Journal of Physiotherapy 64 (2018) 16–23

Journal of PHYSIOTHERAPY

journal homepage: www.elsevier.com/locate/jphys

Research

Few promising multivariable prognostic models exist for recovery of people with non-specific neck pain in musculoskeletal primary care: a systematic review

Roel W Wingbermühle [a,b], Emiel van Trijffel [a,c], Paul M Nelissen [a], Bart Koes [b], Arianne P Verhagen [b]

[a] SOMT University of Physiotherapy, Amersfoort; [b] Erasmus University Medical Centre, Rotterdam, The Netherlands; [c] Vrije Universiteit Brussel, Brussels, Belgium

KEY WORDS

Primary care
Systematic review
Neck pain
Multivariable prognostic models
Prognosis

ABSTRACT

Question: Which multivariable prognostic model(s) for recovery in people with neck pain can be used in primary care? **Design**: Systematic review of studies evaluating multivariable prognostic models. **Participants**: People with non-specific neck pain presenting at primary care. **Determinants**: Baseline characteristics of the participants. **Outcome measures**: Recovery measured as pain reduction, reduced disability, or perceived recovery at short-term and long-term follow-up. **Results**: Fifty-three publications were included, of which 46 were derivation studies, four were validation studies, and three concerned combined studies. The derivation studies presented 99 multivariate models, all of which were at high risk of bias. Three externally validated models generated usable models in low risk of bias studies. One predicted recovery in non-specific neck pain, while two concerned participants with whiplash-associated disorders (WAD). Discriminative ability of the non-specific neck pain model was area under the curve (AUC) 0.65 (95% CI 0.59 to 0.71). For the first WAD model, discriminative ability was AUC 0.85 (95% CI 0.79 to 0.91). For the second WAD model, specificity was 99% (95% CI 93 to 100) and sensitivity was 44% (95% CI 23 to 65) for prediction of non-recovery, and 86% (95% CI 73 to 94) and 55% (95% CI 41 to 69) for prediction of recovery, respectively. Initial Neck Disability Index scores and age were identified as consistent prognostic factors in these three models. **Conclusion**: Three externally validated models were found to be usable and to have low risk of bias, of which two showed acceptable discriminative properties for predicting recovery in people with neck pain. These three models need further validation and evaluation of their clinical impact before their broad clinical use can be advocated. **Registration**: PROSPERO CRD42016042204. **[Wingbermühle RW, van Trijffel E, Nelissen PM, Koes B, Verhagen AP (2018) Few promising multivariable prognostic models exist for recovery of people with non-specific neck pain in musculoskeletal primary care: a systematic review. *Journal of Physiotherapy* 64: 16–23]**

Introduction

Globally, neck pain is one of the main contributors to years lived with disability.[1,2] Improvements in pain and disability typically occur in the first weeks after the onset of an episode of neck pain, but residual pain and disability beyond this time are often of substantial severity and persist for at least 1 year.[3] High baseline neck pain intensity and disability scores have been identified as predictors for poor outcome in people with neck pain.[4] Cost-effectiveness and short-term beneficial effects of non-invasive primary care treatment have been reported, while long-term effects are still limited.[5–8] Subgrouping of people with neck pain based on their prognosis may enhance treatment outcomes by enabling tailored treatment and management strategies.[9–11] High-quality research on neck pain prognosis has been a research priority for over a decade.[12]

A fundamental shift in clinical practice has been proposed towards the prospective relationships between phenotypic, genomic, and environmental assessment of patients.[13] It is argued that prognostic profiles allow a more wholistic view and can better manage subjectively reported health problems than diagnostic labels.[13] These prognostic profiles should also more accurately mirror daily practice.[14]

Prognostic factors can be developed based on demographic factors, disease characteristics, or factors derived from history taking, physical examination, or additional examinations (such as imaging, blood assays, urine tests or other biological measurements).[15] Multiple factors are likely to interact with each other, so multivariable prognostic models that consider correlations between predictors have been proposed.[4,16–18] Development of multivariable prognostic models consists of three consecutive stages: developing the model (derivation); validating its performance in new patients (external validation); and studying its clinical impact (impact analysis).[17,19]

Numerous multivariable prognostic models in musculoskeletal primary care for people with neck pain have been developed. To our knowledge, these models have not been evaluated systematically using tools specifically designed to assess quality and

https://doi.org/10.1016/j.jphys.2017.11.013

usability of primary multivariable prognostic model studies included in a systematic review.

Several systematic reviews have been conducted to summarise the value of prognostic models in the musculoskeletal domain,[20–22] with one focusing on neck pain alone.[23] These reviews concluded that the methodological quality of the included studies was often poor to moderate, validation studies are rare, and routine clinical use is therefore not supported. Methodological quality was assessed in these systematic reviews using tools not specifically designed for assessing the quality of prediction models. Only recently, PROBAST (Prediction model study Risk Of Bias Assessment Tool) has become available; it is designed to assess the risk of bias and concerns about applicability of studies that develop and/or validate a multivariable prediction model when they are included in systematic reviews.[24–26]

To our knowledge, no systematic review on multivariable prognostic models for recovery (pain reduction, reduced disability, or perceived recovery) of people of all ages presenting in primary care with neck pain has been conducted using up-to-date methodology. The aim of this systematic review was to summarise the validity and applicability of multivariable prognostic models for recovery in people with neck pain in primary care.

Therefore, the specific research question for this systematic review was:

> Which multivariable prognostic model(s) for recovery in people with neck pain can be used in primary care?

Box 1. Inclusion criteria.

Models
- Constructed with multivariable analysis
- Combination of at least two predictors
- Any stage of development

Design
- Prospective cohort studies
- Randomised, controlled trials

Participants
- People of any age
- Non-serious specific or non-specific neck pain at any stage[a]

Determinants
- Baseline characteristics at intake
- Applicable to and easily obtained in non-invasive musculoskeletal primary care

Outcome to be predicted
- Pain
- Disability
- Perceived recovery

[a] Neck pain was defined as pain located in the anatomic region of the neck from the linea nuchea superior to the spina scapula, with or without radiation to the trunk or upper limb.[33,34] Non-specific neck pain was defined as neck pain without an identified pathological basis. Non-serious neck pain was defined as neck pain with an identified pathological basis, but with no contra-indication for musculoskeletal primary care.[35]

Method

Identification and selection of studies

MEDLINE, EMBASE, and CINAHL databases were searched to retrieve all relevant studies on multivariable prognostic models for recovery of neck pain from inception up to May 3, 2016. This search was based on a validated strategy adapted for the purpose of this study.[20,27,28] The full search strategy is listed in Appendix 1 on the eAddenda. De-duplication was performed in Mendeley and hand-checked.[29] No language restrictions were imposed. Additional manual searching of reference lists of all included studies was performed.

To be eligible for inclusion, studies had to generate multivariable prognostic models using data from prospective cohort studies and randomised, controlled trials on participants of any age with non-serious specific and non-specific neck pain. Models in all stages of their development were considered. Models were defined as those constructed by multivariable analysis from a combination of at least two predictors associated with a particular outcome, while derived models could contain one remaining variable.[17,30,31,32] All baseline characteristics that are feasible to measure in primary care were considered as potential predictors. Studies were included when the outcome concerned pain reduction, reduced disability, or perceived recovery at any time of follow-up. The inclusion criteria are summarised in Box 1. Studies aimed at (cost-)effectiveness, side effects, or developing a questionnaire were excluded. Studies using clinical procedures involving skin penetration like injection, acupuncture, or dry needling were also excluded.

Two reviewers (RW, PN) independently screened records for possibly relevant studies based on title and abstract. Subsequently, full texts of potentially relevant articles were independently assessed for eligibility. Discrepancies between reviewers were resolved through discussion or by a third reviewer (APV).

Assessment of characteristics of studies

Quality

Quality of the selected studies was assessed using the pre-publication version of PROBAST.[36] PROBAST was developed using a Delphi process involving 40 experts in the fields of systematic review methodology and prediction research. It was designed to assess risk of bias, applicability, and usability of multivariable prediction model studies included in a systematic review using a similar domain-based approach as the revised tool for the quality assessment of diagnostic accuracy studies (QUADAS-2). Judgements on high, low, or unclear risk of bias for reported estimates of the model's predictive performance were made for five key domains (participant selection, predictors, outcome, sample size and participant flow, and analysis) after judgement of signalling questions. As the signalling question was to determine whether there was a reasonable number of outcome events in a logistic regression, the number of events in the smallest group was divided by the total degrees of freedom used during the whole modelling process. Counting degrees of freedom was based on each time a variable or its category was tested on the outcome. Univariable predictors were considered here as part of the whole modelling process if they were selected based on their p-value. Rating was according to the 'rule of thumb' of 10 events per variable.[37] For linear regression, the number of participants was divided by the number of predictors. High, low, or unclear concerns about applicability regarding the review question were made in a similar structure for three key domains (participant selection, predictors, and outcome). An overall judgement about risk of bias and applicability of the prediction model evaluation was reached based on a separate summative rating across all domains for derivation and validation studies according to the PROBAST criteria. Finally, a model's usability was rated for its presentation with sufficient detail to be used in the intended context and target population.

Two reviewers (RW, PN) independently assessed the quality of the selected studies. Discrepancies and unclear items were resolved through discussion or, if necessary, adjudication by a third reviewer (APV). Percentage agreement and Cohen's kappa in a 2x2 contingency table were used to describe the level of agreement between the two reviewers for the judgements of the risk of bias and applicability domains. For this purpose, 'high' and 'unclear' ratings were collapsed into one category. Rating of models within the same study were combined into one variable per reviewer, if ratings were the same.

Data extraction

In accordance with the CHARMS (CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) checklist for prediction model development studies, data were extracted from the included studies on: study setting, country, and dates; participants' condition and treatment; number of variables and events; predictors in the model; predicted outcome and follow-up; model performance and stage; clinical measures; and model presentation.[38] Data extraction was performed independently by two reviewers (RW, PN), and randomly crosschecked by a third reviewer (APV).

Data analysis and evaluation

A qualitative synthesis was performed to evaluate whether a model was ready for clinical use by analysing the model's risk of bias, applicability, and usability as related to its performance accuracy. Analyses were conducted separately for derivation studies and validation studies. For subdividing the studies according to study stage, validation performed with non-random split data (type 2b) was considered as external validation.[18,39]

A model was judged to be ready for clinical use if it was usable and externally validated in a study with overall low risk of bias, while showing acceptable discriminative performance. Prediction models were accurate if they were able to discriminate people with and without the outcome.[40] Model discriminative performance was considered acceptable if the area under the curve (AUC) of the receiver operating characteristic (ROC) curve for continuous outcome or c-statistic for binary outcome was $\geq$0.7.[41] The ROC curve plots the model's true-positive prediction rate (sensitivity) versus the false-positive prediction rate (one minus the specificity) over all possible discrimination thresholds of predicted probability of the occurrence of the outcome. The c-statistic is comparable to the AUC for binary outcome, and is the proportion of pairs – one individual with and one individual without the outcome – in which the individual who experienced the outcome had a higher probability of experiencing the outcome than the individual who did not experience the outcome, as predicted by the model.[40] In addition, we searched for prognostic factors consistently appearing in final models from low risk of bias studies.

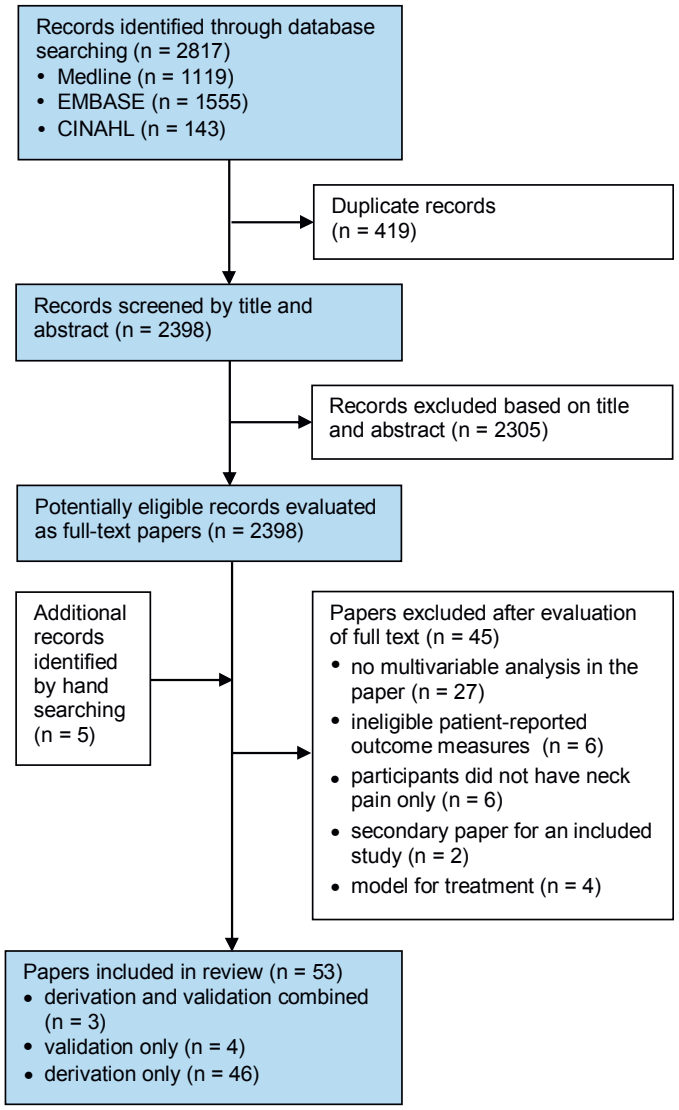

Figure 1. Flow of studies through the review.

Results

Flow of studies through the review

Searching MEDLINE, EMBASE, and CINAHL initially yielded 1119, 1554, and 143 records, respectively. After removal of duplicate citations, 2398 remained. Of these, 2305 records were excluded based on title and abstract. Hand searching added five potentially relevant publications, so a total of 98 full-text articles were evaluated for eligibility. Forty-five studies, of which 27 did not involve multivariable analysis, were excluded. Fifty-three studies met the selection criteria; 46 of these were derivation studies, while four were validation studies only, and three combined derivation and validation in one publication (Figure 1).

Characteristics of included studies

The characteristics of the 46 included derivation studies are presented in Appendix 2. The characteristics of the four validation studies and the three combined studies are presented in Appendix 3. See the eAddenda for Appendices 2 and 3.

Derivation studies

The mean age of participants in the derivation studies ranged from 30 to 65 years. Mean symptom duration at baseline ranged from 60 days to 108 months. Follow-up for outcome measurement among the included derivation studies ranged between 1 week and 5 years. Outcomes were measured using various patient-rated disability scales, global rating of change, or pain scales.

In total, 99 models were derived in 49 studies (excluding two models newly developed in a validation study).[42] Twenty-six studies described 58 models concerning participants with Whiplash-Associated Disorders (WAD); 35 acute, three subacute, six chronic, and 14 of any duration or unknown. Twenty-three studies described 41 models concerning participants with neck pain conditions; three acute, five subacute, six chronic, four with or without arm symptoms, two nerve-related arm pain, and 21 of any duration or unknown. The number of predictors in the final models varied from 1 to 10. The included derivation studies assessed a variety of types of predictors, such as history variables (eg, age, gender, pain/symptoms, symptom duration, disability, psychosocial, contextual) and physical examination variables (eg, range of motion, pain provocation, pain or temperature threshold). Twelve models were presented as a score chart, nomogram, prediction or decision rule.

Quality

Judgements about risk of bias, applicability, and usability are shown in Tables 1 and 2. Agreement between the two reviewers for judgements of the five risk of bias domains (participant selection, predictors, outcome, sample size and participant flow, and analysis) was 71, 69, 51, 98 and 92%, respectively. In the outcome

Table 1
Risk of bias, applicability, and model usability among the studies with development models.[11,43–71,75,82–99]

Study	Risk of bias, signalling questions					Applicability			Overall judgement		
	Participant selection	Predictors	Outcome	Sample size and participant flow	Analysis	Participant selection	Predictors	Outcome	Risk of bias	Applicability	Usability
Angst[56]	L	L	L	H	H	L	H	L	H	H	Y
Åsenlöf[68]	L	H	L	H	L	L	L	L	H	L	N
Atherton[82]	L	L	L	H	H	L	L	L	H	L	N
Baltov[69]	U	L	U	H	H	L	L	L	H	L	N
Bohman[61]	L	L	L	H	U	L	L	L	H	L	N
Buitenhuis[83]	L	L	U	H	H	L	L	L	H	L	Y
Bunketorp[84]	L	H	L	H	H	L	L	L	H	L	N
Cai[47]	L	U	H	H	H	L	L	L	H	L	Y
Carstensen[64]	L	L	L	H	L	L	L	L	H	L	N
Cecchi[85]	L	L	1U/1L	H	H	L	L	L	H	L	N
Chiarotto[57]	L	L	U	H	H	L	L	L	H	L	Y
Cleland[11]	L	L	L	H	H	L	L	L	H	L	Y
Cleland[86]	L	L	L	H	H	L	L	L	H	L	Y
Cobo[43]	L	L	L	H	H	L	L	L	H	L	Y
Dagfinrud[70]	L	U	L	H	H	L	L	L	H	L	N
Gun[87]	L	L	L	H	H	L	L	L	H	L	N
Hanney[48]	L	L	L	H	H	L	L	L	H	L	Y
Hartling[88]	L	L	L	H	H	L	L	L	H	L	Y
Hendriks[63]	L	L	L	H	H	L	L	L	H	L	N
Hill[49]	L	L	L	H	H	L	L	L	H	L	N
Hoving[58]	L	L	L	H	H	L	L	L	H	L	N
Keating[75, a]	U	L	L	H	H	U	L	L	H	U	N
Kjelmann[89]	L	L	L	H	H	L	L	L	H	L	N
Kyhlbäck[44]	L	U	L	H	H	L	L	L	H	L	N
Landers[50]	L	U	L	H	H	L	L	L	H	L	Y
Lankester[45]	U	U	U	H	H	L	L	U	H	U	N
Michaelson[90]	L	L	L	H	H	L	L	L	H	L	N
Nederhand[65]	L	L	L	H	H	L	L	L	H	L	Y
Nee[66]	L	L	L	H	U	L	L	L	H	L	Y
Nieto[91]	L	L	U	H	H	L	L	L	H	L	N
Pape[92]	H	L	L	H	H	U	L	L	H	L	N
Peterson[67]	U	L	H	H	H	L	2U/2L	L	H	H/L	N
Pool[93]	L	L	L	H	H	L	L	L	H	L	N
Puentedura[51]	L	L	L	H	H	L	L	L	H	L	Y
Radanov[52, b]	U	U	H	H	H	U	U	U	H	U	Y
Raney[94]	L	L	L	H	H	L	L	L	H	L	Y
Rebbeck[95]	L	L	U	H	H	L	L	L	H	L	N
Ritchie[96]	L	L	L	H	H	L	L	L	H	L	Y
Rubinstein[59]	L	L	L	H	H	L	L	L	H	L	N
Schellingerhout[53, b]	L	L	L	H	L	L	L	L	H	L	Y
Saavedra-Hernández[54]	L	L	L	H	H	L	L	L	H	L	Y
Sterling[62]	L	L	L	H	U	L	L	L	H	U	1Y/2N
Sterling[71]	L	L	L	H	H	L	L	L	H	U	1Y/2N
Sterner[46]	L	U	U	U	H	L	L	L	H	L	N
Sturzenegger[97]	L	L	U	H	H	L	L	L	H	L	N
Tseng[55]	L	L	L	H	H	L	L	L	H	L	Y
Vos[60]	L	L	L	H	H	L	L	L	H	L	N
Walton[98]	L	U	L	H	H	L	L	L	H	L	Y
Williamson[99]	L	L	L	H	H	L	L	L	H	L	Y

H = high, L = low, N = no, U = unclear, Y = yes.
Green shading = favourable result, Yellow shading = unclear or mixed results, Red shading = unfavourable result.
[a] Type 2b study, intermediate (temporal) validation.
[b] Type 3 study, development and validation using separate data set.

domain, reviewers disagreed mainly due to their interpretation of the impact of predictors that were not excluded from the outcome definition. Agreement between the two reviewers for judgements of the three applicability domains (participant selection, predictors, and outcome) was 74, 90 and 84%, respectively. In two instances, the third reviewer had to make a decision. Cohen's kappa appeared not applicable, due to consistent very low or zero prevalence. All 49 studies had a high risk of bias and every study

Table 2
Risk of bias, applicability, and model usability among the studies with validation models.[42,52,53,72–75]

Study	Risk of bias, signalling questions					Applicability			Overall judgement		
	Participant selection	Predictors	Outcome	Sample size and participant flow	Analysis	Participant selection	Predictors	Outcome	Risk of bias	Applicability	Usability
Cleland[72, c]	L	L	L	U	U	L	L	L	H	L	Y
Fritz[73, c]	L	L	L	U	H	L	L	L	H	L	Y
Keating[75, a]	U	L	L	U	H	U	L	L	H	U	N
Radanov[52, b]	U	U	H	H	H	U	U	U	H	U	Y
Ritchie[74, c]	L	L	L	L	L	L	L	L	L	L	Y
Schellingerhout[53, b]	L	L	L	L	L	L	L	L	L	L	Y
Sterling[42, c/d]	L	L	L	L	L	L	L	L	L	U	Y

H = high, L = low, N = no, U = unclear, Y = yes.
Green shading = favourable result, Yellow shading = unclear or mixed results, Red shading = unfavourable result.
[a] Type 2b study, intermediate (temporal) validation.
[b] Type 3 study, development and validation using separate data set.
[c] Type 4 study, validation only.
[d] Contains two regression models developed in validation study.

had a high risk of bias in the sample size and participant flow domain, while 43 were biased in the analysis domain. In 42 studies, models were judged to have low concerns regarding their applicability. Four studies contained 11 models with a reasonable number of outcome events according to the definition based on events per variable or participants per predictor.[43–46] All enrolled participants were included in the analysis in nine studies for 12 models.[47–55] Missing data were handled appropriately in seven studies for 17 models.[51,53,56–60] Two derivation studies performed internal validation.[61,62]

The model's overall performance was described in 34 studies by some form of R-squared statistic (R^2). In 11 studies, calibration and/or discrimination measures were described for 19 models. Two studies checked internal validity by cross validation bootstrapping; one of them computed a shrinkage factor.[61,62] Some form of treatment was performed in 29 studies, of which eight described that participants received a specific therapy, like manual therapy, a multi-modal program, standardised physiotherapy, or neural tissue management.

Performance

Seven models reported discriminative ability (AUC or c-statistic) ranging from 0.66 to 0.93.[53,61,63–67] The number of events per variable was >5 in two of these studies,[53,64] one of which was subsequently validated and upheld its model performance.[53] Ten studies presented 15 models with an R^2 or adjusted $R^2 \geq 0.5$.[11,50,51,56,62,63,68–71] For two of these models, external validation studies were subsequently performed,[42,72] one of which concluded that the model could not be validated.[72]

Validation studies

Among the validation studies, the sample size ranged from 16 to 315 and the mean age of participants ranged from 32 to 49 years. Outcomes were measured between 1 week and 12 months, mostly with the Neck Disability Index (NDI) scale or Global Rating of Change.

One study concerned an insurance company population[52] and six studies concerned populations from physiotherapy care, four of which combined a physiotherapy setting with other settings.[42,53,73,74] In two studies, models were tested in a different country than the derivation study.[42,53] Four studies contained models on neck pain,[53,72,73,75] while three studies concerned models for WAD.[42,52,74] Two studies reported that the models could not be validated,[72,73] and one study reported no improvement based on positive predictive value and only weak improvement based on negative predictive value.[75] Two studies reported support for their models based on model performance measures.[53,74] One study reported support based on percentage correct predictions only, and did not give any model performance measures.[52] One WAD study concluded that the model was not accurate because it overestimated the NDI score, and reported discriminative ability if the outcome was dichotomised.[42]

Quality

Agreement between the two reviewers for judgements of the five risk of bias domains (participant selection, predictors, outcome, sample size and participant flow, and analysis) was 57, 86, 57, 71 and 85%, respectively. In the participant selection and outcome domains, reviewers disagreed mainly due to their interpretation of the impact of selection criteria and predictors that were not excluded from the outcome definition. Agreement between the two reviewers for the three applicability domains (participant selection, predictors, and outcome) was 86, 71 and 100%, respectively. Cohen's kappa appeared not to be applicable, due to consistent very low or zero prevalence. Four studies had an overall high risk of bias in one or more domains;[52,72,73,75] among these studies, two models were judged as having unclear concerns regarding applicability[52,75] and one was judged as not usable.[75] One study performed type 2b non-random split validation.[75] High risk of bias was consistent in the analysis domain, mostly due to dichotomised variables and lack of information. Three studies with a low risk of bias generated usable models.[42,53,74] Two of these models were judged to have low concerns regarding their applicability.[53,74]

Performance

In the three validation studies with a low risk of bias overall, one model was intended for use in people with non-specific neck pain,[53] while two concerned people with WAD.[42,74] Discriminative ability of the non-specific neck pain model was AUC 0.65 (95% CI 0.59 to 0.71) and that of the corresponding score chart was 0.66 (95% CI 0.59 to 0.72).[53] Applicability concerns were low and the score chart was clinically usable. Discriminative ability of the first WAD model was AUC 0.85 (95% CI 0.79 to 0.91), and for calibration the study reported an overestimation of the NDI outcome.[42] This

study did not recalibrate the validated model but used its predictors for developing a new model, presenting AUC 0.89 (95% CI 0.84 to 0.94) and 0.91 (95% CI 0.86 to 0.95) if adjusted for study site. The second WAD study tested a prediction rule for two of its three recovery pathways, one moderate to severe path with outcome of NDI $\geq$ 30%, and one full recovery path with outcome NDI $\leq$ 10%.[74] For the path of NDI $\geq$ 30%, specificity was 99% (95% CI 93 to 100) and sensitivity was 44% (95% CI 23 to 65). For the path of NDI $\leq$ 10%, specificity was 86% (95% CI 73 to 94) and sensitivity was 55% (95% CI 41 to 69). Applicability concerns were low and the model was clinically usable.

Consistent prognostic factors in these three models were age, and initial NDI score for WAD. Age lost its significance initially during a low risk of bias derivation study but it regained significance after adjusting for research site.[42]

Discussion

This systematic review included 53 studies of 99 derivation models and seven models tested for validation for prediction of recovery in people with neck pain. Two WAD models and one non-specific neck pain model were found to be promising for use in primary care settings.

These findings are in line with previous systematic reviews on prognostic models for neck pain recovery. One review included six studies and concluded that most models were in the developmental stage, often with moderate study quality.[20] Another review on clinical prediction rules included 18 studies with four models at derivation stage and no neck pain models appearing validated.[21] A second review on clinical prediction rules concluded that two out of the three neck pain studies met their quality criteria. However, quality criteria for prognostic studies were used instead of ones specifically developed for prognostic models.[22] The most recent review on clinical prediction rules for prognosis and treatment prescription in neck pain found that 11 out of 15 clinical prediction rules were at the initial stage of development and seven models had undergone validation.[23] All previous reviews concluded that the methodological quality of the original studies was generally low and few models had undergone validation. Therefore, broad routine clinical use was not recommended yet, which was a conclusion shared with other reviews within the spinal musculoskeletal field.[20,76,77]

Evaluating the studies with up-to-date criteria using the PROBAST tool, a large number of derivation studies with high risk of bias was found, especially in the analysis and sample size/participant flow domains. Studies with a high risk of bias may find inflated discriminative performance. Reporting and methodological standards were often not met, for instance, with respect to reporting of missing data and model performance measures (eg, calibration, discrimination), appropriate handling of missing data (eg, multiple imputation), or correction for overfitting (eg, bootstrapping, shrinkage). Overfitting is one of the biggest concerns and occurs when too many predictors are included in the analysis, especially in small data sets resulting in derived models fitting the data too closely.[78] In that case, the model could obtain idiosyncratic features that are specific to the derivation data itself, resulting in a model that predicts accurately in a derivation sample but performs poorly when applied to other individuals.[38] Too many predictors and categorical variables were often selected in derivation studies and the sample size became very low, resulting in high risk for overfitting. Few studies corrected for overfitting using techniques such as bootstrapping and shrinkage. To reduce overfitting, it is recommended that future researchers collect more data, if possible, select predictors based on former knowledge, and use bootstrapping and shrinkage techniques.[78]

This is the first study that systematically evaluated multivariable prognostic models for recovery of people of any age presenting in primary care with neck pain, using a tool specifically designed for assessing the risk of bias and applicability of prognostic model studies. Using PROBAST – instead of tools not specifically designed for assessing prognostic model studies – facilitates evaluating items specific for prognostic models such as overfitting, data complexities, and a model's performance. However, PROBAST does not provide a guideline for scoring of items as yet and we had to construct our own. For example, we interpreted the signalling question on reasonable number of outcome events on the 'rule of thumb' of 10 events per variable; this was rigorous because it was based on degrees of freedom used. A less rigorous interpretation would probably result in the review spuriously concluding lower risk of bias for the derivation studies.

Another limitation was that WAD studies were included with populations that included primary care patients, people recruited from hospital emergency departments and recruited via general advertisements. It might be possible that predictors for recovery differ between patients in primary care versus emergency departments, or the general population. Another potential limitation could have been publication bias. Although a large number of studies without language restriction were included, no non-English studies were obtained, which may have potentially yielded more negative results.

The vast majority of the models cannot be used in a clinical situation yet, because their derivation studies had high risk of bias and validation was not executed or unsuccessful. Nevertheless, this review found three validated models that are considered to be promising and may provide support for clinicians in their decision-making process.

The Ritchie two-way WAD model predicted full recovery by NDI $\leq$ 32% and age $\leq$ 25 years, and ongoing moderate/severe disability by NDI $\geq$ 40%, age $\geq$ 35 years, and hyperarousal (Post-traumatic Diagnostic Scale subscale $\geq$ 6).[74] The Sterling WAD model predicted disability by initial NDI, age, left rotation range of motion, cold pain threshold, Impact of Events Scale, and blood flow (Quotient of Integrals).[42] The Schellingerhout non-specific neck pain model predicted recovery by age, pain intensity, headache, radiation to elbow/shoulder, previous neck complaints, low back pain, employment status, and quality of life (EuroQOL).[53]

Baseline disability appeared to be a consistent prognostic factor in WAD and could support treatment decision-making because disability can effectively be reduced by primary care interventions in WAD and neck pain.[79,80]

Rather than development of new models, (further) validating, adjusting, or updating of existing (high-quality) models is advocated.[19,81] For the three promising models, further validation and evaluation of clinical impact is advised before their broad clinical use can be advocated. The neck pain model showed a small pre-test to post-test probability shift, and testing the model or its chart in a comparable setting with other prevalence rates is recommended. Further, testing the performance of the two WAD models in a primary care setting alone is required.

What is already known on this topic: Improvements in pain and disability typically occur in the first weeks after the onset of an episode of neck pain, but residual pain and disability beyond this time are often of substantial severity and persist for at least 1 year. Subgrouping of people with neck pain based on their prognosis may enhance treatment outcomes by enabling tailored treatment and management strategies.
What this study adds: Although many models have been developed and investigated for their ability to predict recovery of people with neck pain, few are suitable to use. However, two models for whiplash-asociated disorders and one model for non-specific neck pain were found to be suitable for use in primary care settings.

eAddenda: Appendices 1,2 and 3 can be found online at: https://doi.org/10.1016/j.jphys.2017.11.013.

Ethics approval: Not required.

Competing interests: The authors declare that there are no competing interests.

Source of support: This study was partly funded by a program grant of the Dutch Arthritis Foundation and SOMT University of Physiotherapy.

Acknowledgements: Nil.

Provenance: Not invited. Peer reviewed.

Correspondence: Roel Wingbermühle, SOMT University of Physiotherapy, Amersfoort, The Netherlands. Email: r.wingbermuhle@somtuniversity.nl

References

1. Hoy D, March L, Woolf A, Blyth F, Brooks P, Smith E, et al. The global burden of neck pain: Estimates from the global burden of disease 2010 study. *Ann Rheum Dis.* 2014;73:1309–1315.
2. Vos T, Allen C, Arora M, Bhutta Z, Brown A, Carter A, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet.* 2016;388:1545–1602.
3. Hush JM, Lin CC, Michaleff ZA, Verhagen A, Refshauge KM. Prognosis of acute idiopathic neck pain is poor: a systematic review and meta-analysis. *Arch Phys Med Rehabil.* 2011;92:824–829.
4. Walton DM, Carroll LJ, Kasch H, Sterling M, Verhagen AP, MacDermid JC, et al. An overview of systematic reviews on prognostic factors in neck pain: Results from the International Collaboration on Neck Pain (ICON) Project. *Open Orthop J.* 2013;7:494–505.
5. van der Velde G, Yu H, Paulden M, Côté P, Varatharajan S, Shearer HM, et al. Which interventions are cost-effective for the management of whiplash-associated and neck pain-associated disorders? A systematic review of the health economic literature by the Ontario Protocol for Traffic Injury Management (OPTIMa) Collaboration. *Spine J.* 2016;16:1582–1597.
6. Vincent K, Maigne J-Y, Fischhoff C, Lanlo O, Dagenais S. Systematic review of manual therapies for nonspecific neck pain. *Joint Bone Spine.* 2013;80:508–515.
7. Gross A, Kay T, Paquin J, Blanchette S, Lalonde P, Christie T, et al. Exercises for mechanical neck disorders (Review). *Cochrane Database Syst Rev.* 2015;(1).
8. Hurwitz EL, Carragee EJ, van der Velde G, Carroll LJ, Nordin M, Guzman J, et al. Treatment of Neck Pain: Noninvasive Interventions. Results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *J Manipulative Physiol Ther.* 2009;32:S141–S175.
9. Fritz JM, Brennan GP. Preliminary examination of a proposed treatment-based classification system for patients receiving physical therapy interventions for neck pain. *Phys Ther.* 2007;87:513–524.
10. Carroll LJ, Hogg-Johnson S, van der Velde G, Haldeman S, Holm LW, Carragee EJ, et al. Course and prognostic factors for neck pain in the general population. *Eur Spine J.* 2008;17:75–82.
11. 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. *Phys Ther.* 2007;87:9–23.
12. Carroll LJ, Hurwitz EL, Côté P, Hogg-Johnson S, Carragee EJ, Nordin M, et al. Research priorities and methodological implications: the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *J Manipulative Physiol Ther.* 2009;32:244–251.
13. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ.* 2013;346:e5595–e5595.
14. Moons KGM, Biesheuvel CJ, Grobbee DE. Test Research versus Diagnostic Research. *Clin Chem.* 2004;50:473–476.
15. Moons KGM, Altman DG, Reitsma JB, Collins GS. New guideline for the reporting of studies developing, validating, or updating a prediction model. *Clin Chem.* 2015;61:565–566.
16. van Trijffel E, Lindeboom R, Bossuyt PM, Schmitt MA, Lucas C, Koes BW, et al. Indicating spinal joint mobilisations or manipulations in patients with neck or low-back pain: protocol of an inter-examiner reliability study among manual therapists. *Chiropr Man Therap.* 2014;22:22.
17. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ.* 2009;338:b375–b375.
18. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med.* 2015;162:55.
19. Steyerberg E, Moons KGM, van der Windt D, Hayden JA, Perel P, Schroter S, et al. Prognosis research strategy (PROGRESS) series 3: prognostic models. *PLoS Med.* 2013;10:e1001381.
20. van Oort L, van den Berg T, Koes BW, de Vet RHCW, Anema HJR, Meymans MW, et al. Preliminary state of development of prediction models for primary care physical therapy: A systematic review. *J Clin Epidemiol.* 2012;65:1257–1266.
21. Stanton TR. Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. *Am Phys Ther Assoc.* 2010;90:177–201.
22. Beneciuk JM, Bishop MD, George SZ. Clinical prediction rules for physical therapy interventions: a systematic review. *Am Phys Ther Assoc.* 2009;89:114–124.
23. Kelly J, Ritchie C, Sterling M. Clinical prediction rules for prognosis and treatment prescription in neck pain: A systematic review. *Musculoskelet Sci Pract.* 2017;27:155–164.
24. Ensor J, Riley RD, Moore D, Snell KIE, Bayliss S, Fitzmaurice D. Systematic review of prognostic models for recurrent venous thromboembolism (VTE) post-treatment of first unprovoked VTE. *BMJ Open.* 2016;6:e011190.
25. Braun C, Hanchard NC, Batterham AM, Handoll HH, Betthauser A. Prognostic models in adults undergoing physical therapy for rotator cuff disorders: systematic review. *Am Phys Ther Assoc.* 2016;96:961–971.
26. Halligan S, Boone D, Bhatnagar G, Ahmad T, Bloom S, Rodriguez-Justo M, et al. Prognostic biomarkers to identify patients destined to develop severe Crohn's disease who may benefit from early biological therapy: protocol for a systematic review, meta-analysis and external validation. *Syst Rev.* 2016;5:206.
27. Ingui BJ, Rogers MAM. Searching for clinical prediction rules in MEDLINE. *J Am Med Informatics Assoc.* 2001;8:391–397.
28. Pillastrini P, Vanti C, Curti S, Mattioli S, Ferrari S, Violante FS, et al. Using PubMed search strings for efficient retrieval of manual therapy research literature. *J Manipulative Physiol Ther.* 2015;38:159–166.
29. Kwon Y, Lemieux M, McTavish J, Wathen N. Identifying and removing duplicate records from systematic review searches. *J Med Libr Assoc.* 2015;103:184–188.
30. Clark GM. Prognostic factors versus predictive factors: Examples from a clinical trial of erlotinib. *Mol Oncol.* 2008;1:406–412.
31. Hidalgo B, Goodman M. Multivariate or multivariable regression? *Am J Public Health.* 2013;103:39–40.
32. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.* New York: Springer Science and Business Media; 2009.
33. Guzman J, Hurwitz EL, Carroll LJ, Haldeman S, Côté P, Carragee EJ, et al. A New Conceptual Model of Neck Pain. *Eur Spine J.* 2008;17:14–23.
34. Guzman J, Hurwitz EL, Carroll LJ, Haldeman S, Côté P, Carragee EJ, et al. A new conceptual model of neck pain: linking onset, course, and care: the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine.* 2008;32:S14–S23.
35. Hoving JL, de Vet HCW, Koes BW, van der Windt DA, Assendelft WJ, van Mameren H, et al. Manual therapy, physical therapy, or continued care by the general practitioner for patients with neck pain: long-term results from a pragmatic randomized clinical trial. *Clin J Pain.* 2006;22:370–377.
36. Wolff R, Whiting P, Mallet S, Riley R, Westwood M, Kleijnen K, et al. PROBAST: a risk of bias tool for prediction modelling studies | The 23rd Cochrane Colloquium. http://2015.colloquium.cochrane.org/abstracts/probast-risk-bias-tool-prediction-modelling-studies. Published 2015 [accessed 28/07/2016].
37. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat Methods Med Res.* 2017;26:796–808.
38. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS Checklist. *PLoS Med.* 2014;11:e1001744.
39. Moons KGM, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med.* 2015;162:W1–W73.
40. Austin PC, Steyerberg EW. Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable. *BMC Med Res Methodol.* 2012;12:82.
41. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression.* 3rd ed. Hoboken, NJ: Wiley; 2013.
42. Sterling M, Hendrikz J, Kenardy J, Kristjansson E, Dumas JP, Niere K, et al. Assessment and validation of prognostic models for poor functional recovery 12 months after whiplash injury: A multicentre inception cohort study. *Pain.* 2012;153:1727–1734.
43. Cobo EP, Mesquida MEP, Fanegas EP, Atanasio EM, Pastor MBS, Pont CP, et al. What factors have influence on persistence of neck pain after a whiplash? *Spine.* 2010;35:E338–E343.
44. Kyhlbäck M, Thierfelder T, Söderlund A. Prognostic factors in whiplash-associated disorders. *Int J Rehabil Res.* 2002;187:181–187.
45. Lankester BJA, Garneti N, Gargan MF, Bannister GC. Factors predicting outcome after whiplash injury in subjects pursuing litigation. *Eur Spine J.* 2006;15:902–907.
46. Sterner Y, Toolanen G, Gerdle B, Hildingsson C. The incidence of whiplash trauma and the effects of different factors on recovery. *J Spinal Disord.* 2003;16:195–199.
47. 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. *Eur Spine J.* 2011;20:912–922.
48. 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. *Int J Sports Phys Ther.* 2013;8:756–776.
49. Hill JC, Lewis M, Sim J, Hay EM, Dziedzic K. Predictors of poor outcome in patients with neck pain treated by physical therapy. *Clin J Pain.* 2007;23:683–690.
50. Landers MR, Creger RV, Baker CV, Stutelberg KS, Landers M, Creger R, et al. The use of fear-avoidance beliefs and nonorganic signs in predicting prolonged disability in patients with neck pain. *Man Ther.* 2008;13:239–248.
51. 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. *J Orthop Sports Phys Ther.* 2012;42:577–592.
52. Radanov BP, Sturzenegger M. Predicting recovery from common whiplash. *Eur Neurol.* 1996;36:48–51.
53. Schellingerhout JM, Heymans MW, Verhagen AP, Lewis M, de Vet HCW, Koes BW. Prognosis of patients with nonspecific neck pain: development and external validation of a prediction rule for persistence of complaints. *Spine.* 2010;35:E827–E835.
54. Saavedra-Hernández M, Castro-Sánchez AM, Fernández-De-Las-Peñas 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. *J Manipulative Physiol Ther.* 2011;34:144–152.

55. Tseng Y, Wang WTJ, Chen W-Y, Hou T-J, Chen T-C, Lieu F-K. Predictors for the immediate responders to cervical manipulation in patients with neck pain. *Man Ther.* 2006;11:306–315.
56. Angst F, Gantenbein AR, Lehmann S, Gysi-Klaus F, Aeschlimann A, Michel BA, et al. Multidimensional associative factors for improvement in pain, function, and working capacity after rehabilitation of whiplash associated disorder: a prognostic, prospective outcome study. *BMC Musculoskelet Disord.* 2014;15:130.
57. Chiarotto A, Fortunato S, Falla D. Predictors of outcome following a short multimodal rehabilitation program for patients with whiplash associated disorders. *Eur J Phys Rehabil Med.* 2015;51:133–141.
58. Hoving JL, de Vet HCW, Twisk JWR, Deville WLJM, van der Windt DAWM, Koes BW. Prognostic factors for neck pain in general practice. *Pain.* 2004;110:639–645.
59. Rubinstein SM, Knol DL, Leboeuf-Yde C, de Koekkoek TE, Pfeifle CE, van Tulder MW. Predictors of a favorable outcome in patients treated by chiropractors for neck pain. *Spine.* 2008;33:1451–1458.
60. Vos CJ, Verhagen AP, Passchier J, Koes BW. Clinical course and prognostic factors in acute neck pain: an inception cohort study in general practice. *Pain Med.* 2008;9:572–580.
61. Bohman T, Cote P, Boyle E, Cassidy JD, Carroll LJ, Skillgate E. Prognosis of patients with whiplash-associated disorders consulting physiotherapy: development of a predictive model for recovery. *BMC Musculoskelet Disord.* 2012;13:264.
62. Sterling M, Jull G, Vicenzino B, Kenardy J, Darnell R. Physical and psychological factors predict outcome following whiplash injury. *Pain.* 2005;114:141–148.
63. Hendriks EJM, Scholten-Peeters GGM, Van Der Windt DAWM, Neeleman-Van Der Steen CWM, Oostendorp RAB, Verhagen AP. Prognostic factors for poor recovery in acute whiplash patients. *Pain.* 2005;114:408–416.
64. Carstensen TBW, Fink P, Oernboel E, Kasch H, Jensen TS, Frostholm L. Sick leave within 5 years of whiplash trauma predicts recovery: a prospective cohort and register-based study. *PLoS One.* 2015;10:e0130298.
65. 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. *Arch Phys Med Rehabil.* 2004;85:496–501.
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. *J Orthop Sports Phys Ther.* 2013;43:379–391.
67. Peterson C, Bolton J, Humphreys BK. Predictors of outcome in neck pain patients undergoing chiropractic care: comparison of acute and chronic patients. *Chiropr Man Therap.* 2012;20:27.
68. Åsenlöf P, Bring A, Söderlund A. The clinical course over the first year of Whiplash Associated Disorders (WAD): Pain-related disability predicts outcome in a mildly affected sample. *BMC Musculoskelet Disord.* 2013;14.
69. Baltov P, Cote J, Truchon M, Feldman DE. Psychosocial and socio-demographic factors associated with outcomes for patients undergoing rehabilitation for chronic whiplash associated disorders: a pilot study. *Disabil Rehabil.* 2008;30:1947–1955.
70. Dagfinrud H, Storheim K, Magnussen LH, Ødegaard T, Hoftaniska I, Larsen LG, et al. The predictive validity of the Orebro Musculoskeletal Pain Questionnaire and the clinicians' prognostic assessment following manual therapy treatment of patients with LBP and neck pain. *Man Ther.* 2013;18:124–129.
71. Sterling M, Jull G, Kenardy J. Physical and psychological factors maintain long-term predictive capacity post-whiplash injury. *Pain.* 2006;122:102–108.
72. Cleland JA, Mintken PE, Carpenter K, Fritz JM, Glynn P, Whitman J, 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. *Phys Ther.* 2010;90:1239–1250.
73. 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. *J Orthop Sports Phys Ther.* 2014;44:45–57.
74. 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. *J Orthop Sports Phys Ther.* 2015;45:242–250.
75. 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. *J Whiplash Relat Disord.* 2005;4:43–55.
76. Haskins R, Osmotherly PG, Rivett DA. Validation and impact analysis of prognostic clinical prediction rules for low back pain is needed: a systematic review. *J Clin Epidemiol.* 2015;68:821–832.
77. 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;38:762–769.
78. Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med.* 2004;66:411–421.
79. Gross AR, Paquin JP, Dupont G, Blanchette S, Lalonde P, Cristie T, et al. Exercises for mechanical neck disorders: A Cochrane review update. *Man Ther.* 2016;24:25–45.
80. Sutton D, Cote P, Wong JJ, Varatharajan S, Randhawa K, Yu H, et al. Is multimodal care effective for the management of patients with whiplash-associated disorders or neck pain and associated disorders? A systematic review by the Ontario Protocol for Traffic Injury Management (OPTIMa) Collaboration. *Spine J.* 2014;34–61.
81. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart.* 2012;98:691–698.
82. Atherton K, Wiles NJ, Lecky FE, Hawes SJ, Silman AJ, Macfarlane GJ, et al. Predictors of persistent neck pain after whiplash injury. *Emerg Med J.* 2006;23:195–201.
83. Buitenhuis J, de Jong PJ, Jaspers JPC, Groothoff JW. Relationship between posttraumatic stress disorder symptoms and the course of whiplash complaints. *J Psychosom Res.* 2006;61:681–689.
84. Bunketorp L, Lindh M, Carlsson J, Stener-Victorin E. The perception of pain and pain-related cognitions in subacute whiplash-associated disorders: Its influence on prolonged disability. *Disabil Rehabil.* 2006;28:271–279.
85. Cecchi F, Molino-Lova R, Paperini A, Boni R, Castagnoli C, Gentile J, et al. Predictors of short- and long-term outcome in patients with chronic non-specific neck pain undergoing an exercise-based rehabilitation program: a prospective cohort study with 1-year follow-up. *Intern Emerg Med.* 2011;6:413–421.
86. Cleland JA, Fritz JM, Whitman JM, Heath R. Predictors of short-term outcome in people with a clinical diagnosis of cervical radiculopathy. *Phys Ther.* 2007;87:1619–1632.
87. Gun RT, Osti OL, O'Riordan A, Mpelasoka F, Eckerwall CGM, Smyth JF. Risk factors for prolonged disability after whiplash injury: a prospective study. *Spine.* 2005;30:386–391.
88. Hartling L, Pickett W, Brison RJ. Derivation of a clinical decision rule for whiplash associated disorders among individuals involved in rear-end collisions. *Accid Anal Prev.* 2002;34:531–539.
89. Kyhllman G, Skargren E, Oberg B. Prognostic factors for perceived pain and function at one-year follow-up in primary care patients with neck pain. *Disabil Rehabil.* 2002;24:364–370.
90. Michaelson P, Sjolander P, Johansson H. Factors predicting pain reduction in chronic back and neck pain after multimodal treatment. *Clin J Pain.* 2004;20:447–454.
91. Nieto R, Miro J, Huguet A. Pain-related fear of movement and catastrophizing in whiplash-associated disorders. *Rehabil Psychol.* 2013;58:361–368.
92. Pape E, Brox JI, Hagen KB, Natvig B, Schirmer H. Prognostic factors for chronic neck pain in persons with minor or moderate injuries in traffic accidents. *Accid Anal Prev.* 2007;39:135–146.
93. Pool JJM, Ostelo RWJG, Knol D, Bouter LM, de Vet HCW. Are psychological factors prognostic indicators of outcome in patients with sub-acute neck pain? *Man Ther.* 2010;15:111–116.
94. Raney NH, Petersen EJ, Smith TA, Cowan JE, Rendeiro DG, Deyle GD, et al. Development of a clinical prediction rule to identify patients with neck pain likely to benefit from cervical traction and exercise. *Eur Spine J.* 2009;18:382–391.
95. Rebbeck T, Sindhusake D, Cameron ID, Rubin G, Feyer AM, Walsh J, et al. A prospective cohort study of health outcomes following whiplash associated disorders in an Australian population. *Inj Prev.* 2006;12:93–98.
96. 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;154:2198–2206.
97. Sturzenegger M, Radanov BP, Di Stefano G. The effect of accident mechanisms and initial findings on the long-term course of whiplash injury. *J Neurol.* 1995;242:443–449.
98. Walton DM, Macdermid JC, Nielson W, Teasell RW, Reese H, Levesque L. Pressure pain threshold testing demonstrates predictive ability in people with acute whiplash. *J Orthop Sports Phys Ther.* 2011;41:658–665.
99. 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;101:34–43.

Websites

PROBAST www.systematic-reviews.com/probast