

# BMJ Open Development and validation of a screening tool to predict the risk of chronic low back pain in patients presenting with acute low back pain: a study protocol

Adrian Traeger,<sup>1,2</sup> Nicholas Henschke,<sup>3</sup> Markus Hübscher,<sup>1,2</sup> Christopher M Williams,<sup>4,5</sup> Steven J Kamper,<sup>5</sup> Chris G Maher,<sup>5</sup> G Lorimer Moseley,<sup>2,6</sup> James H McAuley<sup>1,2</sup>

**To cite:** Traeger A, Henschke N, Hübscher M, *et al*. Development and validation of a screening tool to predict the risk of chronic low back pain in patients presenting with acute low back pain: a study protocol. *BMJ Open* 2015;**5**:e007916. doi:10.1136/bmjopen-2015-007916

► Prepublication history and additional material for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2015-007916>).

Received 10 February 2015  
Revised 28 May 2015  
Accepted 12 June 2015



CrossMark

For numbered affiliations see end of article.

## Correspondence to

Dr James McAuley;  
j.mcauley@neura.edu.au

## ABSTRACT

**Introduction:** Around 40% of people presenting to primary care with an episode of acute low back pain develop chronic low back pain. In order to reduce the risk of developing chronic low back pain, effective secondary prevention strategies are needed. Early identification of at-risk patients allows clinicians to make informed decisions based on prognostic profile, and researchers to select appropriate participants for secondary prevention trials. The aim of this study is to develop and validate a prognostic screening tool that identifies patients with acute low back pain in primary care who are at risk of developing chronic low back pain. This paper describes the methods and analysis plan for the development and validation of the tool.

**Methods/analysis:** The prognostic screening tool will be developed using methods recommended by the Prognosis Research Strategy (PROGRESS) Group and reported using the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement. In the development stage, we will use data from 1248 patients recruited for a prospective cohort study of acute low back pain in primary care. We will construct 3 logistic regression models to predict chronic low back pain according to 3 definitions: any pain, high pain and disability at 3 months. In the validation stage, we will use data from a separate sample of 1643 patients with acute low back pain to assess the performance of each prognostic model. We will produce validation plots showing Nagelkerke R<sup>2</sup> and Brier score (overall performance), area under the curve statistic (discrimination) and the calibration slope and intercept (calibration).

**Ethics and dissemination:** Ethical approval from the University of Sydney Ethics Committee was obtained for both of the original studies that we plan to analyse using the methods outlined in this protocol (Henschke *et al*, ref 11-2002/3/3144; Williams *et al*, ref 11638).

## INTRODUCTION

Acute non-specific low back pain (LBP) is widely reported to have a favourable

## Strengths and limitations of this study

- First prognostic tool in low back pain to follow, a priori, the Prognosis Research Strategy (PROGRESS) framework, and Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) reporting guidelines for prognostic research.
- Prespecifies statistical analysis plan and informative levels of tool performance to increase transparency of results and of the final report.
- Restricted to only the use of predictor variables measured in previous data sets.
- Minor differences in the way variables are measured between the development and the validation data sets.

prognosis;<sup>1</sup> pain intensity reduces rapidly in the first few weeks,<sup>2</sup> and around 60% have fully recovered (return to work, no disability, and no pain) by 3 months.<sup>3</sup> However, for the 40% of patients who continue to report pain at 3 months, or ‘chronic’ LBP, the prognosis is much poorer.<sup>4</sup> Despite this difference in prognosis, many randomised trials apply the same treatment to all patients with acute LBP, which is an inefficient approach. An alternative approach is to target specific interventions to those at higher risk of developing chronic LBP. Research evaluating this ‘stratified’ approach to acute LBP management has been identified as a priority.<sup>5</sup>

Secondary prevention in LBP refers to preventing patients with acute LBP from developing chronic LBP.<sup>6</sup> The first step in secondary prevention is to identify factors that are associated with poor outcome, or prognostic factors.<sup>7</sup> Recently, the Prognosis Research Strategy (PROGRESS) group proposed a framework for conducting prognosis

studies that included standards for identifying prognostic factors.<sup>8</sup> Once prognostic factors have been identified they can be combined to produce prognostic models<sup>9 10</sup> to assist clinicians in management decisions and researchers in trial design. For example, prognostic models applied to patients with LBP have predicted outcome better than chance or clinical judgment alone.<sup>11 12</sup> Models that are used to screen patients into risk groups are known as prognostic screening tools.

The prognostic screening tools that are currently available for LBP are of limited use in secondary prevention.<sup>13</sup> For example, the majority of existing tools were developed in secondary care,<sup>14 15</sup> and in groups of patients with acute LBP and chronic LBP,<sup>16 17</sup> making their routine application in primary care problematic without further testing. In addition, the predictive validity of these tools is invariably poorer when used outside of the sample in which they were developed.<sup>18</sup>

These limitations demonstrate the importance of external validation,<sup>19</sup> which is rarely reported.<sup>20 21</sup> To assess the external validity of a prognostic model, discrimination (the probability that patients who develop the health outcome of interest are allocated higher risk scores) and calibration (how closely predicted outcomes match actual outcomes) should ideally be assessed in a separate sample from that used to initially derive the model. For the few tools that have been externally validated in acute LBP,<sup>22–24</sup> discrimination ranged from poor to moderate, and calibration was only reported in one study.<sup>23</sup>

There is also variability in the outcomes that prognostic tools have been developed to predict. Although these tools should predict reliable and clinically important outcomes,<sup>25</sup> in LBP there is no consensus on which outcomes are the most important.<sup>26</sup> Most published tools predict disability or return to work outcomes, rather than pain, which is the most common outcome used to define chronic LBP.<sup>21 27</sup> Von Korff and Saunders<sup>28</sup> have argued that pain, in particular high pain intensity and/or disability, is the most important outcome to assess at 3 months. Three months is also the time at which a marked change in prognosis occurs.<sup>3</sup> A recently formed research task force<sup>27</sup> agreed on a uniform definition of chronic LBP: 3 months' worth of pain days in the past 6 months. The task force also emphasised the importance of grading the *impact* of chronic pain and disability, though validated cut-offs were not available. The recent progress made to define chronic LBP highlights the need to identify not only the patients at risk of ongoing pain for 3 months, but also the patients at risk of developing high-impact chronic pain and disability. Such patients are logical targets for early secondary prevention strategies. We are not aware of any tool that was developed to predict the onset of chronic pain.

The aim of this protocol is to describe the method and analysis plan for the development and validation of a prognostic screening tool for acute LBP that is suitable for secondary prevention.

## METHODS

Our study design is informed by the PROGRESS framework and specific recommendations for statistical approaches to prognostic research (table 1). We will report the study in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement for prediction studies.<sup>34</sup>

The PROGRESS framework outlines four types of prognostic research: (1) fundamental prognosis research; (2) prognostic factor research; (3) prognostic model research; (4) stratified medicine research. The proposed study is Type 3, prognostic model research (table 1). To develop the prognostic screening tool, we will consider the recommendations of Kamper *et al.*<sup>29</sup> and Royston *et al.*<sup>9</sup> To validate the prognostic screening tool, we will use the method suggested by Altman *et al.*<sup>30</sup> and Steyerberg *et al.*<sup>35</sup>

### Development of the tool

#### Development sample

The development sample data for the proposed study is from a cohort study that assessed the prognosis of acute LBP in primary care. Details of this study have been published elsewhere.<sup>3</sup> In short, consecutive patients were recruited between November 2003 and July 2005 from primary care practices in the Sydney Metropolitan area, Australia. In the Australian healthcare system, 'primary care' includes first contact care provided by general practitioners, allied health practitioners (physiotherapists, chiropractors), and pharmacists. The original study recruited an inception cohort of 1248 patients with acute LBP (<4 weeks' duration), 973 of which had pain of less than 2 weeks' duration. All patients were managed according to the Australian National Health and Medical Research Council guidelines for acute LBP.<sup>36</sup> Outcomes were measured via telephone at 6 weeks, 3 months and 12 months. Key design features of the study are described in table 2 below.

### Measures

#### Outcomes

The primary outcome will be the presence of chronic LBP. Chronic LBP will be defined as having greater than 'very mild' pain 3 months after the initial assessment. In the development sample, this was classified as >2 on a 6-point Likert scale (How much back pain have you had in the past week? 1='none', 2='very mild', 3='mild', 4='moderate', 5='severe', or 6='very severe').<sup>41</sup> The validation sample used a different pain rating scale, so we have defined greater than 'very mild' pain on this numeric rating scale as  $\geq 3/10$ . We chose these cut-offs because a large proportion of patients with 'very mild' pain (1–2/10) pain consider themselves to be recovered.<sup>26</sup>

We will also develop two additional prognostic screening tools that consider the impact of pain and disability when defining chronic LBP.<sup>27</sup> For the tool predicting

**Table 1** Recommendations for prognostic research

<b>PROGRESS type of clinical research</b>	<b>Types 1 and 2<sup>7 29</sup></b>	<b>Type 3<sup>8 9 30–33</sup></b>			<b>Type 4<sup>10</sup></b>
<b>Research theme</b>	<b>Original prognosis research and prognostic factor research</b>	<b>Development of a prognostic model</b>	<b>Validation of a prognostic model</b>	<b>Refinement of a prognostic model for clinical use</b>	<b>Impact study</b>
Current recommendations (PROGRESS recommendations in parentheses)	<ol style="list-style-type: none"> <li>1. Register study and publish a protocol (Recommendation 10)</li> <li>2. Recruit a representative and well-defined sample at a common, early time point (Recommendation 14)</li> <li>3. Ensure complete follow-up of sufficient length</li> <li>4. Choose prognostic factors measured based on sound theory</li> <li>5. Blind outcome assessors</li> <li>6. Account for covariates statistically</li> <li>7. Ensure sample size is large enough to assess multiple prognostic factors (10 outcome-events-per-predictor-rule)</li> <li>8. Validate the model</li> <li>9. Report all results explicitly and transparently (Recommendation 15)</li> </ol>	<ol style="list-style-type: none"> <li>1. Select candidate predictors that are clinically relevant</li> <li>2. Evaluate data quality (Recommendation 20)</li> <li>3. Describe data handling decisions for example, continuous variables should be analysed on their continuous scale (Recommendation 13)</li> <li>4. Select variables to be included in the final model using a prespecified strategy</li> <li>5. Assess the performance of the model (internal validation), ie, overall performance, discrimination and calibration</li> </ol>	<ol style="list-style-type: none"> <li>1. Prespecify acceptable performance of the model</li> <li>2. Assess overall performance, discrimination, and calibration in the validation sample</li> <li>3. Include a validation plot</li> <li>4. Update and recalibrate the model</li> </ol>	<ol style="list-style-type: none"> <li>1. Adjust the tool for clinical use</li> <li>2. Use a simple interface</li> <li>3. Do not refer to the tool as a 'rule'</li> <li>4. Make sure all aspects of the tool are clear and unambiguous</li> <li>5. Include uncertainty interval (95%CI) around posterior probability estimates</li> </ol>	<ol style="list-style-type: none"> <li>1. Take care with underpowered statistical analyses that are not pre-planned</li> <li>2. Report all subgroup findings</li> <li>3. Subgroup analyses should be replicated in new data</li> <li>4. Analyse continuous outcomes on their continuous scale</li> <li>5. Design RCTs to be 4 arm, ie, intervention and control in groups +ve and -ve on rule (Recommendation 22)</li> <li>6. Studies should compare 'stratified' vs 'all-comer' approaches (Recommendation 23)</li> </ol>

PROGRESS, Prognosis Research Strategy.

**Table 2** Prognosis study by Henschke *et al*<sup>8</sup> adherence to PROGRESS recommendations

Recommendation		Notes
Study was registered and protocol published (Recommendation 10)	Y	Protocol published and registered prior to original statistical analysis <sup>57</sup>
Sample was representative and recruited at a common, early time point (Recommendation 14)	Y	Patients were eligible if presenting to primary care with acute non-specific low back pain <sup>37</sup>
Follow-up was complete and of sufficient length	Y	98.4% follow-up at 3 months
Prognostic factors were based on sound theory	Y	6 groups of putative prognostic factors were measured: current history, past LBP history, sociodemographic characteristics, general health, psychological factors and work <sup>38–40</sup>
Outcome assessors were blinded	N	Unlikely to have introduced bias because this study did not originally have a hypothesis regarding prediction
Covariates were accounted for	NA	Not relevant for tool development.
Sample size was large enough to assess multiple prognostic factors (10 outcome-events-per-predictor-rule)	Y	1248 patients were recruited for a total of 21 predictors
Model was validated	NA	See 'Validation of the tool' below
Results were explicitly and transparently reported (Recommendation 15)	Y	Full protocol and results are published and available

LBP, low back pain; N, not achieved; NA, not applicable to the present study; Y, achieved.

high pain, patients will be classified as having 'chronic LBP (high pain)' if they reported 'moderate', 'severe' or 'very severe' pain intensity on a 6-point scale, or  $\geq 5$  on the 11-point scale in the validation sample, at 3 month follow-up.<sup>42</sup> Because there is still no consensus on what constitutes 'high impact' or 'severe' chronic LBP, we have selected our own cut-off in line with Von Korff and Saunders'<sup>28</sup> recommendation of 'moderate' or greater pain intensity at 3 months. For the tool predicting disability, patients will be classified as having 'chronic LBP (disability)' if they reported  $\geq 2/5$  on a 5-point disability scale (During the past week, how much has LBP interfered with your normal work (including both work outside the home and housework)? 1='not at all', 2='a little bit', 3='moderately', 4='quite a bit', 5='extremely')<sup>41</sup> at 3 month follow-up. Once again, because no established cut-offs exist,<sup>27</sup> we selected our own cut-off for what might constitute clinically important disability. We chose a disability score of  $\geq 2/5$  on the 5-point scale used in the development sample,<sup>41</sup> or  $\geq 7/24$  on the 24-point Roland Morris Disability scale used in the validation sample. When both are converted to a 0–10 scale, these values approximate each other ( $7/24 \times 10 = 3/10$  or  $1.5/5$ ).<sup>2</sup> Choosing this disability cut-off will allow comparison to two recently published prognostic screening tools, which also selected a cut-off of 7 on the Roland scale.<sup>16 24</sup>

### Predictors

Candidate predictors will be selected from those measured at baseline in the cohort study if they are: (1) simple and reliable to measure in practice and (2) have a theoretical association with the development of chronic LBP. Candidate predictors are listed in online supplementary appendix A.

### Statistical analysis

Cases with missing values will be removed from the dataset if follow-up rates are higher than 95%. If missing data exceeds 5%, 'single imputation'<sup>43</sup> will be used. As per the recommendation from the PROGRESS group (table 1), we will ensure at least 10 cases per candidate predictor variable, to adequately power the regression analysis.<sup>9 44</sup>

### Variable selection

A logistic regression analysis will be used to investigate the relationship between the prognostic variables at baseline and the measures of chronic LBP. Age, sex and duration of the pain episode will be entered into block 1 of a multivariate analysis to reduce the complexity of the final model for clinical use, and ensure maximum sensitivity of the final models. In block 2, each potential prognostic factor will be added stepwise to the model by using an automated forward selection procedure. We will set a liberal significance level ( $p < 0.10$ ) to select variables that remain in the model. The predicted probability of chronic LBP will be modelled using the logistic regression equation

$$P(\text{chronic LBP}) = 1/[1 + \exp - (a_0 + b_1x_1 + \dots + b_jx_j)]$$

Continuous predictor variables will be treated as linear in the first multivariable regression model. After the initial predicted probabilities of chronic LBP are calculated, the linearity of continuous predictor variables will be examined with the predicted probability of chronic LBP as the dependent variable using scatter plots and the Box-Tidwell transformation.<sup>45</sup> The Box-Tidwell transformation involves log transforming each continuous predictor variable, producing an interaction term between the original variable and its log (eg, pain $\times$ LN), and including this term in the regression analysis. If the

interaction term is significant, this indicates significant non-linearity. Continuous predictor variables that demonstrate a non-linear relationship with the dependent variable (predicted probability of chronic LBP) will be transformed by using fractional polynomial procedure.<sup>46</sup> Regression analyses will be performed using SPSS.<sup>46</sup>

### Scoring

Each individual will be allocated a risk score. The risk score will be calculated by a sum of the products of individual values of each predictor variable and its regression coefficient.<sup>47</sup> The full algorithm will be used to produce a score in the first instance to maximise predictive capacity. For the purpose of examining the performance of the predictive tool, patients will be classified as low, medium and high risks, based on their quartile of risk. Those in the highest quartile will be classified as high risk and those in the lowest quartile as low risk. The middle two quartiles will be classified as medium risk.

### Performance

To examine the apparent performance (internal validity) of the prognostic screening tools, we will assess measures of overall performance, calibration and discrimination. Overall performance will be assessed using the Nagelkerke  $R^2$  and Brier score. The Brier score is a method of quantifying differences between actual binary outcomes and their predictions, that is, average prediction error.<sup>35</sup> The Brier score ranges from 0 to 0.25, values close to 0 represent a useful model and values close to 0.25 a non-informative model. Calibration, that is, the agreement between observed and predicted frequencies of a given outcome, will be determined by plotting the mean predicted versus observed cases of chronic LBP for 10 risk stratification levels. The calibration slope and calibration-in-the-large statistic (intercept) will be calculated by constructing calibration plots. Discrimination, that is, the ability of the tool to discriminate between patients who did (+ve case) or did not (-ve case) develop chronic LBP, will be determined by using a Receiver Operator Characteristic Curve analysis, by calculating discrimination slope (box plots) and by examining risk-stratified likelihood ratios. Performance indices and plots will be calculated using R software.<sup>48</sup> Rules for interpretation of these statistics are presented in the Discussion.

After the performance indices have been calculated, we will internally validate the model using bootstrapping techniques suggested by Moons *et al.*<sup>33</sup> (see online supplementary appendix B, Table B). Bootstrapping will be performed in SPSS using syntax available at <http://gjyp.nl/marta/BootstrapValidationOfLogisticRegression.SPS>. To assess model fit and optimism, bootstrapped estimates of the Nagelkerke  $R^2$  and its SE will be compared with the original model estimates. We will conduct a sensitivity analysis to assess performance of the tool for patients in different settings (general practice, physiotherapy, chiropractic).

## Validation of the tool

### Validation sample

The validation sample consists of 1643 participants from a randomised trial conducted over 235 primary care centres in Greater Metropolitan Sydney, Australia. The trial, published elsewhere,<sup>49–51</sup> found no effect of paracetamol on recovery from acute LBP compared with placebo. In short, all participants were adults with acute non-specific LBP (<6 weeks duration) who had presented to primary care (GP, physiotherapist or pharmacist) between November 2009 and March 2013. Participants received up to 4 weeks of regular paracetamol, as-needed paracetamol, or placebo, and were followed up for 3 months on measures of pain and recovery.

### Measures

#### Outcomes

Data on chronic LBP incidence (pain and disability at 3 months) will be extracted from the validation sample. Where necessary, all outcomes extracted from the validation sample will be transformed to match the format of the development sample. Patients will be classified as having chronic LBP, chronic LBP (high pain) and chronic LBP (disability), using the same definitions as were used in the development stage.

### Statistical analysis

Each prognostic model will be assessed in the validation sample using the same statistical procedures as in the development stage. We will calculate estimates of overall performance (Nagelkerke  $R^2$ , Brier Score) discrimination (area under the curve (AUC), likelihood ratios, discrimination slope), and calibration (Calibration plot, Hosmer-Lemeshow test).

### Posterior probability

We will calculate posterior probability along with 95% confidence limits according to the method recommended by Haskins *et al.*<sup>52</sup> The 95% confidence limits will be calculated in Microsoft Excel using the following  $\beta$  distribution equations:

$$\text{Lower boundary} = \text{beta.inv}(0.025, \text{true positive} + 1, \text{false positive} + 1)$$

$$\text{Upper boundary} = \text{beta.inv}(0.975, \text{true positive} + 1, \text{false positive} + 1)$$

### Validation plots

Validation plots of predicted versus observed risks showing the intercept, slope, AUC, scaled Brier score, and  $R^2$  will be produced to summarise and compare performance of the tool in the development and validation samples.

### Updating the tool

Because the validation sample is more recent, we will consider updating and extending the model in the

validation sample. We will update the model using a recalibration method described by Steyerberg *et al.*<sup>53</sup> which involves multiplying the regression coefficients and intercept from the original logistic equation by the calibration slope ( $\beta$ ) and the calibration intercept ( $\alpha$  or 'calibration in the large statistic') from the validation data, to produce a new logistic equation. To extend the model, we will add new, potentially useful predictors from the validation sample for example, sleep quality.

If, based on our prespecified criteria below, we find the predictive validity of the prognostic model to be informative, we will consider simplifying it for clinical use (table 1). This process may involve steps such as refining and specifying measurement of the predictors, simplifying and clearly describing calculation of the prediction score/strata, and producing an electronic or paper-based form designed for clinical application.

## DISCUSSION

We have described the methods and statistical analysis plan to develop and validate a prognostic screening tool for acute LBP. To our knowledge, this tool will be the first of its kind in LBP to follow, a priori, the PROGRESS framework and TRIPOD reporting guidelines for prognostic research. Importantly for secondary prevention, the tool will be developed specifically to predict the onset chronic LBP at its inception at 3 months. The study is limited to using predictors measured in previously collected data sets.

We will use contemporary statistical methods to assess the calibration and discrimination of the screening tool in the two samples. The relative importance of calibration and discrimination ultimately depends on the purpose of the screening tool. For example, if the purpose of the tool is to aid clinical decision-making and provide accurate estimates of risk to patients, then calibration is an important consideration. If a clinician were to inform their patient that they had a 10% chance of chronic LBP, this estimate would be misleading if the tool was not well calibrated and, for example, 40% of patients with the same level of risk actually developed chronic LBP. If, on the other hand, the purpose of the tool is to select appropriate patients to include in a randomised trial, for example, in a trial to prevent chronic LBP, then adequate discrimination is important. In this example, a poorly discriminating tool would misclassify a large proportion of patients, including a number of inappropriate (low risk) patients and excluding appropriate (high risk) patients.

Ideally, a useful screening tool should have discrimination and calibration that is considered informative for its purpose. In general, Hosmer and Lemeshow suggest that for a logistic regression model, an AUC statistic of <0.7 represents poor discrimination, 0.7–0.8 acceptable discrimination, 0.8–0.9 excellent discrimination and 0.9–1.0 outstanding discrimination.<sup>54</sup> However, Steyerberg points out that for clinical decisions that are close to a 'toss-up',

a tool with an AUC of 0.6 can be informative.<sup>35</sup> Published LBP tools report AUC values that range between 0.6<sup>23</sup> and 0.75<sup>55</sup> for pain, and between 0.68<sup>56</sup> and 0.83<sup>55</sup> for disability at 3–6 months of follow-up. On the basis of the reference standard values and those of previous work, we will consider AUC values of less than 0.6 to be non-informative. We also plan to use our additional measures of discrimination (likelihood ratios, discrimination plots) to determine whether or not the tool is informative. For example, overlapping likelihood ratio estimates among low-risk, medium-risk and high-risk groups would indicate poor discrimination. If the posterior probability CI includes the prevalence rate, we will consider the tool to not be informative, and not likely to be clinically useful.

Acceptable calibration of the tool will be based on the results of the calibration plots. If observed frequencies of the chronic LBP in the validation sample fall within 5% of predicted frequencies, the tool will be considered to have acceptable calibration. Calibration in the large statistic (intercept) should be close to 0 and the slope close to 1. With such a large sample, the p value of the Hosmer-Lemeshow test will be interpreted with caution.

## CONCLUSION

This protocol outlines the design of development and validation studies for a prognostic screening tool in acute LBP. Results coming from this study will be interpreted for both clinical and research purposes.

### Author affiliations

<sup>1</sup>School of Medical Sciences, University of New South Wales, Sydney, New South Wales, Australia

<sup>2</sup>Neuroscience Research Australia, Sydney, New South Wales, Australia

<sup>3</sup>Institute of Public Health, University of Heidelberg, Heidelberg, Germany

<sup>4</sup>Hunter Medical Research Institute and School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia

<sup>5</sup>The George Institute for Global Health, University of Sydney, Sydney, New South Wales, Australia

<sup>6</sup>Sansom Institute for Health Research, University of South Australia, Adelaide, South Australia, Australia

**Contributors** CGM, NH, SJK, JM and CMW acquired the original data for this research. AT, JM, MH, CMW, SJK, GLM and NH formulated the methods and designed the protocol. AT drafted the manuscript. All authors contributed to revisions and approved the final version of the manuscript.

**Funding** AT and HL are supported by a National Health and Medical Research Council PhD Scholarships. GLM, CGM and SJK are supported by National Health and Medical Research Council research fellowships. JM and MH are supported by a National Health and Medical Research Council research grant ID 1047827. The two original studies to be used in the planned analysis were funded by National Health and Medical Research Council research grants ID 253634<sup>3</sup> and ID 352576.<sup>23</sup>

**Competing interests** None declared.

**Ethics approval** University of Sydney Ethics Committee.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

- Balaguer F, Mannion AF, Pellise F, *et al.* Non-specific low back pain. *Lancet* 2012;379:482–91.
- Menezes Costa LC, Maher CG, Hancock MJ, *et al.* The prognosis of acute and persistent low-back pain: a meta-analysis. *CMAJ* 2012;184:E613–E24.
- Henschke N, Maher CG, Refshauge KM, *et al.* Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. *BMJ* 2008;337:a171.
- Menezes Costa LC, Maher CG, McAuley JH, *et al.* Prognosis for patients with chronic low back pain: inception cohort study. *BMJ* 2009;339:b3829.
- Koes BW, van Tulder MW, Thomas S. Diagnosis and treatment of low back pain. *BMJ* 2006;332:1430–4.
- Linton SJ. Early identification and intervention in the prevention of musculoskeletal pain. *Am J Ind Med* 2002;41:433–42.
- Hemingway H, Croft P, Perel P, *et al.* Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 2013;346:e5595.
- Riley RD, Hayden JA, Steyerberg EW, *et al.* Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med* 2013;10:e1001380.
- Royston P, Moons KG, Altman DG, *et al.* Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009;338:b604.
- Hingorani AD, Windt DA, Riley RD, *et al.* Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ* 2013;346:e5793.
- Hill JC, Vohora K, Dunn KM, *et al.* Comparing the STarT back screening tool's subgroup allocation of individual patients with that of independent clinical experts. *Clin J Pain* 2010;26:783–7.
- Hancock MJ, Maher CG, Latimer J, *et al.* Can rate of recovery be predicted in patients with acute low back pain? Development of a clinical prediction rule. *Eur J Pain* 2009;13:51–5.
- Traeger A, McAuley JH. STarT Back Screening Tool. *J Physiother* 2013;59:131.
- Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clin J Pain* 1998;14:209–15.
- Hazard RG, Haugh LD, Reid S, *et al.* Early prediction of chronic disability after occupational low back injury. *Spine (Phila Pa 1976)* 1996;21:945–51.
- Hill JC, Dunn KM, Lewis M, *et al.* A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum* 2008;59:632–41.
- Von Korf M, Miglioretti DL. A prognostic approach to defining chronic pain. *Pain* 2005;117:304–13.
- Fritz JM, Beneciuk JM, George SZ. Relationship between categorization with the STarT Back Screening Tool and prognosis for people receiving physical therapy for low back pain. *Phys Ther* 2011;91:722–32.
- Nicholas MK, Linton SJ, Watson PJ, *et al.* Early identification and management of psychological risk factors ("yellow flags") in patients with low back pain: a reappraisal. *Phys Ther* 2011;91:737–53.
- Kent PM, Keating JL. Can we predict poor recovery from recent-onset nonspecific low back pain? A systematic review. *Man Ther* 2008;13:12–28.
- Chou R, Shekelle P. Will this patient develop persistent disabling low back pain? *JAMA* 2010;303:1295–302.
- Hockings RL, McAuley JH, Maher CG. A systematic review of the predictive ability of the Orebro Musculoskeletal Pain Questionnaire. *Spine (Phila Pa 1976)* 2008;33:E494–500.
- Williams CM, Hancock MJ, Maher CG, *et al.* Predicting rapid recovery from acute low back pain based on the intensity, duration and history of pain: a validation study. *Eur J Pain* 2014;18:1182–9.
- Mehling WE, Avins AL, Acree MC, *et al.* Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain? *Eur J Pain* 2015;19:439–46.
- Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997;277:488–94.
- Hush JM, Refshauge K, Sullivan G, *et al.* Recovery: what does this mean to patients with low back pain? *Arthritis Care Res (Hoboken)* 2009;61:124–31.
- Deyo RA, Dworkin SF, Amtmann D, *et al.* Focus article: report of the NIH Task Force on Research Standards for Chronic Low Back Pain. *Eur Spine J* 2014;23:2028–45.
- Von Korf M, Saunders K. The course of back pain in primary care. *Spine (Phila Pa 1976)* 1996;21:2833–7; discussion 38–9.
- Kamper SJ, Hancock MJ, Maher CG. Optimal designs for prediction studies of whiplash. *Spine (Phila Pa 1976)* 2011;36(25 Suppl):S268–74.
- Altman DG, Vergouwe Y, Royston P, *et al.* Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.
- Steyerberg EW, Moons KGM, van der Windt DA, *et al.* Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10:e1001381.
- Haskins R, Osmotherly PG, Tuyl F, *et al.* Uncertainty in clinical prediction rules: the value of credible intervals. *JOSPT* 2014;44:85–91.
- Moons KGM, Altman DG, Reitsma JB, *et al.* Transparent reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration The TRIPOD Statement: explanation and elaboration. *Ann Intern Med* 2015;162:W1–W73.
- Collins GS, Reitsma JB, Altman DG, *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. *Ann Intern Med* 2015;162:55–63.
- Steyerberg EW, Vickers AJ, Cook NR, *et al.* Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass)* 2010;21:128.
- NHMRC. *Evidence-based management of acute musculoskeletal pain*. Canberra, Australia, 2003.
- de Vet HC, Heymans MW, Dunn KM, *et al.* Episodes of low back pain: a proposal for uniform definitions to be used in research. *Spine (Phila Pa 1976)* 2002;27:2409–16.
- Bekkering GE, Hendriks HJ, van Tulder MW, *et al.* Prognostic factors for low back pain in patients referred for physiotherapy: comparing outcomes and varying modeling techniques. *Spine (Phila Pa 1976)* 2005;30:1881–6.
- Pincus T, Burton AK, Vogel S, *et al.* A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine (Phila Pa 1976)* 2002;27:E109–20.
- Steenstra I, Verbeek J, Heymans M, *et al.* Prognostic factors for duration of sick leave in patients sick listed with acute low back pain: a systematic review of the literature. *Occup Environ Med* 2005;62:851–60.
- Ware JE Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
- Von Korf M, Ormel J, Keefe FJ, *et al.* Grading the severity of chronic pain. *Pain* 1992;50:133–49.
- Steyerberg EW, Ewout WS. *Clinical prediction models a practical approach to development, validation, and updating*. Dordrecht: Springer, 2008.
- Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol* 2007;165:710–18.
- Box GE, Tidwell PW. Transformation of the independent variables. *Technometrics* 1962;4:531–50.
- Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 1999;28:964–74.
- Adams ST, Leveson SH. Clinical prediction rules. *BMJ* 2012;344:R: A language and environment for statistical computing. R Foundation for Statistical Computing [program]. Vienna, Austria, 2014.
- Williams CM, Maher CG, Latimer J, *et al.* Efficacy of paracetamol for acute low-back pain: a double-blind, randomised controlled trial. *Lancet* 2014;384:1586–96.
- Williams CM, Maher CG, Latimer J, *et al.* PACE—the first placebo controlled trial of paracetamol for acute low back pain: statistical analysis plan. *Trials* 2013;14:248.
- Williams CM, Latimer J, Maher CG, *et al.* PACE—the first placebo controlled trial of paracetamol for acute low back pain: design of a randomised controlled trial. *BMC Musculoskelet Disord* 2010;11:169.
- Haskins R, Osmotherly PG, Southgate E, *et al.* Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain. *Man Ther* 2014;19:142–51.
- Steyerberg EW, Borsboom GJ, van Houwelingen HC, *et al.* Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med* 2004;23:2567–86.
- Hosmer DW Jr, Lemeshow S. *Applied logistic regression*. John Wiley & Sons, 2004.
- Linton SJ, Boersma K. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Orebro Musculoskeletal Pain Questionnaire. *Clin J Pain* 2003;19:80–6.
- Grotle M, Vollestad NK, Brox JI. Screening for yellow flags in first-time acute low back pain: reliability and validity of a Norwegian version of the Acute Low Back Pain Screening Questionnaire. *Clin J Pain* 2006;22:458–67.
- Henschke N, Maher CG, Refshauge KM, *et al.* Prognosis of acute low back pain: design of a prospective inception cohort study. *BMC Musculoskelet Disord* 2006;7:54.