

Return to Work in a Cohort of Low Back Pain Patients: Development and Validation of a Clinical Prediction Rule

Martijn W. Heymans · Johannes R. Anema · Stef van Buuren ·
Dirk L. Knol · Willem van Mechelen · Henrica C. W. de Vet

Published online: 18 February 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract *Background* From the viewpoint of cost prevention, it is necessary to identify patients that are of high risk for long-term work disability, production loss and sick-leave. *Methods* Secondary data analysis in a cohort of 628 workers on sick-leave between 3 and 6 weeks due to low back pain (LBP). The association of a broad set of demographic, work, LBP and psychosocial related factors on lasting return to work was studied using Cox regression analysis with backward selection. The most relevant factors were used to derive a clinical prediction rule to determine the risk of sick-leave of more than 6 months. Variable and model selection and clinical model performance were performed with bootstrapping techniques. Also the test characteristics of the clinical model were considered. *Results* Longer work absence is related to “moderate” to “poor” job satisfaction, a higher score of fear avoidance beliefs, higher pain intensity at baseline, a longer duration

of complaints and being of female gender. Calibration and discrimination of the clinical prediction rule were 0.90 (slope) and 0.63 (c-index), respectively. The explained variance of 6% of the prediction rule was low and the clinical performance in terms of sensitivity, specificity, positive and negative predictive values at specific cut-off points was moderate. *Conclusions* Our study confirmed the importance of demographic, work, LBP and psychosocial related factors on the prediction of long-term sick-leave. When these factors were used to derive a clinical prediction rule the performance was moderate. As a consequence, prudence has to be taken when using the prediction rule in practice.

Keywords Low back pain · Clinical prediction rule · Occupational health care · Sick-leave

M. W. Heymans · J. R. Anema · W. van Mechelen
Body@Work, Work and Health, Research Center Physical Activity, TNO-VUmc, Amsterdam, The Netherlands

M. W. Heymans (✉) · D. L. Knol · H. C. W. de Vet
Department of Epidemiology and Biostatistics, EMGO-Institute, VU University Medical Center, Van der Boechorststraat 7, Room D 518, 1081 BT Amsterdam, The Netherlands
e-mail: mw.heymans@vumc.nl

J. R. Anema · W. van Mechelen
Department of Public and Occupational Health, EMGO-Institute, VU University Medical Center, Amsterdam, The Netherlands

S. van Buuren
TNO Quality of Life, Leiden, The Netherlands

S. van Buuren
Department of Methodology and Statistics, University of Utrecht, Utrecht, The Netherlands

Introduction

Low back pain (LBP) is the most common and expensive musculoskeletal disorder in Western countries [1]. The recovery process of persons with chronic LBP is slow, and their demands on the health care system are both large and costly. Costs due to work disability and production losses are even higher. These and medical costs in the United States (US), United Kingdom (UK) and the Netherlands were estimated yearly to be US \$50 billion [2], US \$11 billion [3] and US \$5 billion [4], respectively. For preventive reasons it is needed to identify which patients are responsible for long-term disability, production losses and sick-leave. If these high risk patients can be identified preventive actions can be taken by health care providers. This may reduce costs and is therefore of benefit from both a medical and an economical perspective.

Until now a number of systematic reviews has been conducted to determine which variables are related to the development of chronic prolonged functional disability and/or sick-leave [5–8]. The reviews of Crook et al. [5] and Steenstra et al. [6] both reported strong to moderate evidence for the following factors, previous episodes of LBP, higher levels of disability, older age and female gender. The reviews of Linton [7] and Hartvigsen et al. [8] studied the influence of psychosocial factors at work. Linton [7] found evidence for factors as paincoping and fear avoidance and Hartvigsen et al. [8] found evidence of no association for social support. Inconsistent findings between reviews were found for the variables radiating pain, job satisfaction, and heavier work. Obvious from these reviews is that many studies have been conducted with as aim to examine the prognosis of functional disability and/or sick-leave and that they do not agree on relevant factors. Therefore, it is still unclear which of these factors separately or in combination, contribute most to prolonged sick-leave and disability due to LBP studies on RTW and functional disability need comprehensive multivariate models and have to focus more on the feasibility of using this evidence for clinical practice.

Clinical prediction rules seem attractive to discriminate between high and low risk groups of patients with a poor outcome, which may support treatment decisions [9]. McGinn et al. [10] have defined four levels of evidence to develop good clinical prediction rules. Level 4 is the lowest level, i.e., the prediction rule is derived and validated in the original but not in another sample. Level 3 indicates that the prediction rule is validated in another small sample. For Level 2 it is necessary that the prediction rule has shown to be accurate in a large sample or that it has been validated in several other different settings. Level 1, i.e., the highest level, indicates that the use of the prediction rule lead to a change in clinician behaviour and improvement of patient outcomes (impact analysis). To successfully move through these levels, a prediction rule must preferably be developed by using an outcome measure that is of relevance for the clinical setting and must consist of predictors that can easily be assessed in daily practice. Furthermore, the predictive ability in terms of explained variance, discrimination and calibration has to be tested as well as the generalizability in new patients. However, these aspects of a prediction rule are often not well established [9].

The purpose of our study is to compose a clinical prediction rule in a large dataset of employees on sick-leave due to LBP that determines the risk of prolonged sick-leave of more than 6 months by using information from a wide range of demographic, work, LBP and psychosocial related factors. Also the predictive ability in terms of the explained variance, discrimination and calibration of this rule will be validated internally.

Methods

Study Design

Prospective cohort study with merged data (628 patients) from three different RCTs on LBP. The first trial investigated the effectiveness of a behaviorally oriented graded activity program in comparison to usual care (134 patients) [11]. The second trial compared a workplace intervention and graded activity to usual care (195 patients) [12]. The third trial compared high and low intensity back schools with usual care (299 patients) [13].

Study Population

Study patients visited their occupational physician (OP) at one of participating occupational health services (OHS) when they were on sick-leave for not more than 8 weeks. All studies had a follow-up of at least 1 year. The population consisted of blue and white-collar workers covering a broad range of professions. Patients were eligible for participation if they met the following inclusion criteria: non-specific LBP, defined as pain localised in the lower back without a specific underlying cause; sick-listed due to LBP (completely or partially) for not more than 8 weeks; between 18 and 65 years of age; ability to complete questionnaires written in Dutch. Exclusion criteria were: specific cause of the LBP; pregnancy; serious psychiatric disorders; juridical conflict at work. The study protocol was approved by the Medical Ethics Committee of the VU University Medical Center in Amsterdam and patients who met the eligibility criteria and who were willing to participate signed an informed consent form.

Outcome

The outcome variable lasting return to work (RTW) was defined as the duration of work absenteeism in calendar days from the first day of sick-leave until full return to own work or other work with equal earnings for at least 4 weeks. Next to this variable, a status variable was defined. On this variable, participants were assigned a “1” when they fulfilled the RTW outcome variable and were assigned a “0”, indicating “censoring” when they were on sick-leave at 6 months follow-up. People that went back on sick-leave before the 4 weeks period ended, until the end of the study period, were also classified as being on sick-leave. The sick-leave data were collected continuously from the electronic medical records of the OHS [14].

Selection of Prognostic Factors

The selection of relevant prognostic factors was performed in two steps. First, the literature on prognosis for sub-acute

LBP and sick-leave was reviewed [5–8]. On basis of this and the prognostic factors measured in the three RCTs we composed a list of prognostic factors. Second, we presented this list of potential prognostic factors by e-mail to 42 occupational physicians (OPs) working in different sectors. They were asked to judge, according to their expert opinion, whether each indicator would contribute to the development of chronic LBP. We also asked them their opinion on whether the factors were modifiable by treatment. The indicators: smoking, body weight and height, and years working in current job were not considered relevant after the literature review and the judgment of the OPs, so we excluded these. The prognostic factors were assessed by means of self-reported questionnaires before inclusion in the studies.

Potential Prognostic Factors

The following factors were considered important: age (years), gender (male/female), duration of complaints prior to randomization (in weeks), radiation to one or both legs (yes/no) and treatment during study enrolment (yes/no). Patients were also asked to rate the level of certainty that they would be working fulltime in 6 months time (“self-predicted certainty”). For this prognostic factor they answered the question “How certain are you about full work resumption at 6 months” using a 5-point scale: “completely uncertain”, “a little certain”, “somewhat certain”, “certain”, “completely certain”. Physical activity was measured with the Baecke questionnaire [15]. Pain intensity and functional status at baseline were scored by using a numerical visual analogue scale (VAS, range 0–10) [16] and the roland disability questionnaire (RDQ, range 0–24) [17], respectively. Potential job-related physical factors were measured by the section ‘Musculoskeletal Workload’ of the dutch musculoskeletal questionnaire (DMQ) [18]. The physical factors we examined were daily: lifting, bending and twisting of the trunk, driving a vehicle at work (whole body vibration), and stooping consisting of the answer categories: “never”, “sometimes”, “quite frequently”, “much frequently”. The DMQ is used to identify high and low risk groups for musculoskeletal disorders like LBP and has fair validity. Potential work-related psychosocial factors were measured by means of a Dutch version of Karasek’s Job Content Questionnaire (JCQ). Dimensions of this questionnaire are: quantitative job demands, decision authority, skill discretion, supervisor support and co-worker support [19]. According to Karasek’s model, quantitative job demands were called job demands, decision authority and skill discretion were merged into job control and supervisor and co-worker support were merged into social support. Job satisfaction was assessed by means of a question concerning job task satisfaction consisting of

the answer categories: “no good”, “reasonable”, “moderate” and “good”. The Dutch version of the tampa scale for kinesiophobia (TSK, range 17–68) was used to measure the extent to which people feared that exercise can lead to reinjury [20]. A high score indicates much fear for physical activity or injury. Fear of movement, avoidance of activities and back pain beliefs were measured with the Fear Avoidance Beliefs Questionnaire (range 0–42) [21]. Coping with pain was measured with the pain coping inventory (PCI) questionnaire. The PCI questionnaire measures cognitive and behavioural coping strategies of pain patients. The questionnaire consists of six subscales: transformation of pain, distraction, lowering demands, withdrawal, worrying, and resting [22]. The first three subscales were merged to obtain a measure of active paincoping and the latter three subscales were merged to obtain a passive paincoping measure, respectively.

Analyses

Model Building Process

The Cox proportional hazard model was used to examine the prognostic value of each factor. In each Cox model RTW was fitted as the dependent variable and the prognostic factors as the independent variables. We adjusted for the effects of the interventions. The Cox proportional hazard assumption was affirmed by plotting log minus log survival curves.

To fill in variables with missing values, we applied multiple imputation by using the multiple imputation by chained equation (MICE) package [23]. This is a flexible imputation method, which allows one to specify the multivariate structure in the data as a series of conditional imputation models based on the information of other variables. We generated ten multiple imputed data sets.

Variable and model selection in these ten data sets was performed by a two-stage bootstrap modelling approach [24]. During the first stage backward regression, with a P value selection criterion of 0.157 [25], was applied on 200 bootstrap samples. The bootstrap samples were drawn with replacement from the imputed data sets and were of equal sample size as this original sample. On basis of the selection frequencies in these bootstrap samples prognostic factors that were selected in more than 50% of the regression models were included at the second modelling stage [24]. As a sensitivity analyses we also included factors that were selected in more than 40% of the models. In the second stage (500 bootstrap samples) again backward regression was performed on each of the bootstrap samples with the same P -value selection criterion as in stage one.

Now, the frequency of selected models was calculated by using the factors selected in the first stage.

The Cox regression coefficients, standard errors and 95% confidence intervals of the final model were estimated over the ten imputed data sets according to Rubin's rules [26]. During the modeling process we also considered the balance between the number of variables and events/non-events in the models, which is recommended not to be lower than 10–15 events per variable [27].

To make risk prediction available in daily practice, we transformed the final model into a clinical prediction rule.

Derivation of the Clinical Prediction Rule

1. We calculated the regression coefficients of the Cox model for RTW over a period of 12 months follow-up and used these coefficients to calculate risk scores of prolonged sick-leave for more than 6 months follow-up. We choose for this procedure because regression coefficients that are calculated over a period of 12 months follow-up will provide more accurate estimates of RTW than those calculated over a period of 6 months follow-up. These regression coefficients will be re-used in step 3.
2. The survival function under the Cox model is defined as: $S(t) = S_0(t)^{\exp(XB)}$, where $S(t)$ stands for each patient's probability of prolonged sick-leave, $S_0(t)$ is the time-dependent risk of prolonged sick-leave when all predictor variables are zero, and XB is the linear equation of the predictor variables (X) and the regression weights (B) [28].
3. The survival function can be rewritten into: $S_0(t) = S(t)^{1/\exp(XB)}$. Within this formula the regression coefficients of step 1 are used to calculate the baseline risk of prolonged sick-leave for each patient according to his or her own time of follow-up.
4. Subsequently, the risk of prolonged sick-leave at 6 months could be identified from the patients with a follow-up time of 6 months.
5. Now the survival function was again applied. For $S_0(t)$ the risk of prolonged sick-leave at 6 months (obtained at step 4) was used for all patients. Only the values for the predictor variables (X) and the regression weights (B) of each patient were responsible for the differences in risk of prolonged sick-leave. Thus the following formula was used here: $S(6 \text{ months}) = S_0(6 \text{ months})^{\exp(XB)}$.
6. To derive practical interpretable scores for the prediction rule, the Cox regression coefficients of step 1 were multiplied by 10 and rounded to the nearest integer. These points can be multiplied by the values for the predictors and added up for each patient to calculate the total risk score on prolonged sick-leave.

Performance of the Prediction Rule

Discrimination was calculated by the c-index. The c-index equals the area under the receiver operating characteristic (ROC) curve in logistic regression [29]. Calibration refers to the agreement between the observed probabilities in the original data and the predicted probabilities of the prediction rule. To plot the calibration curve, the predicted and observed probabilities were presented on a scatter diagram. To determine the observed probabilities, the Kaplan–Meier estimate of the 6-months risk of not having returned to work was calculated for each decile of predicted risk [29]. The explained variance was also calculated, which is an indication of how much of the variance between patients in the outcome can be explained by the predictors [30]. We used bootstrapping to correct the regression coefficients, the c-index and the explained variance for over-optimism. Over-optimism is determined by the slope of the linear predictor, which was calculated in 200 bootstrap samples and tested in the original sample. The average difference reflects the amount of overfitting present and was used as a shrinkage factor to preshrink the regression coefficients. The model performance indices were calculated on each imputed data set and averaged over the ten imputed data sets.

Software

The MICE as well as the backward selection procedures were performed with R software [23, 27].

Results

Table 1 gives an overview of the mean values and proportions of the potential prognostic factors and of the % of missing values. Most of the missings occurred because some variables were included in only 2 of the 3 RCTs. After 12 months 577 patients had returned to work (92%), with follow-up times ranging from 4 to 364 days.

The inclusion frequencies of the factors at the first bootstrap selection stage ranged from 9.2 to 99.1% (Table 2). The indicator job satisfaction had the highest inclusion frequency. For bootstrap stage two, six factors with inclusion frequencies of more than 50% were included and with these six factors (plus the treatment factor), 46 different models were selected. As shown by the model with the highest selection frequency, longer work absence is related to job satisfaction, daily stooping, fear avoidance beliefs, pain intensity at baseline, duration of complaints and gender. This model was selected 41.6% of the time. We compared this model with the model obtained with factors selected at least 40% of the time. With this

Table 1 Patient characteristics at baseline ($n = 628$) and the percentage of missing information for the potential prognostic factors of sick-leave

	Value	Missing (%)
Age (mean years \pm SD)*	40.6 (9.5)	0
Gender (male, %)	71.0	0
Self-predicted certainty at 6 months (%)		24.3
Completely uncertain	5.9	
A little certain	12.9	
Somewhat certain	23.1	
Certain	18.5	
Completely certain	15.3	
Physical activity (mean \pm SD)	8.8 (1.0)	44.6
Job satisfaction (%)		2.8
No good	3.5	
Moderate	10.4	
Reasonable	40.3	
Good	43.0	
Job Content Questionnaire (mean \pm SD)		
Job control	56.2 (9.2)	25.6
Job demands	33.1 (4.8)	24.7
Social support	22.5 (4.1)	24.8
Daily exposed to:		
Lifting >25 kg per day (%)		24.2
Never	29.6	
Sometimes	19.7	
Quite frequently	14.8	
Much frequently	11.6	
Bending and twisting of the trunk (%)		24.2
Never	18.8	
Sometimes	18.6	
Quite frequently	23.1	
Much frequently	15.3	
Whole body vibration (%)		26.4
Never	42.0	
Sometimes	13.4	
Quite frequently	12.4	
Much frequently	5.7	
Stooping (%)		25.2
Never	15.1	
Sometimes	22.6	
Quite frequently	22.5	
Much frequently	14.6	
Duration of complaints (months) prior to randomization; median (IQR)**	5.8 (13.3)	33.3
Pain radiation in 1 or both legs (%)	33.8	2.1
Functional status (mean \pm SD)	11.3 (5.2)	5.1
Pain intensity (mean \pm SD)	6.2 (1.9)	3.0
Pain coping, active (mean \pm SD)	6.7 (1.2)	5.6
Pain coping, passive (mean \pm SD)	6.5 (1.3)	7.0

Table 1 continued

	Value	Missing (%)
Fear avoidance beliefs (mean \pm SD)	19.5 (9.7)	48.1
Kinesiophobia (mean \pm SD)	39.8 (6.7)	6.2

* Standard deviation

** InterQuartile range

criterion, the final model included ten factors, compared to the six factors with the “50%” rule. The final model with ten factors was chosen much more infrequently between the stage two bootstrap samples—only 9.3% of the time—and it showed equal performance with regard to calibration and discrimination as the six factor model (data not shown). Therefore, we opted for the more parsimonious “50%” model as our best model.

With this model a clinical prediction rule was derived, which is presented in Table 3. In this prediction rule the variable daily exposure to stooping, unexpectedly, contributed to a lower risk of prolonged sick-leave. We tested the performance of the prediction rule when we in- and excluded this variable and these models performed equally well (data not shown). The prediction rule is therefore presented without the variable daily exposure to stooping. Table 3 shows that longer work absence is related to moderate to poor job satisfaction, a higher score for fear avoidance beliefs, higher pain intensity at baseline, a longer duration of complaints and being of female gender.

With the risk scores presented in Table 3, the risk of developing prolonged sick-leave of more than 6 months can be calculated for each individual patient. For example, a male LBP patient, who experiences moderate job satisfaction, is not extremely fear avoidant (score of 15), has a low pain intensity score (score of 2), and a long duration of complaints (8 weeks) will have a total risk score of -2 (gender) + 2 (job satisfaction) + 3 (fear avoidance) + 1.2 (pain intensity) + 0.08 (duration of complaints) = 4.3 . This patient will have a risk of developing prolonged sick-leave of more than 6 months of 12% (last column of Table 4).

The bootstrap corrected explained variance of the model was 6% (Table 3). This indicates that 6% of the variance between patients in the outcome can be explained by the predictors in the model. The bootstrap corrected c-index was 0.63 and indicates that in 63% of the patients the prediction rule discriminates well between a high and a low risk patient to develop long-term sick-leave. The calibration slope was 0.9. This indicates that a little overoptimism is to be expected when applying the prediction rule in new workers with LBP.

The score of categories and their related mean absence days and observed and predicted probabilities of prolonged

Table 2 Selection frequencies of variables at step 1 and models at step 2 as a result of the two-step bootstrap model averaging (BMA) approach

	Step 1 (%)	Models selected in Step 2 (the first 10 models are shown)									
		1	2	3	4	5	6	7	8	9	10
1. Job satisfaction	99.1	✓	✓	✓	✓	✓	✓	✓	✓	–	–
2. Daily stooping	77.5	✓	✓	✓	✓	–	✓	–	✓	✓	✓
3. Fear avoidance beliefs	68.2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4. Pain intensity	67.0	✓	✓	✓	✓	✓	–	✓	–	✓	✓
5. Duration of complaints	55.5	✓	✓	–	–	✓	✓	✓	✓	✓	✓
6. Gender	50.4	✓	–	✓	–	–	–	✓	✓	–	✓
7. Daily bending and twisting	48.6										
8. Pain radiation	48.1										
9. Prediction at 6 months	48.0										
10. Job control	43.5										
11. Daily whole body vibration	36.4										
12. Kinesiophobia	35.4										
13. Physical activity	32.3										
14. Daily lifting	31.7										
15. Social support	23.3										
16. Age	22.5										
17. Job demands	19.9										
18. Treatment	16.6										
19. Active pain coping	15.9										
20. Passive pain coping	15.7										
21. Functional disability	9.2										
Frequency of that specific model		41.6	19.5	5.4	4.9	4.0	3.0	2.8	2.7	2.2	1.8

✓, Variables included in the model; –, variable not selected

sick-leave at 6 months is presented in Table 4. Also presented in this table are the test characteristics in terms of sensitivity, specificity and positive and negative predictive values at different risk score categories. The mean observed probability of sick-leave at 6 months averaged over all score categories was 18.5%. Overall, the observed probability was almost similar to the predicted probability. This pattern can also be seen in the calibration curve in Fig. 1. In general, the pairs of predicted and observed probabilities are near the ideal line of perfect calibration.

The sensitivity and specificity of the rule are moderate to low. If a cut-off level of ≥ 10 is chosen as an indication of a high risk of prolonged sick-leave then 32% of the patients who actually progressed to prolonged sick-leave are correctly identified (sensitivity). From the patients with a score of < 10 , i.e., the group with a low risk of prolonged sick-leave, 89% who does not develop prolonged sick-leave is correctly classified (specificity). More important for clinical practice are the predictive values of the rule. The negative predictive value (NPV) of 84% in the patients with a low risk score of < 10 means that these patients will not receive an intervention, and this is correct. However,

16% (1-NPV) of these patients will develop prolonged sick-leave but are not targeted for an intervention. The positive predictive value (PPV) of the rule at the score level of ≥ 10 is 41%. This means that in this high risk group of patients an intervention is justified in 41% of the cases because they actually will develop prolonged sick-leave. However, this also means that in 59% (1-PPV) of these high risk patients an intervention is applied, while they will not develop prolonged sick-leave.

Discussion

In this study we developed a model for the prediction of prolonged work absence at 6 months follow-up in a cohort of workers on sick-leave due to LBP by the use of a broad spectrum of demographic, work, LBP and psychosocial related prognostic factors. We identified that moderate to poor job satisfaction, a higher score of fear avoidance beliefs, higher pain intensity at baseline, a longer duration of complaints and female gender were associated with a higher risk of not returning to work at 6 months.

Table 3 Factors included in the multivariable model together with the index for discrimination (c-index) and calibration (slope)

Variable	Coefficient ^a	HR ^b	95% CI	Risk score ^c
Job satisfaction				
Good	0	Reference category		0
Reasonable	0.07	1.07	0.90–1.27	–1
Moderate	–0.22	0.80	0.61–1.04	2
Poor	–0.52	0.59	0.37–0.94	5
Fear avoidance beliefs (0–42)	–0.02	0.98	0.97–1.00	0.2 × FA ^e
Pain intensity (0–10)	–0.06	0.94	0.90–0.98	0.6 × PI ^f
Duration of complaints (weeks)	–0.001	1.00	0.97–1.00	0.01 × DU ^g
Gender (male)	0.15	1.16	0.98–1.39	–2
Explained variance (R^2)				7/6% ^d
C-index (discrimination)				0.64/0.63 ^d
Slope (calibration)				0.90

^a Values are regression coefficients after bootstrap correction for overoptimism

^b HR > 1 means a shorter time until RTW and a HR < 1 means a longer time until RTW

^c Risk score of prolonged sick-leave of more than 6 months

^d Apparent and bootstrap corrected c-index respectively

^e FA = value of patient for fear avoidance beliefs

^f PI = value of patient for pain intensity

^g DU = value of patient for duration of complaints

Table 4 Score categories, associated observed and predicted risk score probabilities of prolonged sick-leave and test characteristics of prediction rule at 6 months (%)

Score categories	N (%)	Mean absence days	Observed probability (%)	Predicted probability (%)	Se ^a	Sp ^b	PPV ^c	NPV ^d
<2	110 (17.5)	53.6	0.02	0.04				
≥2 (2, 5)	199 (31.7)	81.5	0.12	0.12	98	21	23	98
≥6 (6, 9)	223 (35.5)	99.2	0.21	0.21	73	55	28	89
≥10	96 (15.3)	132.2	0.39	0.38	32	89	41	84

^a Sensitivity

^b Specificity

^c Positive predictive value

^d Negative predictive value

Comparison with Findings in the Literature

In our study “moderate” and “poor” job satisfaction was found to be responsible for a longer time until RTW. Job satisfaction in relation to RTW is part of the evidence that work-related psychosocial characteristics might be relevant factors for workers to maintain work [8]. In a recent review it was concluded that there is strong evidence that job satisfaction is not associated with a longer time off work [6]. Adding our study to this review would result in conflicting evidence of job satisfaction as an important factor associated with sick-leave due to LBP. This indicates that job satisfaction might influence the course of disabling LBP and that evidence of this factor has to be confirmed.

There is a lack of prognostic studies in occupational health care that examine associations between psychological factors, like fear avoidance beliefs, and RTW [31]. Psychological beliefs have been identified as important factors in relation to the course of LBP and accompanying sick-leave [32]. The effect of fear avoidance beliefs on prolonged sick-leave in our study was small. A more pronounced relationship of fear-avoidance with future LBP work absence was shown by Fritz et al. [33]. Although the role of psychosocial beliefs in association with RTW is inconclusive [7], the presence of this association in the context of other potential prognostic factors in an occupational setting needs further attention.

A higher level of pain intensity at baseline and a longer duration of complaints at study inclusion have been

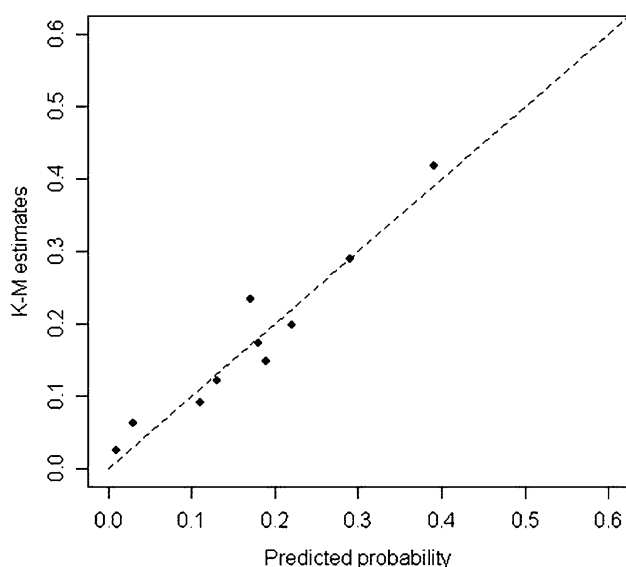


Fig. 1 Calibration curve of the prediction rule of the risk of prolonged sick-leave at 6 months (shrunken regression coefficients). The dotted line indicates perfect calibration

identified as relevant prognostic factors in several studies before [32, 34]. The duration of complaints can be related to the severity of the LBP, which in turn may be responsible for a longer work absence. Von Korff et al. [34] stated that pain intensity is related to pain severity and limitations in functioning and work. Von Korff et al. showed that back pain does not have to be present all the time. The pain may be present in the background for a longer time at a lower level of pain intensity and may flare-up. Flare-ups are frequently seen in chronic LBP patients and are in combination with higher levels of functional limitations, like problems with work activities, responsible for higher pain severity and consequently work absence [34].

In the current study gender, i.e., female sex, was associated with a higher risk of sick-leave at 6 months follow-up. The review of Steenstra et al. [6] showed that female gender is associated with a longer time off work. An explanation of this gender difference on sick-leave duration can be found in the vulnerability hypothesis. This hypothesis states that due to differences in biological (e.g., hormone physiology) or psychological factors (e.g., coping strategies) similar exposures at work might have a larger negative effect on women than on men [35].

In our study, a more frequent exposure to daily stooping did not contribute to a higher risk of prolonged sick-leave. Daily exposure to stooping is an aspect of the high workload that workers experience if they are frequently exposed to it. This factor was associated with more sickness absence due to LBP in an etiological study [36]. However, the role of this factor on the prognosis of sick-leave in LBP patients is less obvious. Some prognostic studies did find an effect

of high workload on longer sick-leave [37, 38], but others did not [39, 40]. We tested in the current study if the inclusion and exclusion of the variable stooping was responsible for a change in the estimates of the regression coefficients of other variables in the final model or if it caused a change in the model performance. In both test situations similar results were produced. Therefore, we choose to report the model without the inclusion of the variable daily exposure to stooping.

The explained variance of our prediction rule was 6%. In the study of Pransky et al. [41] an explained variance of 12% was reported. Dionne et al. [42] did not present a value for the explained variance. The explained variance of the prediction model in the study of Heymans et al. [43] was, 23.7%. Other studies that developed prognostic models for RTW in an occupational setting reported explained variances of 18–30% [39, 44]. Obvious is that the values of the explained variance strongly differ between studies but that they are not high in general. A low value for the explained variance means that prognostic factors can only explain a small fraction of the variance between individual patients. We still might have missed variables that may play a role in the complex environment of occupational health care and that may influence the prognosis of sick-leave, e.g., the maintenance of contact with the employer during the sick-leave period turned to be important in a recent study [45]. It also has to be noted that the explained variance of a Cox regression model is low in general even if there are strong and highly significant predictors in the model [46].

With respect to calibration, the slope index of our prediction model was 0.90. This means that the observed and predicted probabilities are well in agreement. The bootstrap-corrected c-index of our model of 0.63 was moderate. We found three similar prediction rules for LBP patients in an occupational setting that used the outcome measure RTW. Dionne et al. [42] developed a prediction tool to identify workers at high risk of adverse occupational outcomes. Pransky et al. [41] developed a practical screening model to predict length of disability after acute occupational LBP. Heymans et al. [43] developed a nomogram to predict work status in chronic LBP patients. Their nomogram had presented a slope index of 0.91 and a c-index of 0.76. Both slightly higher than in our current study. The studies of Pransky et al. [41] and Dionne et al. [42] did not report on calibration or discrimination of their prediction tools. Pransky et al. also did not report on PPV or NPV. Dionne et al. reported PPVs of 33–57% and NPVs for their model of 74–91%. The study of Heymans et al. [43] reported PPVs in the range of 70–95% and NPVs in the range of 33–100%. With respect to the practical implication of our prediction rule we are aware that prudence has to be taken when using the prediction rule in practice.

Considering an adequate cut point, at the score level of 2 the NPV is 98%, which means that the majority of patients with a score lower than 2 will not receive an intervention, which is correct. At the score levels of 6 and 10 the NPV of 89 and 84%, respectively are still high. The PPV of 41% at the score level of 10 is moderate, which means that 41% of the patients with a score of 10 and higher (i.e., patients with a poor prognosis) correctly receive an intervention. However, 59% (100%-PPV) of the patients with a score of 10 and higher receive an intervention despite their good prognosis. This may lead to misusing health care resources, high treatment costs, potential for iatrogenesis or that sick-leave duration takes longer than expected because patients think that treatments should be completed before full RTW is reached. However, from an employers' perspective it seems more attractive to refer each patient with a score of 10 or higher to a specific treatment to stimulate work ability and work resumption.

Remarks on Our Study: Choices in the Data Analysis

In our prediction model we adjusted for the treatment effects that were examined in each RCT that delivered the patient data. Therefore, treatment effects may have influenced prediction of RTW. However, the treatment effect of usual care [11–13], and low and high intensity back schools were small [13]. The graded activity intervention showed a beneficial effect on RTW in one RCT [11] and an opposite effect on RTW in another study [12]. These interventions will therefore have a small impact on prediction of RTW. A workplace intervention proved to stimulate RTW [12]. The prognostic score of a patient may slightly improve when he will be referred to this intervention. Whether the risk scores obtained by our prediction rule can be improved in new patients by specific interventions has to be confirmed in a future RCT. We therefore present our prediction rule without involvement of the treatment variable.

Study Strengths and Limitations

The strength of our study is that we were able to include almost all variables that are mentioned in the literature as prognostic factors. Moreover, the selection of relevant predictors was not only based on evidence in the literature but also on clinical expertise of OPs. This may enhance clinical applicability of the prediction rule. According to the Dutch occupational health care guidelines, workers have to visit their OP when they are on sick-leave for not more than 8 weeks due to LBP, which was also the case in our study. Therefore, our study results are generalizable to Dutch occupational health care practice.

The success of using clinical prediction rules in practice depends among other things on how much time it will take

for the patient and/or clinician to determine the final risk score and probability of outcome. Our prediction rule consists of five variables that can easily be answered by the patient. Most variables consist of one question only, i.e., job satisfaction, pain intensity, duration of complaints and gender. For the variable fear avoidance beliefs 16 questions have to be answered, which makes a total of 20 questions. The prediction rule can easily be offered via a desktop computer, in an Excel format or as a web application. In this form it will take 5–10 min to fill in the prediction rule and might it be feasible to administer the rule on a routine basis just before or during the appointment with the clinician.

A limitation of our study is that we did not test the generalizability of the prediction rule in new similar patients (external validation) [29]. Even after correction for optimism the performance of the model may decrease due to other population characteristics. The need for external validation after adjusting for optimism was for example reported in the study of Bleeker et al. [47]. However, there are good indications that internal validation by using bootstrap-corrected indices of discrimination produce estimates that can also be expected in future patients [48].

A limitation of our study is the low explained variation and moderate clinical performance of the prediction rule. This means that not each individual risk profile might be accurate enough so that it can be used in practice to guide treatment decisions. As a consequence prudence has to be taken when using the prediction rule in practice.

Conclusion

Our study confirmed the importance of demographic, work, LBP, and psychosocial related factors on the prediction of long-term sick-leave. These factors were used to derive a clinical prediction rule to make risk prediction of prolonged sick-leave of more than 6 months due to LBP. By taking a cut-off point of the risk score of larger than ten, the NPV is high, thereby preventing treatment in persons at a low risk of prolonged sick-leave. However, the performance of the prediction is only moderate and the explained variance is low.

Acknowledgments We would like to thank Ivan Steenstra, Bart Staal and Hynek Hlobil for providing their data. **Funding:** This study is granted by: STECR, the platform for occupational health and reintegration at work. The funding body had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication. There are no competing interests.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Andersson GBJ. Epidemiological features of chronic low-back pain. *Lancet*. 1999;354:581–5. doi:10.1016/S0140-6736(99)01312-4.
2. Frymoyer JW. Quality: an international challenge to the diagnosis and treatment of disorders of the lumbar spine. *Spine*. 1993;18:2147–52. doi:10.1097/00007632-199311000-00001.
3. Maniadakis N, Gray A. The economic burden of back pain in the UK. *Pain*. 2000;84:95–103. doi:10.1016/S0304-3959(99)00187-6.
4. van Tulder MW, Koes BW, Bouter LM. A cost-of-illness study of back pain in the Netherlands. *Pain*. 1995;62:233–40. doi:10.1016/0304-3959(94)00272-G.
5. Crook J, Milner R, Schultz IZ, et al. Determinants of occupational disability following a low back injury: a critical review of the literature. *J Occup Rehabil*. 2002;12:277–95. doi:10.1023/A:1020278708861. Review.
6. Steenstra IA, Verbeek JHAM, Heymans MW, et al. Prognostic factors for return to work in patients sick listed with acute low-back pain, a systematic review of the literature. *OEM*. 2005;62:851–60. doi:10.1136/oem.2004.015842. Review.
7. Linton SJ. A review of psychosocial risk factors in back and neck pain. *Spine*. 2000;25:1148–56. doi:10.1097/00007632-200005010-00017. Review.
8. Hartvigsen J, Lings S, Leboeuf-Yde C, et al. Psychosocial factors at work in relation to low back pain and consequences of low back pain; a systematic, critical review of prospective cohort studies. *Occup Environ Med*. 2004;61:e2. Review.
9. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144:201–9.
10. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-based medicine working group. *J Am Med Assoc*. 2000;284:79–84. doi:10.1001/jama.284.1.79.
11. Staal JB, Hlobil H, Twisk JWR, et al. The effects of graded activity for low back pain in occupational health on sick leave, functional status and pain: a randomized controlled trial. *Ann Intern Med*. 2004;140:77–84.
12. Steenstra IA, Anema JR, Bongers PM, et al. Cost effectiveness of a multi-stage return to work program for workers on sick leave due to low back pain, design of a trial. *BMC Musculoskelet Disord*. 2003;4:26. doi:10.1186/1471-2474-4-26.
13. Heymans MW, de Vet HCW, Bongers PM, et al. Back schools in occupational health care: design of a randomized controlled trial and cost-effectiveness study. *J Manipulative Physiol Ther*. 2004;7:457–65. doi:10.1016/j.jmpt.2004.06.004.
14. van Poppel MN, de Vet HCW, Koes BW, et al. Measuring sick-leave: a comparison of self-reported data on sick-leave and data from company records. *Occup Med (Oxford, England)*. 2002;52:485–90.
15. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *Am J Clin Nutr*. 1982;36:936–42.
16. Carlsson AM. Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale. *Pain*. 1983;16:87–101. doi:10.1016/0304-3959(83)90088-X.
17. Gommans IHB, Koes BW, van Tulder MW. Validity and responsiveness of the Dutch roland disability questionnaire. *Ned Tijdschr Fysioth*. 1997;107:28–33. (In Dutch: Validiteit en responsiviteit van de Nederlandstalige Roland Disability Questionnaire).
18. Hildebrandt VH, Bongers PM, van Dijk FJ, et al. Dutch Musculoskeletal Questionnaire: description and basic qualities. *Ergonomics*. 2001;44:1038–55. doi:10.1080/00140130110087437.
19. Karasek RA, Brisson C. The job content questionnaire (JCQ): an instrument for internationally comparative assessments of psychosocial job characteristics. *J Occup Health Psychol*. 1998;3:322–55. doi:10.1037/1076-8998.3.4.322.
20. Roelofs J, Goubert L, Peters ML, et al. The tampa scale for kinesiophobia: further examination of psychometric properties in patients with chronic low back pain and fibromyalgia. *Eur J Pain (London, England)*. 2004;8:495–502. doi:10.1016/j.ejpain.2003.11.016.
21. Waddell G, Newton M, Henderson I, et al. A fear-avoidance beliefs questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain*. 1993;52:157–68. doi:10.1016/0304-3959(93)90127-B.
22. Kraaimaat FW, Bakker A, Evers AWM. Pain coping strategies in chronic pain patients: the development of the pain-coping-inventory list. *Gedragstherapie*. 1997;30:185–201.
23. Van Buuren S, Oudshoorn K. Flexible multivariate imputation by MICE. Technical report. Leiden, The Netherlands: TNO Quality of Life; 1999.
24. Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: application to the Cox regression model. *Stat Med*. 1992;11:2093–109. doi:10.1002/sim.4780111607.
25. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. *Akademia Kiado: Budapest*; 1973. p. 267–281.
26. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
27. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361–87. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4. Review.
28. Hosmer DW, Lemeshow S. *Applied survival analysis: regression modelling of time to event data*. New York: Wiley; 1999.
29. Vergouwe Y, Steyerberg EW, Eijkemans MJ, et al. Validity of prognostic models: when is a model clinically useful? *Semin Urol Oncol*. 2002;20:96–107. doi:10.1053/suro.2002.32521.
30. Schemper M, Henderson R. Predictive accuracy and explained variation in Cox regression. *Biometrics*. 2000;56:249–55. doi:10.1111/j.0006-341X.2000.00249.x.
31. Reiso H, Nygard JF, Jorgensen GS, et al. Back to work: predictors of return to work among patients with back disorders certified as sick: a two-year follow-up study. *Spine*. 2003;28:1468–73. doi:10.1097/00007632-200307010-00020.
32. Fritz JM, George SZ, Delitto A. The role of fear-avoidance beliefs in acute low back pain: relationships with current and future disability and work status. *Pain*. 2001;94:7–15. doi:10.1016/S0304-3959(01)00333-5.
33. Thomas E, Silman AJ, Croft PR, et al. Predicting who develops chronic low back pain in primary care: a prospective study. *BMJ (Clinical Research Ed.)*. 1999;318:1662–7.
34. Von Korf M, Ormel J, Keefe FJ, et al. Grading the severity of chronic pain. *Pain*. 1992;50:133–49. doi:10.1016/0304-3959(92)90154-4.
35. Strazdins L, Bammer G. Women, work and musculoskeletal health. *Soc Sci Med*. 2004;58:997–1005. doi:10.1016/S0277-9536(03)00260-0.
36. Hoogendoorn WE, Bongers PM, de Vet HC, et al. High physical work load and low job satisfaction increase the risk of sickness absence due to low back pain: results of a prospective cohort study. *Occup Environ Med*. 2002;59:323–8. doi:10.1136/oem.59.5.323.
37. Lötters F, Burdorf A. Prognostic factors for duration of sickness absence due to musculoskeletal disorders. *Clin J Pain*. 2006;22:212–21. doi:10.1097/01.ajp.0000154047.30155.72.

38. van den Heuvel SG, Ariens GA, Boshuizen HC, et al. Prognostic factors related to recurrent low-back pain and sickness absence. *Scand J Work Environ Health*. 2004;30:459–67.
39. Heymans MW, de Vet HC, Knol DL, et al. Workers' beliefs and expectations affect return to work over 12 months. *J Occup Rehabil*. 2006;16:685–95. doi:[10.1007/s10926-006-9058-8](https://doi.org/10.1007/s10926-006-9058-8).
40. Kuijer W, Groothoff JW, Brouwer S, et al. Prediction of sickness absence in patients with chronic low back pain: a systematic review. *J Occup Rehabil*. 2006;16:439–67. Review.
41. Pransky GS, Verma SK, Okurowski L, et al. Length of disability prognosis in acute occupational low back pain: development and testing of a practical approach. *Spine*. 2006;31:690–7. doi:[10.1097/01.brs.0000202761.20896.02](https://doi.org/10.1097/01.brs.0000202761.20896.02).
42. Dionne CE, Bourbonnais R, Fremont P, et al. A clinical return-to-work rule for patients with back pain. *Can Med Assoc J*. 2005;172:1559–67. doi:[10.1503/cmaj.1041159](https://doi.org/10.1503/cmaj.1041159).
43. Heymans MW, Ford JJ, McMeeken JM, Chan A, de Vet HC, van Mechelen W. Exploring the contribution of patient-reported and clinician based variables for the prediction of low back work status. *J Occup Rehabil*. 2007;17:383–97. doi:[10.1007/s10926-007-9084-1](https://doi.org/10.1007/s10926-007-9084-1).
44. Steenstra IA, Koopman FS, Knol DL, Kat E, Bongers PM, de Vet HC, et al. Prognostic factors for duration of sick leave due to low-back pain in dutch health care professionals. *J Occup Rehabil*. 2005;15:591–605. doi:[10.1007/s10926-005-8037-9](https://doi.org/10.1007/s10926-005-8037-9).
45. Nordqvist C, Holmqvist C, Alexanderson K. Views of laypersons on the role employers play in return to work when sick-listed. *J Occup Rehabil*. 2003;13:11–20. doi:[10.1023/A:1021841722613](https://doi.org/10.1023/A:1021841722613).
46. Schemper M. Predictive accuracy and explained variation. *Stat Med*. 2003;22:2299–308. doi:[10.1002/sim.1486](https://doi.org/10.1002/sim.1486).
47. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol*. 2003;56:826–32. doi:[10.1016/S0895-4356\(03\)00207-5](https://doi.org/10.1016/S0895-4356(03)00207-5).
48. Steyerberg EW, Harrell FE Jr, Borsboom GJ, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54:774–81. doi:[10.1016/S0895-4356\(01\)00341-9](https://doi.org/10.1016/S0895-4356(01)00341-9).