	N
Kasch et al (2001)[59]	Y	Y	N	Y	N	Y	Y	N	N	Y	Y	N	N	N	N	N	N	N	Y
Landers et al (2008)[46]	Y	N	N	N	N	Y	Y	N	Y	Y	Y	N	N	N	N	Y	Y	N	Y
Nederhand et al (2004)[43]	Y	N	N	N	Y	Y	Y	N	Y	N	Y	N	Y	N	N	NA[b]	Y	N	Y
Radanov & Sturzenegger (1996)[55]	Y	N	N	N	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N
Ritchie et al (2013)[50]	N[a]	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	Y
Schellingerhout et al (2010)[54]	N[a]	N	N	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Vos et al (2009)[47]	Y	Y	N	Y	N	Y	Y	Y	N	N	Y	Y	Y	Y	NA	NA[b]	N	N	Y
Williamson et al (2015)[41]	Y	N	Y	Y	N	Y	Y	N	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y
Prescriptive CPRs																			
Cai et al (2011)[64]	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	N	N	N	Y	N	Y
Cleland et al (2007)[72]	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	Y	Y	Y
Hanney et al (2013)[65]	Y	N	N	Y	N	N	Y	N	Y	Y	Y	N	Y	N	N	N	Y	N	Y
Keating et al (2005)[37]	Y	Y	N	N	N	Y	Y	N	N	Y	Y	N	Y	N	N	N	N	N	N
Nee et al (2013)[66]	Y	N	N	Y	N	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y	Y	N	N
Puentedura et al (2012)[62]	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	N	N	N	Y	N	Y	Y	Y
Raney et al (2009)[73]	Y	Y	N	N	N	Y	Y	N	Y	Y	Y	N	N	N	N	N	Y	N	Y
Saavedra-Hernandez et al (2011)[63]	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	N	Y	N	N	N	Y	N	Y
Schellingerhout et al (2008)[74]	N[a]	N	N	Y	N	Y	Y	N	N	Y	Y	Y	Y	Y	N	Y	Y	Y	N
Tseng et al (2006)[61]	Y	N	N	N	Y	Y	Y	Y	N	Y	Y	N	Y	N	N	N	Y	N	Y

1 = Prospective design, 2 = Study site described, 3 = Justification for number of subjects reported, 4 = Representative sample, 5 = Important patient characteristics described, 6 = Selection of candidate predictor variables justified, 7 = Blinded predictor assessment, 8 = Predictor variables have demonstrated reliability, 9 = Outcome measure or reference standard has demonstrated reliability and validity, 10 = Blinded outcome assessment or reference standard, 11 = Mathematical techniques described, 12 = Reporting and handling of missing data described, 13 = At least 10 outcome events per independent variable in the final multivariable model, 14 = At least 10 outcome events per candidate predictor variable, 15 = Co-linearity of predictor variables tested, 16 = Continuous predictor variables are kept continuous in the multivariable analysis, 17 = Uncertainty in the accuracy of the CPR is described, 18 = Uncertainty in the post-test probability is described, 19 = CPR performance is non-paradoxical, Y = yes, N = no , NA = not applicable , [a] secondary analysis of prospective study , [b] multivariable analysis not performed , [c] included discrete variables only

Table 5 Methodological quality of CPR validation studies by CPR type

Study	Prospective validation in new patient population	Different clinical setting to derivation study	Representative sample	The rule is applied accurately	Assessment of the inter-rater reliability of the rule	Complete follow-up	Reporting and handling of missing data described	Accuracy of the rule described	Uncertainty in the accuracy of the CPR described	Uncertainty in the post-test probability is described
Prognostic CPRs										
Dagfinrud et al (2013)[19]	Y	Y	N	Y	NA[a]	Y	N	Y	Y	N
Kasch et al (2008)[53]	Y	Y	Y	N	N	N	N	N	N	Y
Kasch et al (2011)[56]	N	N	N	N	N	N	N	N	N	N
Kasch et al (2013)[52]	Y	N	Y	Y	N	N	N	Y	N	N
Radanov & Sturzenegger (1996)[55]	Y	N	N	Y	N	Y	N	N	N	N
Ritchie et al (2015)[49]	N[c]	N	Y	Y	NA[a]	Y	Y	Y	Y	Y
Schellingerhout et al (2010)[54]	N	N	Y	Y	NA[a]	Y	N	Y	Y	Y
Prescriptive CPRs										
Cleland et al (2010)[36]	Y	N	Y	N	N	N	Y	N	Y	N
Fritz & Brennan (2007)[34]	NA[b]	NA[b]	Y	Y	Y	Y	N	Y	N	N
Fritz et al (2014)[35]	Y	Y	Y	Y	N	N	Y	Y	N	N
Keating et al (2005)[37]	Y	N	N	Y	NA[a]	Y	N	Y	N	N

Y = yes, N = no, NA = not applicable, [a] self-report measures not requiring interpretation, [b] no derivation, [c] secondary analysis of prospective study

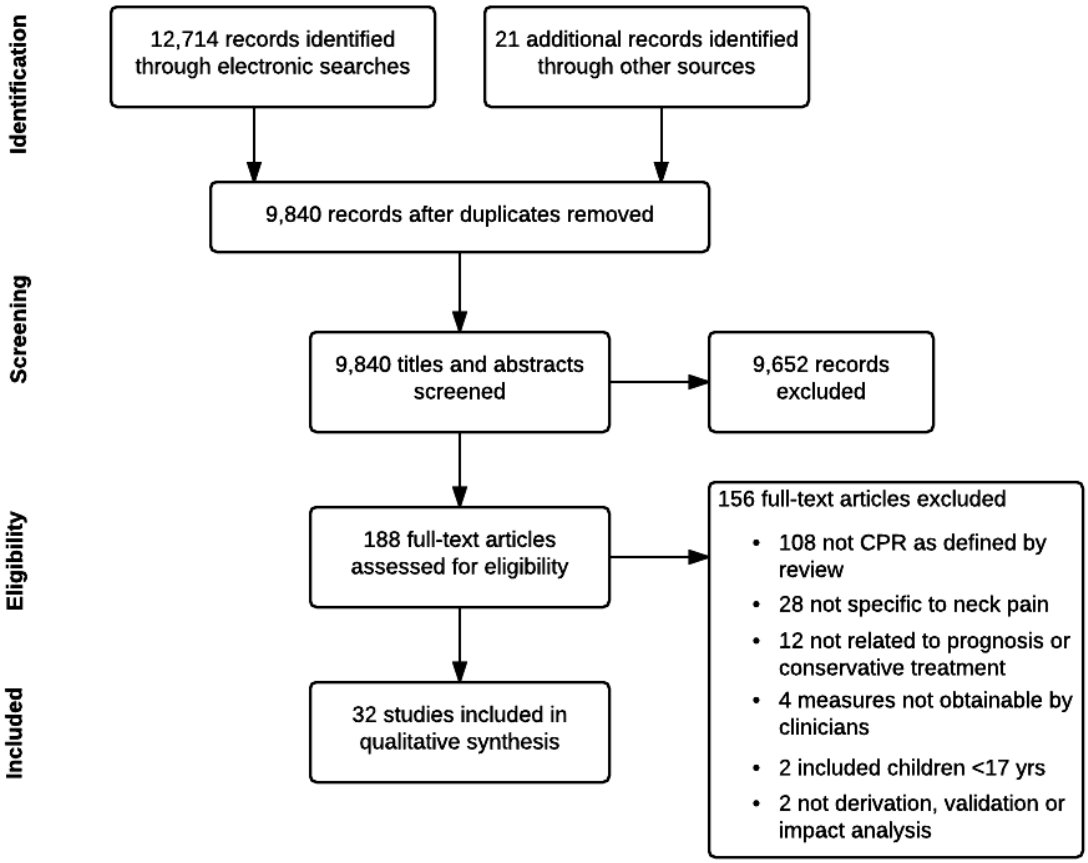

Figure 1 Flow chart of search strategy and study selection process

Supplementary Table 1 Study characteristics of prognostic CPRs by population: acute whiplash, non-specific neck pain, and cervical radiculopathy

CPR no	Publication	Stage	Study design	Sample	Included variables (n)	Primary outcome measures	Accuracy (95% CI)	Post-test probability (95% CI)
Acute whiplash								
1	Atherton et al (2006)[39]	Derivation	Cohort	n = 535, whiplash (QTF I-III), acute (median time since injury 8 days), presenting to emergency department, UK, median age 34yrs, 56% female	Pre-collision widespread pain, vehicle type other than car, ≥5 WAD symptoms, NDI ≥19, GHQ ≥6 (5 variables)	12 month self-report neck pain (yes/no) lasting ≥1 day in last week. Prevalence = 27% (n = 128)	Not reported	Not reported
2	Carroll et al (2006)[44]	Derivation	Cohort	n = 1858, whiplash, acute (≤6 weeks post injury and self-report recovery ≥3 on 6 point Likert scale (feeling some improvement to getting much worse), from motor vehicle insurance claimants and health providers, Canada, Mean age 39yrs, 68% female	PMI-passive coping subscale 21-30 & CES-D ≥16 (2 variables)	Time to recovery (specific measure not reported). Prevalence not reported	75% slower recovery if PMI 21-30 and CES-D ≥16 (adjusted HRR 0.25 (0.17-0.38))	Not reported
3	Gabel et al (2008)[42]	Derivation	Cohort	n = 30, WAD (no definition provided), duration of injury not reported, from general practitioner & primary care, location not reported, mean age 37yrs, 77% female	modified OMPQ ≥109 & cervical rotation at impact (2 variables)	6 month NDI >28% or self-report symptoms or impairments. prevalence = 30% (n = 9)	Sensitivity 100%, specificity 87%, +LR 7.7 (CIs not reported)	Not reported
4	Hartling et al (2002)[40]	Derivation	Cohort	n = 334, whiplash (not fracture or dislocation), acute (≤2 weeks post injury), presenting to emergency department, Canada, age range 18-70yrs, 64% female	Decision tree with: MVC occur other than at an intersection in the city, Upper back pain since MVC, Still experience neck pain, Still experience shoulder pain (4 variables)	6 month WAD pain classification[75] ≥3. Prevalence = 35% (n = 118)	Sensitivity 92% (87-97), specificity 51% (45 -58),	PPV 51% (44-57), NPV 92% (87-97)
5	Kasch et al (2001)[59]	Derivation	Case-control	n = 141, whiplash (traumatic neck injury), acute (≤1 week post injury), from emergency department, Denmark, mean age males (34yrs) and females (35yrs), 52% female	Total CROM <266 degrees, ≥7/15 non-painful complaints, baseline pain VAS ≥54/100mm (3 variables)	12 month self-report work handicap: reduced work hours & capacity from injury, dismissed, change in job or receiving disability	Sensitivity 30%, specificity 99% (CIs not reported)	PPV 75%, NPV 95% (CIs not reported)

						pension. Prevalence = 8% (n = 11)		
	Kasch et al (2008)[53]	Narrow validation *	Secondary analysis RCT	n = 625, whiplash (QTF I-III), acute (median duration 5 days), from emergency department or general practitioner, Denmark, mean age high risk (35yrs) and low risk (35yrs), 64% female,	Score chart with high risk ≥4/10. Includes total CROM, number of non-painful complaints (0-11), baseline pain VAS (0-10), gender (4 variables)	12 month self-report work handicap: >3 months sick leave during last 6 months, OR work inability in last month OR not working anymore because of accident. Prevalence: high risk = 19% (n = 78), low risk = 2% (n = 5)	Not reported	Not reported
	Kasch et al (2011)[56]	Narrow validation *[J]	Case-control	As per Kasch et al (2001)[59].	Score chart (0-10) to create strata (1-7). Includes total CROM, number of non-painful complaints (0-11), baseline pain VAS (0-10) (3 variables)	12 month work disability (specific measure not reported). Prevalence not reported	Risk assessment score: AUC 0.90 (0.74-1.0). Number of sick days and work disability at 12/12 associated with 7 strata (both p <0.001).	Not reported
	Kasch et al (2013)[52]	Narrow validation	Secondary analysis RCT	As per Kasch et al (2008)[53].	As per Kasch et al (2011)[56]	As per Kasch et al (2008)[53]. Prevalence not reported	+LR/-LR Stratum 1 = 1.0/-, Stratum 2 = 1.1/0.18, Stratum 3 = 1.8/0.24, Stratum 4 = 2.3/0.29, Stratum 5 = 2.9/0.43, Stratum 6 = 3.5/0.57 Stratum 7 = 7.8/0.73	Stratum 1 = 13%, Stratum 2 = 24%, Stratum 3 = 55%, Stratum 4 = 68%, Stratum 5 = 76%, Stratum 6 = 81%, Stratum 7 = 87%) (CIs not reported)
6	Nederhan	Derivation	Cohort	n = 90, whiplash (QTF I-II), acute (mean	NDI >15/50 & TSK ≥40	24 week NDI	+LR 4.3, AUC	83% (70-

	d et al (2004)[43]			duration 8.1 days), from emergency department, location not reported, mean age disabled (38yrs), recovered (33yrs), 57% female	(2 variables)	≥15/50. Prevalence = 33% (n = 27)	0.77(0.63-0.91)	91)
7	Radanov & Sturzenegger (1996)[55]	Derivation	Cohort	n = 117, whiplash (not fracture/dislocation), acute (mean 7 days post injury) referred from general practitioner, Switzerland, mean age 41yrs, 58% female	Equation including impaired neck movement, pre-traumatic headache, history of head trauma, age, initial neck pain intensity (0-10), initial headache intensity (0-10), FPI-N nervousness, FPI-N neuroticism, focused attention (measure not reported) (9 variables)	12 month recovered or symptomatic (specific measure not reported). Prevalence (symptomatic)= 24% (n = 28)	Not reported	96% (CIs not reported)
	As above	Narrow validation	Cohort	n = 16, whiplash (not fracture/dislocation), acute (mean 23 days post injury), referred from insurance company, Switzerland, mean age 33yrs, 25% female	As above	Recovered or symptomatic (specific measure not reported). Prevalence (symptomatic)= 44% (n = 7)	Not reported	88% (CIs not reported)
8	Ritchie et al (2013)[50]	Derivation	Secondary analysis cohort	n = 262, whiplash (QTF I-III), acute (<4 weeks duration), from emergency department, primary care practices & community, location not reported, mean age 37yrs, 65% female	Full recovery: initial NDI ≤32, age ≤35 yrs (2 variables) Ongoing disability: NDI ≥40, age ≥35 yrs, PDS hyperarousal subscale ≥6 (3 variables)	Full recovery: 12 month NDI ≤10%. 50% prevalence (n = 51) Ongoing disability: 12 month NDI ≥30%. Prevalence = 23% (n = 23)	Full recovery: sensitivity 45% (35-54), specificity 85% (77-90), +LR 2.9 (1.9-4.5), -LR 0.6 (0.5-0.8) Ongoing disability: sensitivity 44% (31-55), specificity 94% (89-96), +LR 7.0 (3.8-13), -LR 0.6 (0.5-0.7)	Full recovery: PPV 71% (59-80), NPV 65%) Ongoing disability: PPV 71% (55-84), NPV 82% (76-87)
	Ritchie et al (2015)[49]	Narrow validation^	Secondary analysis cohort	n = 101, whiplash (QTF II – different to derivation), acute (<4 weeks duration), from emergency department, primary care practices & community, Australia, mean age 34yrs, 66% female	As per Ritchie et al (2013)[50]	Full recovery: 6 month NDI ≤10%. Prevalence = 46% (n = 120) Ongoing disability: 6 month NDI ≥30%. Prevalence	Full recovery: sensitivity 55% (41-69), specificity 86% (73-94), +LR 3.9 (1.9-8.1), -LR 0.5 (0.4-0.7) Ongoing disability:	Full recovery: PPV 80% (63-92) Ongoing disability PPV 91%

						= 26% (n = 69)	sensitivity 44% (23-65), specificity 99% (93-100), +LR 34 (4.6-251), -LR 0.6 (0.4-0.8)	(59-99)
9	Williamson et al (2015)[41]	Derivation	Cohort	n = 430, whiplash (QTF I-III), acute (duration ≤6 weeks) from emergency department/referred to physiotherapy for RCT, location not reported, mean age 41yrs, 65% female	Baseline NDI ≥50%, self-predicted long (>6 months) or non-recovery, GHQ ≥4/12, PCQ-passive coping ≥5/12, CSOQ ≥6/15 (5 variables)	12 month NDI ≥30%. Prevalence = 30% (n = 136)	Relative risk for 1 factor = 3.5 (1.0-11), 2 factors = 7.4 (2.4-23), 3 factors = 8.1 (2.6-25), ≥4 factors = 16 (5.4-49)	Not reported
Non-specific neck pain								
10	Dagfinrud et al (2013)[19]	Broad validation [#]	Cohort	n ≈ 81, neck pain (no treatment in past 4 weeks), symptom duration 0 weeks to >1 year, seeking care of manual therapist, Norway, mean age 43yrs, 72% female	OMPQ ≥105 (1 variable, 21 items)	8 week NDI improvement ≥10%. Prevalence not reported	Sensitivity 18%, specificity 86%, -LR 0.95, +LR 1.3, AUC 0.60 (0.44, 0.75)	Not reported
11	Grooten et al (2007)[45]	Derivation	Cohort	n = 803, neck and shoulder pain in workers, symptom duration not reported, 'presenting to care', Sweden, mean age 42yrs, 65% female	≥2 of 3 biomechanical exposures: manual handling ≥50 newton^2 ≥60 minutes per day, working with hands above shoulders ≥30 minutes per day, working with vibrating tools ≥60min per day (3 variables)	4-6 year symptom free: pain intensity <3 on 11 point scale and pain-related disability <1/10[76]. Prevalence = 36% (n = 289)	Adjusted relative chance if ≥2 factors (compared with those unexposed to all 3) 0.61 (0.40-0.94).	Not reported
12	Landers et al (2008)[46]	Derivation	Cohort	n = 79, neck pain (not congenital instability), symptom duration <7 days to >7 weeks, presenting to physiotherapy, location not reported, mean age 50yrs, 71% female	≥2 of 5 CNOS variables[77]: palpation, simulation, cervical ROM, regional disturbances, overreaction (5 variables)	12 week NDI ≥15/50. Prevalence = 37% (n = 29)	Sensitivity 48%, specificity 97%, +LR 16, -LR 0.54, AUC 0.78 (0.67-0.90)	96% (CIs not reported)
13	Schellingerhout et al (2010)[54]	Derivation	Secondary analysis RCT	n = 468, non-specific neck pain (no disc herniation or rheumatological condition), symptom duration <1 month to >3 months, consulting physician in primary care, Netherlands, mean age 45yrs, 61% female	Score chart with ≥35 points: age, accompanying low back pain, traumatic cause neck pain, EuroQOL, headache, radiation of pain to elbow or shoulder,	6 month global perceived recovery ≥2 on 7 point Likert scale (slightly improved to worse than ever). Prevalence = 43% (n = 199)	Sensitivity 61% (55-68), specificity 61% (55-67), overall discriminative ability 0.66 (0.62-0.71)	PPV 54% (47-60)

					previous neck complaints, paid employment, pain intensity (9 variables)			
	As above	Broad validation	Secondary analysis RCT	n = 315, non-specific neck pain (no disc herniation or rheumatological condition), symptom duration <1 month to >3 months, consulting physicians in primary care, UK, mean age 49yrs, 64% female	As above	6 month global perceived recovery ≥2 on 7 point Likert scale (slightly improved to worse than ever). Prevalence = 39% (n = 124)	Sensitivity 63% (54-71), specificity 60% (53-67), overall discriminative ability 0.66 (0.59-0.72)	PPV 51% (43-59)
14	Vos et al (2009)[47]	Derivation	Cohort	n = 143, neck pain, acute (duration ≤6 weeks), from general practice, Netherlands, mean age 40yrs, 64% female	Modified ALBPSQ ≥72/200 (1 variable, 20 items)	12 month sick leave >7 days. Prevalence = 22% (n = 31)	Sensitivity 77%, specificity 62%, AUC 0.66 (0.56-0.76)	PPV 81%, NPV 57% (CIs not reported)
Cervical Radiculopathy								
15	Cleland et al (2007)[48]	Derivation	Cohort	n = 96, cervical radiculopathy or neck/arm pain and positive on diagnostic CPR, mean symptom duration 80 days, referred to physiotherapy from physician, location not reported, mean age 51yrs, 64% female	≥3 of 4 variables: age <54yrs, dominant arm not affected, looking down does not worsen symptoms, multimodal treatment at least 50% visits (manual therapy, cervical traction, deep neck flexor strengthen) (4 variables)	28 day improvement in all of: NDI ≥7, NPRS ≥ 2 points, PSFS ≥2 points, GROC ≥+5. Prevalence = 52% (n = 31)	Sensitivity 68% (55-81), specificity 87% (77-97), +LR 5.2 (2.4-11)	85% (CIs not reported)

* Validation of derived CPR that is altered from its original form, [+] Derivation of CPR not successful by review definition, [^] Follow-up period altered from derivation model, [#] Validation of model derived for use in different population, [∫] Used same population as derivation study, CPR = clinical prediction rule, CI = confidence interval, WAD = Whiplash Associated Disorder, NDI = Neck Disability Index, GHQ = General Health Questionnaire, QTF = Quebec Task Force Whiplash Classification, IQR = Interquartile Range, PMI = Pain Management Inventory, CES-D = Centre for Epidemiologic Studies Depression Scale, OMPQ = Orebro Musculoskeletal Pain Questionnaire, +LR = positive likelihood ratio, -LR = negative likelihood ratio, PPV = positive predictive value, NPV = negative predictive value, MVC = motor vehicle crash, CROM = cervical range of motion, RCT = randomised controlled trial, VAS = visual analogue scale, AUC = area under the curve, TSK = Tampa Scale for Kinesiophobia, FPI = Freiburg Personality Inventory, PDS = Posttraumatic Stress Diagnostic Scale, PCQ = Pain Coping Questionnaire, CSOQ = Cervical Spine Outcomes Questionnaire, CNOS = cervical nonorganic signs, ALBPSQ = Acute Low Back Pain Screening Questionnaire, NPRS = Numeric Pain Rating Scale, PSFS = Patient Specific Functional Scale, GROC = Global Rating of Change Scale

Supplementary Table 2 Study characteristics of prescriptive CPRs by primary treatment modality: spinal manipulation, spinal traction, exercise program and mixed modes

CPR no	Publication	Stage	Study design	Sample	Treatment	Included variables (n)	Primary outcome measure	Accuracy (95% CI)	Post-test probability (95% CI)
Spinal manipulation									
16	Cleland et al (2007)[72]	Derivation	Single group	n = 78, mechanical neck pain with NDI ≥10% (not cervical canal stenosis or nerve root compression), mean symptom duration 80 days, referred to physiotherapy, USA, mean age 42.0yrs, 68% female	Thoracic manipulation (plus CROM exercise)	≥3 of 6 variables: symptom duration <30 days, no symptoms distal to shoulder, looking up does not aggravate symptoms, FABQ-PA <12, diminished upper thoracic spine kyphosis, extension CROM <30 degrees (6 variables)	GROC ≥+5 at end of 2nd session (or 3rd if unsuccessful at 2nd). Prevalence = 55% (n = 42)	Sensitivity 76% (67-82), specificity 86% (75-93), +LR 5.5 (2.7-12)	86% (74-94)
	Cleland et al (2010)[36]	Broad validation*	RCT	n = 140, neck pain with NDI ≥10% (not cervical canal stenosis or nerve root compression), mean symptom duration 64 days, referred to physiotherapy, USA, mean age 40yrs, 69% female	As per Cleland et al (2007)[72]	As per Cleland et al (2007)[72]	Change in NDI and NPRS at 1 week, 4 weeks and 6 months (different to derivation). Prevalence not reported	Non-significant difference between those positive and negative on CPR (except for disability at 4/52, p=0.05). No difference between treatment effects for each treatment by status on CPR	Results did not support the validity of the CPR
17	Puentedura et al (2012)[62]	Derivation	Single group	n = 82, mechanical neck pain with NDI ≥10/50 (not rheumatological condition, acute whiplash, nerve root compression or CNS involvement), mean symptom duration 293 days, from 4 physiotherapy clinics, USA & Spain, mean age 38yrs, 59% female	C3-7 cervical manipulation (plus rotation CROM exercise)	≥3 of 4 variables: symptom duration <38 days, positive expectation that manipulation will help, side-to-side difference in rotation CROM ≥10 degrees, pain on posterior-anterior spring testing to middle cervical spine (4 variables)	GROC ≥+5 at end of 1 week (1-2 sessions). Prevalence = 39% (n = 32)	Sensitivity 81% (63-92), specificity 94% (82-98), +LR 14 (4.5-41)	90% (74-96)

18	Saavedra-Hernandez et al (2011)[63]	Derivation	Single group	n = 81, mechanical neck pain (not whiplash, radiculopathy, fibromyalgia, CNS involvement, nerve root compression or previous treatment with spinal manipulation), mean symptom duration 1703 ± 1726 days, referred for therapy, Spain, mean age 39yrs, 70% female	Cervical and thoracic manipulation	≥4 of 5 variables: NPRS >4.5/11, extension CROM <46 degrees, hypomobility at T1, negative UL tension test, female sex (5 variables)	GROC ≥+5 at end of 2nd session (or 3rd if unsuccessful at 2nd). Prevalence = 62% (n = 50)	Sensitivity 12% (5-25), specificity 94% (77-99), +LR 1.9 (0.40-8.6)	75% (CIs not reported)
19	Tseng et al (2006)[61]	Derivation	Single group	n = 100, neck pain (diagnosis of spondylosis ± radiculopathy, herniated disc, myofascial pain syndrome or cervicogenic headache), symptom duration <3 weeks to >3 months, referred to physiotherapy, location not reported, mean age 46yrs, 66% female	C0-7 cervical manipulation	≥4 of 6 variables: initial NDI <11.5/50, having bilateral involvement pattern, not performing sedentary work >5 hours per day, feeling better while moving neck, not feeling worse while extending neck, diagnosis of spondylosis without radiculopathy (6 variables)	NPRS-11 reduction ≥50% OR perceived improvement ≥4 on 15 point Likert scale (much improved) OR satisfaction level 5 of 5 point scale (very satisfied) at end of session. Prevalence = 60% (n = 60)	Sensitivity 40% (28-52), specificity 93% (84-100), +LR 5.3 (1.7-17)	89% (CIs not reported)
Spinal traction									
20	Cai et al (2011)[64]	Derivation	Single group	n = 103, neck pain (and/or cervical numbness with radicular pain, numbness or headache, diagnosis of spondylosis or degenerative change), mean duration 30.6 weeks, referred to physiotherapy, location not reported, mean age 49yrs, 40% female	Home-based mechanical cervical traction	≥3 of 4 variables: FABQ-W <13, baseline NPS ≥7, positive cervical distraction test, pain below shoulder (4 variables)	NPS reduction ≥50% OR NDI reduction ≥50% OR global rating of perceived recovery ≥much improved (7 point Likert scale, much worse to completely recovered) at 2 weeks#. Prevalence = 46% (n = 47)	Sensitivity 51% (36-66), specificity 89% (77-96), +LR 4.8 (2.1-10.7)	80%

21	Raney et al (2009)[73]	Derivation	Single group	n = 68, neck pain with NDI ≥20%, mean duration 292.4 days, referred to physiotherapy, USA, mean age 48yrs, 56% female	Clinic-based mechanical cervical traction (+ postural & deep neck flexor exercise)	≥3 of 5 variables: patient reported peripheralisation on C4-7 mobility tests, positive shoulder abduction test, age ≥55 yrs, positive upper limb tension test A, positive neck distraction test (5 variables)	GROC ≥+6 at end of 6th session (3 weeks). Prevalence = 44% (n = 30)	Sensitivity 63% (46-78), specificity 87% (73-94), +LR 4.8 (2.2-11), -LR 0.42 (0.25-0.65)	79% (CIs not reported)
	Fritz et al (2014)[35]	Broad validation	RCT	n = 54, neck pain with NDI ≥10% (different to derivation) and symptoms extending to acromioclavicular joint or superior scapular, median duration 53 days, from physicians and physiotherapists, USA, mean age 47yrs, 54% female		As per Raney et al (2009)[73]	Change in NDI at 4 weeks and 6 and 12 months. Prevalence not reported	No significant difference between those positive and negative on CPR (p ≥0.05)	Results did not support the validity of the CPR
Exercise program									
22	Hanney et al (2003)[65]	Derivation	Single group	n = 91, non-specific neck pain with NDI ≥10/50, mean symptom duration 286.6 days, from physiotherapy, location not reported, mean age 46yrs, 76% female	Standardised neck and shoulder stretching and strengthening	≥4 of 5 variables: NDI <18/50, shoulder protraction during static postural assessment, patient does not bicycle for exercise, side bending CROM <32 degrees, FABQ-PA subscale <15 (5 variables)	GROC ≥+4 at treatment completion (6 weeks). Prevalence = 55% (n = 50)	Sensitivity 58% (42-70), specificity 81% (66-98), +LR 3.0 (1.5-5.7), -LR 0.52 (0.36-0.74)	78% (CIs not reported)
23	Keating et al (2005)[37]	Derivation	Single group	n = 97, chronic neck pain, median symptom duration 36 months, seeking physiotherapy care, Australia, mean age 41yrs, 65% female	Tailored neck strengthening	NDI reading question >1/5 and NDI lifting question >1/5 (2 variables)	Reduction in NDI ≥14 at discharge (median time 6 weeks). Prevalence = 56% (n = 54)	Not reported	PPV 64%, NPV 74% (CIs not reported)
	As above	Narrow validation	Single group	n = 192, chronic neck pain, median symptom duration 60 months, seeking physiotherapy care, Australia, mean age 41yrs, 67% female	As above	As above	Reduction in NDI ≥14 at discharge (median time 6 weeks). Prevalence not	Not reported	PPV 56%, NPV 74% (CIs not reported

							reported		)
Mixed modes									
24	Fritz & Brennan (2007)[34]	Not applicable [+]	Single group	n = 274, neck pain, median symptom duration 48 days, receiving physiotherapy treatment, USA, 44yrs, 74% female	Classification based physiotherapy	Classification into 1 of 5 subgroups with matched intervention: mobility (manual therapy), exercise & conditioning (strengthening), pain control (mobilization & ROM), headache (strengthening & manipulation), centralisation (retraction & traction)	NDI reduction ≥8 points, change in NPRS-11, number of visits. Prevalence = 73% (n = 83)	Matched intervention superior to unmatched for change in NDI (mean difference 5.6 (2.6-8.6)) and NPRS (0.74 (0.21-1.3)), but not sessions required	Not reported
25	Nee et al (2013)[66]	Derivation	Single group (secondary analysis RCT)	n = 40, neck and nerve-related unilateral arm pain with NPRS ≥3, duration ≥4 weeks and ≥4 weeks pain-free prior (not traumatic onset or ≥2 neurological signs), from general community, location not reported, duration median 26 weeks, mean age 47yrs, 63% female	Neural tissue management (education, manual therapy and nerve gliding exercises)	>89/120 (responder) or ≤71/120 (non-responder) points from score chart including S-LANSS, age and ULNT1(median) elbow extension deficit	Responder: GROC ≥+4 at 3-4 weeks. Prevalence = 53% (n = 21)	Responder: sensitivity 43% (25-64), specificity 95% (75-99), +LR 8.1 (1.1-58) Non-responder: sensitivity 5% (1-23), specificity 47% (27-68), +LR 0.09 (0.01-0.64). AUC 0.85 (0.72-0.98)	Responder: 90% (56-98)[ſ] Non responder: 9% (1-42)[ſ]
26	Schellingerhout et al (2008)[74]	Derivation	Secondary analysis 3 RCTs	n = 329, non-specific neck pain (no disc herniation or rheumatological condition), symptom duration <1 month to >3 months, consulting primary care physician, Netherlands, mean age 46yrs, 61% female	Manual therapy, physiotherapy or advice	Decision tree matched to therapy including pain (NRS ≤7/>7), accompanying LBP (yes/no), age ≤50/>50 (3 variables)	Global perceived recovery <2 on 7 point scale (0 = completely recovered, 6 = worse than ever) at end of treatment (short-term) and 52 weeks (long-term). Short term prevalence = 52% (n = 165), long term	Short-term: probability of recovery increased by 26% (14-38) from 32%. Long-term: probability of recovery increased by 17% (3.8-29) from 50%	Not reported

							prevalence = 62% (n = 195)

*Included further sessions of strength and stretch exercises additional to derivation model treatment, [#]Reduction by 5 points required if initial NDI score >5 and <10. Reduction by 2 points required if initial NPS score = 2. Measure not used if NDI score ≤5 or NPS ≤1, [+] Examination of non-mathematically derived classification system
[j] Values in figure inconsistent with those in text, CPR = clinical prediction rule, no = number, CI = confidence interval, CROM = cervical range of motion, NDI = Neck Disability Index, GROC = Global Rating of Change Scale (15 point: -7 (a very great deal worse) to +7 (a very great deal better)), RCT = Randomised Controlled Trial, FABQ-PA = Fear Avoidance Belief Questionnaire – Physical activity subscale, FABQ-W =, Fear Avoidance Belief Questionnaire – Work subscale, +LR = positive likelihood ratio, CNS = central nervous system, NPRS = Numeric Pain Rating Scale, NPS = Numerical Pain Scale, IQR = Interquartile Range, S-LANSS = Self-report Leeds Assessment of Neuropathic Symptoms and Signs , AUC = area under the curve, NRS = Numerical Rating Scale

APPENDIX A. Search strategies

PubMed	
1	(((Neck[Mesh] OR Neck OR Necks OR Cervical Vertebrae[Mesh] OR "Cervical Vertebrae" OR Neck Muscles[Mesh] OR "Neck Muscles") AND (Pain[Mesh] OR Pain OR Pains OR Aches OR Ache OR Sore OR Disability)) OR (Neck Injuries[Mesh] OR Whiplash Injuries[Mesh] OR Radiculopathy[Mesh] OR "Neck Injuries" OR "Neck Injury" OR Whiplash OR Radiculopathies OR Radiculopathy))
2	((validat* OR predict*[Title] OR rule*) OR (predict* AND (outcome* OR risk OR risks OR model OR modelling OR models OR modelled)) OR ((history OR variable* OR criteria OR score OR scores OR scoring OR scored OR characteristic* OR finding* OR factor*) AND (predict* OR model OR modelling OR models OR modelled OR decision* OR identif* OR prognos*)) OR (decision* AND (model OR modelling OR models OR modelled OR clinical* OR logistic regression[Mesh])) OR (prognostic AND (history OR variable* OR criteria OR score OR scores OR scoring OR scored OR characteristic* OR finding* OR factor* OR model OR modelling OR models OR modelled)))
3	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
4	Fractures, Bone[Mesh] OR Neoplasms[Mesh] OR Fracture OR Fractures OR Neoplasm OR Neoplasms OR Cancer OR Cancers
5	(1 AND (2 OR 3)) NOT 4
6	Limit 5 to English language

Embase	
1	'neck'/exp OR neck OR 'cervical spine'/exp OR 'cervical spine' AND ('pain'/exp OR pain OR ache OR sore OR 'disability'/exp OR disability) OR ('neck'/exp OR neck AND ('injury'/exp OR injury)) OR 'cervicobrachial neuralgia'/exp OR 'cervicobrachial neuralgia'
2	validat* OR predict*:ti OR rule*OR (predict* AND (outcome* OR 'risk'/exp OR risk OR risks OR 'model'/exp OR model OR 'modelling'/exp OR modelling OR models OR modelled)) OR ('history'/exp OR history OR variable* OR criteria OR score OR scores OR scoring OR scored OR characteristic* OR finding* OR factor* AND (predict* OR 'model'/exp OR model OR 'modelling'/exp OR modelling OR models OR modelled OR decision* OR identif* OR prognos*)) OR (decision* AND ('model'/exp OR model OR 'modelling'/exp OR modelling OR models OR modelled OR clinical* OR 'logistic regression analysis'/exp OR 'logistic regression analysis')) OR (prognostic AND ('history'/exp OR history OR variable* OR criteria OR score OR scores OR scoring OR scored OR characteristic* OR finding* OR factor* OR 'model'/exp OR model OR 'modelling'/exp OR modelling OR models OR modelled))
3	'stratified sample'/exp OR 'stratified sample' OR 'receiver operating characteristic'/exp OR 'receiver operating characteristic' OR discrimination OR 'kappa statistics'/exp OR 'kappa statistics' OR 'statistical significance'/exp OR 'statistical significance' OR 'area under the curve'/exp OR 'area under the curve' OR 'auc'/exp OR auc OR 'calibration'/exp OR 'calibration' OR indices OR 'classification algorithm'/exp OR 'classification algorithm' OR multivariable
4	'fracture'/exp OR fracture OR 'neoplasm'/exp OR neoplasm
5	(1 AND (2 OR 3)) NOT 4
6	limit 5 to humans, Embase, English language, articles or articles in press

CINAHL Plus via EBSCO	
1	((MH "Neck") OR Neck OR Necks OR (MH "Cervical Vertebrae") OR "Cervical Vertebrae" OR (MH "Neck Muscles") OR "Neck Muscles") AND ((MH "Pain") OR Pain OR Pains OR Aches OR Ache OR Sore OR Disability)) OR ((MH "Neck Injuries") OR (MH "Whiplash Injuries") OR (MH "Radiculopathy") OR "Neck Injuries" OR "Neck Injury" OR Whiplash OR Radiculopathies OR Radiculopathy)
2	(validat* OR ti predict* 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 (MH "logistic regression+"))) OR (prognostic AND (history OR variable* OR criteria OR scor* OR characteristic* OR finding* OR factor* OR model*)) OR (stratification OR (MH "ROC Curve") 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)
3	(MH "Fractures, Bone") OR (MH "Neoplasms") OR Fracture OR Fractures OR Neoplasm OR Neoplasms OR Cancer OR Cancers
4	(1 AND 2) NOT 3
5	Limit 4 to humans, academic journals, English language and exclude MEDLINE records

AMED via Ovid	
1	((exp neck/ OR neck$ OR exp cervical vertebrae/ OR cervical vertebrae OR cervical spine OR exp neck muscles/ OR neck muscle$) AND (exp pain/ OR pain$ OR ache$ OR Sore OR Disability)) OR (exp neck injuries/ or neck injur$ OR exp whiplash injuries/ OR whiplash injur$ OR whiplash OR cervical

	radiculopathy$ or radiculopathy$)
2	(Validat$ OR Predict$.ti. OR Rule$) 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$))
3	Stratification OR ROC Curve OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	Fracture$ or exp neoplasms/ or neoplasm$ or cancer$
5	(1 AND (2 OR 3)) NOT 4
6	Limit 5 to English language and journal articles

PEDro	
1	“clinical prediction” (abstract and title)
2	head and neck (body part)
3	clinical trial (method)
4	1 AND 2 AND 3

Cochrane Library	
1	(((MeSH descriptor "Neck" or Neck* or MeSH descriptor "Cervical Vertebrae" or "Cervical Vertebrae" or MeSH descriptor "Neck Muscles" or "Neck Muscles") and (MeSH descriptor Pain or Pain* or Ache* or Sore or Disability)) or (MeSH descriptor "Neck Injuries" or MeSH descriptor "Whiplash Injuries" or MeSH descriptor "Radiculopathy" or "Neck Injuries" or "Neck Injury" or Whiplash or Radiculopathy))
2	(validat* or "predict":ti 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 (MeSH descriptor Logistic Models, explode all trees))) or (prognostic and (history or variable* or criteria or scor* or characteristic* or finding* or factor* or model*))
3	stratification or MeSH descriptor "ROC Curve" 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
4	MeSH descriptor "Fractures, Bone" or MeSH descriptor "Neoplasms" or Fracture* or Neoplasm* or Cancer*
5	(1 AND (2 OR 3)) NOT 4
6	Limit 5 to trials