

Exploring the Contribution of Patient-Reported and Clinician Based Variables for the Prediction of Low Back Work Status

Martijn W. Heymans · Jon J. Ford · Joan M. McMeeken ·
Alexander Chan · Henrica C. W. de Vet · Willem van Mechelen

Received: 20 November 2006 / Accepted: 27 March 2007 / Published online: 15 June 2007
© Springer Science+Business Media, LLC 2007

Abstract

Background Successful management of workers on sick leave due to low back pain by the general physician and physiotherapist depends on reliable prognostic information on the course of low back pain and work resumption.

Methods Retrospective cohort study in 194 patients who were compensated because of chronic low back pain and who were treated by a physiotherapy functional restoration program. Patient-reported and clinician based prognostic indicators were assessed at baseline before patients entered the functional restoration program. We investigated the predictive value of these indicators on work status at 6 months. Relationships were studied using logistic regression analysis in a 2-step bootstrap modelling approach and a nomogram was developed. Discrimination and calibration of the nomogram was evaluated internally and the explained variation of the nomogram calculated.

Results Seventy percent of workers were back to work at 6 months. We found that including duration of complaints, functional disability, disc herniation and fear avoidance beliefs resulted in the “best” prognostic model. All these factors delayed work resumption. This model was used to construct a nomogram. The explained variation of the nomogram was 23.7%. Discrimination was estimated by the area under

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

M. W. Heymans (✉) · W. van Mechelen
Department of Public and Occupational Health, VU University Medical
Center EMGO-Institute, Van der Boechorststraat 7, Amsterdam 1081 BT, The Netherlands
e-mail: mw.heymans@vumc.nl

M. W. Heymans · H. C. W. de Vet · W. van Mechelen
Institute for Research in Extramural Medicine, VU University Medical Center,
Amsterdam, The Netherlands

J. J. Ford · J. M. McMeeken · A. Chan
School of Physiotherapy, The University of Melbourne, Melbourne, Australia

the receiver operating characteristic curve and was 0.76 and for calibration we used the slope estimate that was 0.91. The positive predictive values of the nomogram at different cut-off levels of predicted probability were good.

Conclusions Knowledge of the predictive value of these indicators by physicians and physiotherapists will help to identify subgroups of patients and will thus enhance clinical decision-making.

Keywords Low back pain · Work status · Prediction · Nomogram

Introduction

Low back pain (LBP) is the most common and expensive musculoskeletal disorder in Western countries [1]. Although most patients recover from LBP within the first 2 months about 10% will develop chronic LBP [2]. The recovery process of persons with chronic LBP is slow, and their demands on the health care system are both large and costly. Total costs in Australia and the Netherlands were estimated yearly to be 400 million Euro in 1993–1994 and 4 billion Euro in 1991, respectively [3, 4]. Urgent identification of those patients for whom treatment is warranted is necessary to protect them from prolonged disability, sick-leave and medical over consumption.

Studies evaluating the ability of single or combined criteria to predict change in work status following intervention can be useful in identifying patient subgroups that need more intensive interventions because of their poor prognosis. Evidence exists for the predictive value of patient-reported LBP indicators such as low back symptom duration, severity of pain, disability and fear avoidance beliefs on treatment outcome [5–7]. However, in general these prognostic models have only been able to explain a relatively small amount of variance in work status change [8]. Other studies have shown that patient examination findings of lumbar segmental mobility, lumbar range of motion and hip rotation can predict treatment outcome [9–11]. It is feasible that the inclusion of both patient-report and examination findings would increase the prognostic ability of existing models, and therefore increase their value for clinicians and researchers.

Disc herniation with associated radiculopathy (DHR) is a category of LBP that has accepted diagnostic criteria [12]. Significant literature exists supporting the validity of criteria, particularly regarding response to repeated movements that may be indicative of discogenic pain [13]. However, mixed results exist concerning the prognostic value of the straight leg raising and Waddell's nonorganic signs test [14–16]. There is no study that has evaluated the ability of clinician based pathoanatomical subgroups in combination with existing patient-reported criteria to predict change in work status in response to conservative treatment.

Systematic reviews have concluded that multi-disciplinary and physiotherapy functional restoration programs are most effective in improving pain, disability and work related outcome measures for chronic LBP [17, 18]. Identifying subgroups of patients that are more or less responsive to functional restoration can be a successful method of enhancing treatment effects and to guide clinical decision-making [19].

The aim of this study was to investigate the predictive value of patient-reported indicators as well as clinician-based subgroups on work status in patients following a

physiotherapy functional restoration (PFR) program. Relevant indicators were used to construct a nomogram to ease practical application.

Methods

The analyses in this study were conducted with retrospective data from patients who were referred by a general physician, rehabilitation provider, employer or medical specialist in Victoria, Australia because of sick-leave due to LBP in the period between 2002 and 2004. As part of standard clinical practice all referred patients routinely completed a comprehensive subjective assessment and clinical examination as well as standardised questionnaires. A standard quality control process aggregated all data into a central database which was used in this study.

Study Population

Patients were included in the study if they had a primary problem of mechanical LBP, were aged between 18 and 65, had a compensation claim relating to their LBP and had some degree of work disability. All patients were referred to undergo a PFR program of between 4–8 weeks duration. As part of standard clinical procedure, no patient was accepted if they were pregnant, used narcotic analgesics, or had any serious and/or specific medical condition requiring further investigation or treatment, e.g., cancer, infection or other systematic non-musculoskeletal disease. The study comprised 194 patients.

Treatment

Treatment consisted of a PFR program conducted three times a week for 4–8 consecutive weeks. All physiotherapists practicing PFR had completed a 2 days training course and had been assigned to a mentor program to ensure competency and consistency in the treatment of chronic compensable LBP patients. The PFR program was structured around progressive functional aerobic and resistance exercises. These exercises were all performed in a graded manner in conjunction with a precise contraction of transverses abdominis and lumbar multifidus in an effort to functionally retrain the stabilisers of the lumbar spine. All exercises were administered using a cognitive behavioural approach. The cognitive-behavioural approach included emphasizing the relative benefits of active exercise and self-management as opposed to passive treatment. Functional goals relating to activities of daily living, including work, and the exercise program were set and the patients were encouraged to maintain focus on the goals throughout the program. When progress towards the goals was achieved, i.e., exercise intensity was increased, positive feedback was given to the patient to reinforce wellness behaviour. Every effort was made not to reinforce excessive illness behaviour [20].

Prognostic Factors

The prognostic factors in our study were classified into 2 categories: the patient-reported prognostic indicators and the clinician based subgroups. The patient-reported prognostic

indicators were assessed by means of questionnaires completed at initial assessment before the commencement of the PFR program. Clinician based subgroup membership was based on combinations of subjective assessment and examination criteria also obtained at initial assessment.

Patient-reported Prognostic Indicators

We included the indicators that were most frequently identified as having a relationship with work status on basis of a literature review and matched these with the indicators present in our data set [5, 6, 21, 22]. The following indicators were considered: gender (male / female); age; the duration of complaints prior to commencement of a PFR program, pain radiation into both legs (yes / no), and functional disability measured by means of the Oswestry Disability Questionnaire (ODQ) [23]. The severity of pain was measured with the pain intensity subscale of the ODQ. Fear of movement, avoidance of activities and back pain beliefs were measured with the Fear Avoidance Beliefs Questionnaire [24].

Clinician Based Prognostic Indicators

To investigate the influence of clinician-based indicators on work status we used three main subgroups of criteria. If patients fulfilled all criteria of a specific subgroup they were assigned a score of “1”, otherwise a score of “0”. The subgroup criteria were:

1. Disc herniation with associated radiculopathy (DHR): herniation or extrusion observed on magnetic resonance imaging (MRI); unilateral pain; unilateral paraesthesia or pain below the knee in one leg; and a straight leg raise discrepancy of at least 15 degrees between legs [25].
2. Discogenic pain responsive to repeated movement (DRRM): lumbar or leg pain eased with walking and pain worse with sitting and forward bending. Absence of criteria related to DHR [26].
3. Discogenic pain unresponsive to repeated movement (DURM): lumbar or leg pain worse with walking, standing, sitting and forward bending; active lumbar flexion less than reaching patella with hands, active lumbar extension no more than 20°. Absence of criteria related to DHR and DRRM [27].

We also explored the association of a positive straight leg raising test in isolation, Waddell’s non-organic signs [28] and clinically determined inflammation. For each of these criteria patients were assigned a score of “1” or “0” if they were positive or negative on that criterion respectively. Patients were positive on the straight leg raising test if they experienced pain in their leg at 60 degrees of flexion in their hip. A positive non-organic score for the purpose of this study was defined as the presence of at least four of the seven signs [29, 30]. There are no validated methods for clinically determining the presence or absence of inflammation in LBP. However features described in the rheumatological literature and commonly in clinical use include: constancy of pain, night pain and morning stiffness of at least 60 min [31]. The patient was deemed to have clinically determined inflammatory LBP if two or more of these criteria were positive.

Outcome

The outcome variable work status was defined as being at work at 6 months follow-up after completion of the PFR program. Patients defined as being at work could be: working full time, part time and on modified duties, because of the LBP they experienced.

Statistical Analysis

Work status at 6 months was included into multivariable logistic regression models as the dependent variable and the potential variables as independent variables. A potential nonlinear behaviour of the continuous indicators with the outcome was examined by using restricted cubic spline functions and spline plots [32]. Restricted cubic spline functions allow continuous indicators to be fitted within the regression model without assuming a linear relation. We did not find a nonlinear relation for any continuous indicator and therefore did not have to include spline functions or other indicator transformations.

To fill in the variables with missing values we applied multiple imputations by using the Multiple Imputation by Chained Equation (MICE) package [33]. 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. Logistic regression is used to impute incomplete dichotomous variables, linear regression to impute continuous variables. We generated five multiply imputed data sets. These data sets were included into a two-step model building process as described beneath.

Model Building

Indicator selection was performed by a 2-step bootstrap model averaging approach proposed by Holländer et al. [34] and Sauerbrei and Schumacher [35]. During the first step of this procedure, bootstrap samples of equal sample size as the original sample were drawn with replacement from the original data set. For indicator selection, backward regression analyses were applied on each bootstrap sample with a P -value of 0.157. A P -value of 0.157 corresponds to using Akaike's Information Criterion (AIC) for variable selection. AIC is much used in variable selection and has good theoretical and statistical properties [36]. The inclusion frequency of each indicator was evaluated by counting the number of times that each indicator was retained in the regression model and by dividing this number by the number of bootstrap samples drawn. On the basis of their inclusion frequencies, prognostic indicators were included into a second modelling step. Holländer et al. [34] recommend omitting indicators for further analyses, which are selected in less than 30% of the regression models. We compared this level with an exclusion frequency level of 20%. Likewise as in the first step, backward regression analyses were performed on each of the bootstrap samples in a second modelling step. In this step the indicators that were selected in more than 20% and 30% of the regression models were used. Next the selection frequency of the multivariable models that appeared in the bootstrap samples was considered. This was done by counting the number of times that a regression model that consisted of the same prognostic indicators appeared in the bootstrap samples and dividing this number by the

number of bootstrap samples drawn. The final “best” model was chosen on the basis of a high model selection frequency, model simplicity and most accurate model performance (see under model performance).

The logistic regression coefficients and standard errors (SE) of the final “best” model were estimated on each imputed data set and then averaged over the five data sets using Rubin’s rules. In this procedure 95% confidence intervals are calculated by taking into account the variance between the imputed data sets [37].

During the modelling process we considered also the balance between the number of variables and events in the models, which is recommended not to be lower than 10 events per variable [38].

To make the risk prediction available in clinical practice we transformed the final “best” model into a nomogram [39]. The nomogram and an explanation of how to use the nomogram can be found in the Appendix.

Model Performance

The explained variation of the multivariable model was estimated according to the (Nagelkerke’s) R^2 [40]. Discrimination was evaluated by the Area Under the receiver operating characteristic Curve (AUC). Discrimination can be interpreted as how well the model distinguishes between patients who have a higher risk of returning to work from patients who have a low risk [41]. The clinical characteristics of the prediction model were also evaluated in terms of sensitivity, specificity and positive and negative predictive values at different cut-off levels of predicted probability. The calibration of the models was considered by calculating the slope index. Calibration refers to the agreement between the observed probabilities in the original data and the predicted probabilities of returning to work, estimated by the nomogram. The slope indicates the statistical overoptimism of the model. When a slope is <1 , low predictions may be too low and high predictions may be too high [41]. Performance measures were estimated by using the regression coefficients that were averaged over the five imputed data sets. Model performance measures were also considered in choosing the final “best” model.

Software

The MICE [33] as well as the backward selection procedures were performed with S-Plus software (version 2000). We developed additional software for S-plus to perform the two-step bootstrap selection approach. Evaluating of the performance of the models was done with the Design Library [42].

Results

At 6 months post PFR program complete data on work status was available for all 194 patients. All were included in the multivariable regression analyses. Their baseline characteristics are presented in Table 1. Most of the prognostic indicators had complete data, however, the indicators, disc herniation (DHR), discogenic pain responsive (DRRM) and unresponsive (DURM) to repeated movement had 15% missing values. In the study sample, 70% had returned to work at 6 months follow-up.

Table 1 Patient characteristics at baseline (*n* = 194)

Patient-reported prognostic indicators:	Value	Missing (%)
Age (mean years ± <i>SD</i>)	41.8 (9.9)	0.0
Gender (male) (%)	33.5	0.0
Pain intensity (mean ± <i>SD</i>)	2.5 (1.5)	4.0
Pain radiation (%)	44.8	0.0
Functional disability (mean ± <i>SD</i>)	45.0 (15.2)	1.0
Fear avoidance beliefs (mean ± <i>SD</i>)	46.9 (9.4)	0.0
Duration of complaints (months, mean ± <i>SD</i>)	21.4 (40.2)	0.0
Clinically evaluated prognostic indicators:		
Positive on straight leg raising test (%)	43.8	0.1
A positive non-organic score (%)	9.3	0.0
Disc herniation (DHR) (%)	5.5	14.9
Discogenic pain responsive to repeated movement (DRRM) (%)	12.2	15.0
Discogenic pain unresponsive to repeated movement (DURM) (%)	24.8	14.9
Inflammation (%)	50.0	0.0

Table 2 shows the results of the two-step bootstrap model averaging approach. The inclusion frequencies of the indicators at this first step ranged between 4.1% and 74.2%. The highest inclusion frequencies, i.e., in >30% of the regression models, were found for the indicators duration of complaints, functional disability, disc herniation and fear avoidance beliefs.

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

Inclusion frequency	Step 1 %	Models selected in Step 2 (only the first 10 models are shown)									
		1	2	3	4	5	6	7	8	9	10
1 Duration of complaints	74.2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2 Functional disability	61.1	✓	✓	✓	✓	✓	✓	–	✓	–	✓
3 DHR	55.7	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
4 Fear avoidance beliefs	48.6	✓	✓	✓	–	–	–	✓	✓	✓	✓
5 DURM	33.8	–	✓	–	✓	–	✓	✓	✓	✓	–
6 Straight leg raising	20.0	–	–	✓	–	✓	✓	✓	✓	–	–
7 Pain intensity	11.2	–	–	–	–	–	–	–	–	–	–
8 DRRM	10.6	–	–	–	–	–	–	–	–	–	–
9 Pain radiation	10.1	–	–	–	–	–	–	–	–	–	–
10 No sign test	6.5	–	–	–	–	–	–	–	–	–	–
11 Age	5.3	–	–	–	–	–	–	–	–	–	–
12 Gender (male)	4.5	–	–	–	–	–	–	–	–	–	–
13 Inflammation	4.1	–	–	–	–	–	–	–	–	–	–
Model selection frequency% ^a		25.4	26.2	#	15.5	#	#	#	#	8.1	5.2
Model selection frequency% ^b		15.2	13.5	12.1	8.9	5.9	5.2	5.1	5.1	4.2	2.9

✓ Indicators included

^a Selected models with indicators included >30% at step 1 (total number of models is 22)

^b Selected models with indicators included >20% at step 1 (total number of models is 42)

–: Not selected

This model with indicators included >30% at step 1 can not be chosen due to the indicator straight leg raising which was selected in 20% of the bootstrap samples at step 1

In modelling step two five indicators with an inclusion frequency of more than 30% were included and eight indicators were excluded. In this step 22 different models were selected. As seen from Table 2, the model with all five indicators included reached the highest selection frequency of 26.2%. We repeated the analyses with indicators who had an inclusion frequency of >20% at step 1. Now 42 different models were selected at step 2. The model with the indicators duration of complaints, functional disability, disc herniation and fear avoidance beliefs was selected most of the time, i.e., 15.2%. The model with the highest selection frequency of 26.2% in the previous analyses now obtained a selection frequency of 13.5%. On the basis of this selection procedure the model performance of models 1 and 2 was compared.

The multivariable models 1 and 2 are presented in Table 3. This table also reports the values for the R^2 , slope and AUC. All indicators for model 1 had a delayed effect on work resumption at 6 months and showed a statistically significant relationship with the outcome.

The strongest effect was found for disc herniation. The multivariable model 1 had a R^2 of 23.7% and an AUC of 0.76. The slope of the model was 0.91. For model 2, where the indicator DURM was added the R^2 and the AUC were similar. The slope for model 2 slightly decreased 0.88 compared to a slope value of 0.91 for model 1. Because of its practical simplicity under the same model performance, model 1 was selected for the construction of a nomogram (see Appendix).

Table 4 shows the diagnostic characteristics of the multivariable prediction model according to the values of sensitivity, specificity and positive and negative predictivity at different cut-off levels of predicted probability. The first row of Table 4 shows that sensitivity is 100% because RTW will eventually be identified in all patients. At this level the PPV or prevalence of RTW is 70%. The PPV increases with increasing levels of predicted probability but the number of patients that RTW decrease. At a probability level of RTW of 80% or higher, 90% of the patients will RTW. However, 10% of the patients will not RTW (1–PPV) but will not receive additional interventions due to their high probability of RTW.

Table 3 Presentation of factors included in the multivariable model together with the R^2 , slope estimate and Area Under the ROC Curve (AUC) ($n = 194$)

	Multivariable model 1		Multivariable model 2	
	Odds Ratio	95% CI	Odds Ratio	95% CI
<i>Patient-reported indicators</i>				
Duration of complaints	0.98	0.97–0.99	0.98	0.97–0.99
Functional Disability	0.96	0.93–0.99	0.96	0.94–0.99
Fear avoidance beliefs	0.95	0.90–0.98	0.95	0.91–0.99
<i>Clinician based indicator</i>				
Disc herniation	0.14	0.02–0.78	0.12	0.02–0.70
Discogenic pain unresponsive to repeated movement (DURM)			0.51	0.22–1.18
	Model performance		Model performance	
Nagelkerke R^2 (%)	23.7		24.0	
Slope	0.91		0.88	
Area Under the Curve (AUC)	0.76		0.76	

Table 4 Clinical characteristics of the prediction model

PB*	Sens ^a (%)	Spec ^b (%)	PPV ^c (%)	NPV ^d (%)	% of patients that RTW
≥ 0%	100	0	70	100	100
≥ 10%	97	28	76	80	97
≥ 20%	88	40	77	59	96
≥ 30%	81	55	81	55	95
≥ 40%	70	64	82	47	92
≥ 50%	62	78	87	46	84
≥ 60%	51	86	90	43	70
≥ 70%	38	88	88	38	58
≥ 80%	26	93	90	35	44
≥ 90%	14	98	95	33	13

*Predicted probability of RTW.

^a Sensitivity; ^bSpecificity; ^cPositive predictive value; ^dNegative predictive value.

Discussion

The most important indicators predictive for work resumption at 6 months following a physiotherapy functional restoration (PFR) program were a shorter duration of complaints, better functional ability, no disc herniation and less fear avoidance beliefs. A nomogram was developed that can easily be used in daily practice to identify individual patients who would be likely to return to work following a PFR program. The explained variation of the nomogram was 23.7% and the discriminative and calibrative abilities measured by the AUC and slope indices were 0.76 and 0.91, respectively. Furthermore, the nomogram showed a good PPV at different cut-off levels of predicted probability.

Pain-related fear of activity is an important factor in relation to the course of LBP and sick leave due to LBP [43]. We found in our study that patients, undergoing a PFR, with higher fear avoidance beliefs were less likely to return to work at 6 months. The findings in our study are supported by the study of Fritz et al. [44] who also found a relation between fear avoidance beliefs and treatment success in a multivariable clinical prediction rule designed for a specific treatment. Therefore, we recommend including fear avoidance beliefs in future prognostic studies to confirm the predictive value of this indicator on longer-term sick-leave due to LBP.

Our findings on the duration of complaints as important prognostic factors for RTW are in agreement with that found in several other reviews [45, 46]. The duration of complaints can be related to the severity of the LBP, which in turn may be responsible for longer work absence. Pain severity is related to pain intensity, functional limitation and pain duration as proposed by von Korff. Von Korff showed that the back pain does not have to be present all the time. The pain may be present at a lower level of pain intensity in the background and may flare-up. Flare-ups are frequently seen in chronic low back pain patients and are in combination with higher levels of functional disability responsible for higher pain severity and consequently longer work absence [47].

We found that greater functional disability contributes to a delay in return to work at 6 months. Functional disability can be a proxy for limitations of daily activities including work. This is consistent with the findings of Truchon [48] that people who are not able to work also have higher levels of function disability.

In our study, the clinical diagnosis of disc herniation, as part of clinician based indicators, was negatively related to work status at 6 months, i.e., if you have a disc herniation you are less likely to be back at work. The criteria used to define disc herniation are well established [49]. There is significant evidence supporting the lumbar disc as a potential pain generator through pathological processes involving posterior migration of the nucleus pulposus and irritation of free nerve endings in the posterior annulus [49]. Our study confirmed that lumbar disc herniated patients have a worse prognosis concerning later work resumption.

We did not find a significant contribution to the prediction of work status of most other clinician-based indicators. This suggests that these indicators, i.e., DRRM, straight leg raising and nonorganic sign test, may require justification. In a recent systematic review it was concluded that the clinical significance of a positive straight leg raising test in LBP patients is still unclear [14]. Polatin et al. [15] did not find that nonorganic signs were predictive for treatment success in chronic LBP patients following a PFR program. This was in line with our study results. Although Gaines et al. [16] found that nonorganic signs were of significant value in predicting a delayed return to work, they included acute LBP patients.

The criteria for DRRM in our study were based on symptoms being aggravated with lumbar flexion activities and improved by extension activities such as walking and standing [31]. Response to repeated movements, activities and positions has been subject to validation studies [13]. However, some LBP cases that are aggravated by flexion activities are also actually aggravated by extension activities [50] and in our study, this group was defined as DURM. This category is based on the hypothesised mechanism of extension activity irritating an inflamed posterior annulus fibrosis [27, 51]. This may be an explanation of the presence of DURM on predicting longer RTW in the final multivariable model 2, compared to DRRM which is not retained in this model.

Missing data were substituted by applying Multiple Imputation (MI). MI accounts for the uncertainty caused by the missing data, and when properly done, MI provides correct statistical inferences [37]. In contrast to naïve missing data techniques, as mean or single imputation, that cause bias, MI replaces each missing value by more than one imputation. We used five imputed data sets, which are enough to generate proper imputations [37]. The spread between the imputed values reflects the uncertainty about the missing data.

In our study we used a new and promising two-step bootstrap model averaging procedure for model building [52]. By applying multiple imputations we were also able to include the information of all patients in our study [37]. With the two-step procedure we accounted for model instability in the selection process. Austin and Tu [53] and Sauerbrei and Schumacher [35] have shown that regression models are subject to variability. A model with 13 predictors, the number of indicators we started with in our analyses, can result in 2^{13} (=8192) different models. Furthermore, by the two-step bootstrap model averaging approach as proposed, variables which have, no or only a weak effect on the outcome can be deleted from further analyses. This reduces the number of all possible models in the next step and makes it easier to find the model that is best supported by the data.

Some limitations in our study have to be considered. Our definition for successful return to work at 6 months after completion of PFR also included patients who were working part time or on modified duties due to the LBP. This means that as part of the return to work process, some patients who were at work (part time or on modified duties) could go on sick-leave completely after the 6 months follow-up period. This

could have influenced the prognostic value of some indicators in our study. It may be better to use a definition for lasting return to work, e.g., return to work to the previous job for at least 4 weeks or more. However, studies have shown that most people who return to work for at least 1 day maintain work for a longer period [8]. Not all patients who were included into the DHR category received a MRI. Conform daily practice, this decision may depend on the clinician's judgment on basis of the patient's symptoms and / or patients demands. This may have led to misclassification of some DHR patients. Possibly more patients could have classified as DHR, which were not DHR. Misclassification usually leads to bias towards the null, which means that the strong predictive value of this variable may have been underestimated. Another limitation is the use of retrospective data in our analyses. It is well-known that retrospective data is of lower quality compared to prospective data, because of the risk of information bias. All predictive information in our study was systematically assessed prior to the interventions as part of a standardized physiotherapy assessment protocol that was routinely used for all patients and from medical records of the patients. Our study may be subject to information bias, dependent on the quality of information in the medical records.

The DRRM subgroup may have exhibited a greater predictive power had a more formal physical examination of therapeutic loading strategies been made [30], and future prospective studies should incorporate such protocols. However, our intention was to explore whether patients exhibiting the features most commonly accepted to be indicative of DRRM (such as response to flexion and extension activities) and therefore we believe the criteria chosen were appropriate. The criteria for the DURM subgroup were based on clinical experience and interpretation of the literature by the authors. Whilst there is limited evidence regarding the clinical features of this subgroup [49] we believe there is sufficient biological plausibility for the criteria selected.

In our opinion this is the first study that determines which combination of patient-reported and clinician based indicators predicts outcome to an evidence-based treatment; in this case a PFR program. The findings in our study are of interest for other researchers and clinicians because the construction of prediction aids in LBP research is still underdeveloped. Variables from both data sources, i.e., clinician based and patient-reported variables, were retained in the final model. Therefore, the main message of our study is that both sources of information are important to predict RTW. We recommend that information from clinical examination and patient-reported information has to be considered in future research and practice with respect to RTW prognosis. We are aware that the methods used in this study need to be applied in well-controlled prospective studies to more conclusively validate the nomogram. Based on such future research, the nomogram could then be broadly implemented in clinical practice and used in future randomised controlled trials to allow for more appropriate patient selection and treatment application.

Conclusion

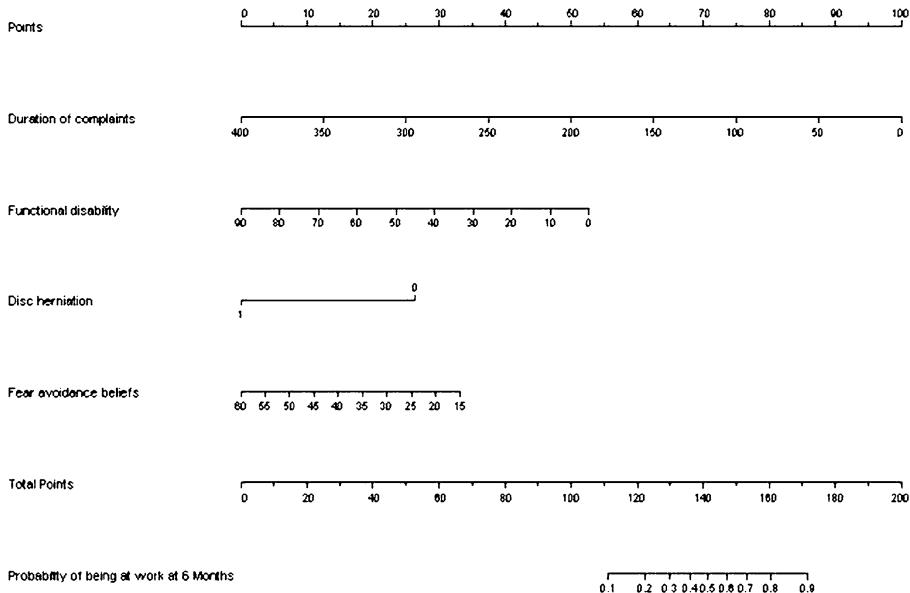
This study presented a predictive model to identify patients who will likely RTW after being treated by a functional restoration program. This predictive model was transformed into a nomogram consisting of patient-reported and clinician-based prognostic indicators that has potential for use in daily practice and clinical research

as a means of identifying subgroups of patients for which a more specific treatment approach is required. Our results need to be replicated in prospective studies and considered when planning future outcome studies.

Appendix

Nomogram Predicting the Probability of Being at Work at 6 Months Follow-up

Instructions for physical therapists to use the nomogram are: first locate the patients' score for an indicator, e.g., a score of 50 for functional disability, on the corresponding horizontal "Functional disability" axis. Then draw a vertical line straight upward from that patient's score on functional disability to the "Points axis" on top of the nomogram. A score of 50 points for disability corresponds to 25 points. Repeat this process for the other indicators, each time drawing a line straight upward to the Points axis and sum the points achieved by each indicator. Find this total score on the "Total Points axis" (last row of the nomogram). Finally, draw a vertical line straight down from the "Total Points axis" to the "Probability of work status at 6 months" axis to find the patient's probability of being at work at 6 months. For example a patient with a duration of low back pain complaints of 125 months (68 points), a functional disability score of 40 (30 points), no disc herniation (25 points) and a fear avoidance score of 55 (3 points) has a probability of around 25% (126 total points) to be at work at 6 months following treatment by a functional restoration program.



References

- Andersson, G. B. J. (1999). Epidemiological features of chronic low-back pain. *Lancet*, *354*, 581–585.
- Frank, J. W., Brooker, A. S., DeMaio, S. E., Kerr, M. S., Maetzel, A., & Shannon, H. S., et al. (1996). Disability resulting from occupational low back pain: Part II: What do we know about secondary prevention? A review of the scientific evidence on prevention after disability begins. *Spine*, *21*, 2918–2929.
- Mathers, C., & Penn, R. (1999). Health system costs of injury, poisoning and musculo-skeletal disorders in Australia 1993–1994. Canberra: Australian Institute of Health and Welfare. AIHW Catalogue No. HWE 12 (Health and Welfare Expenditure Series No. 6). Australia.
- van Tulder, M. W., Koes, B. W., & Bouter, L. M. (1995). A cost-of-illness study of back pain in the Netherlands. *Pain*, *62*, 233–240.
- van der Hulst, M., Vollenbroek-Hutten, M. M., & Ijzerman, M. J. (1995). A systematic review of sociodemographic, physical, and psychological predictors of multidisciplinary rehabilitation-or, back school treatment outcome in patients with chronic low back pain. *Spine*, *30*(7), 813–825. Review.
- Leboeuf-Yde, C., Gronstvedt, A., Borge, J. A., Lothe, J., Magnesen, E., Nilsson, O., Rosok, G., Stig, L. C., & Larsen, K. (2004). The nordic back pain subpopulation program: Demographic and clinical predictors for outcome in patients receiving chiropractic treatment for persistent low back pain. *Journal of Manipulative and Physiological Therapeutics*, *27*(8), 493–502.
- Flynn, T., Fritz, J., Whitman, J., Wainner, R., Magel, J., Rendeiro, D., Butler, B., Garber, M., & Allison, S. (2002). A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. *Spine*, *27*(24), 2835–2843.
- Heymans, M. W., de Vet, H. C., Knol, D. L., Bongers, P. M., Koes, B. W., & van Mechelen, W. (2006). Workers' beliefs and expectations affect return to work over 12 months. *Journal of Occupational Rehabilitation*, *16*(4), 685–695.
- Borge, J. A., Leboeuf-Yde, C., & Lothe, J. (2001). Prognostic values of physical examination findings in patients with chronic low back pain treated conservatively: A systematic literature review. *Journal of Manipulative and Physiological Therapeutics*, *24*(4), 292–295. Review.
- Hunt, D. G., Zuberbier, O. A., Kozlowski, A. J., Berkowitz, J., Schultz, I. Z., Milner, R. A., Crook, J. M., & Turk, D. C. (2002). Are components of a comprehensive medical assessment predictive of work disability after an episode of occupational low back trouble? *Spine*, *27*(23), 2715–2719.
- Childs, J. D., Fritz, J. M., Flynn, T. W., Irrgang, J. J., Johnson, K. K., Majkowski, G. R., & Delitto, A. (2004). A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: A validation study. *Annals of Internal Medicine*, *141*(12), 920–928.
- Andersson, G. B., & Deyo, R. A. (1996). History and physical examination in patients with herniated lumbar discs. *Spine*, *21*(24 Suppl):10S–18S. Review.
- Wetzel, F. T., & Donelson, R. (2003). The role of repeated end-range/pain response assessment in the management of symptomatic lumbar discs. *The Spine Journal*, *3*(2), 146–154.
- Rebain, R., Baxter, G. D., & McDonough, S. (2002). A systematic review of the passive straight leg raising test as a diagnostic aid for low back pain (1989 to 2000). *Spine*, *27*(17), E388–E395. Review.
- Polatin, P. B., Cox, B., Gatchel, R. J., & Mayer, T. G. (1997). A prospective study of Waddell signs in patients with chronic low back pain. When they may not be predictive. *Spine*, *22*(14), 1618–1621.
- Gaines, W. G. Jr, & Hegmann, K. T. (1999). Effectiveness of Waddell's nonorganic signs in predicting a delayed return to regular work in patients experiencing acute occupational low back pain. *Spine*, *24*(4), 396–400.
- Guzman, J., Esmail, R., Karjalainen, K., Malmivaara, A., Irvin, E., & Bombardier, C. (2002). Multidisciplinary bio-psycho-social rehabilitation for chronic low back pain. *Cochrane Database of Systematic Reviews* (1), CD000963. Review.
- Schonstein, E., Kenny, D. T., Keating, J., & Koes, B. W. (2003). Work conditioning, work hardening and functional restoration for workers with back and neck pain. *Cochrane Database of Systematic Reviews*, (1), CD001822. Review.
- Reilly, B. M., & Evans, A. T. (2006). Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. *Annals of Internal Medicine*, *144*(3), 201–209.
- Hahne, A. J., & Ford, J. J. (2006). Functional restoration for a chronic lumbar disk extrusion with associated radiculopathy. *Physical Therapy*, *86*(12), 1668–1680.
- Crook, J., Milner, R., Schultz, I. Z., & Stringer, B. (2002). Determinants of occupational disability following a low back injury: A critical review of the literature. *Journal of Occupational Rehabilitation*, *12*, 277–295. Review.

22. Krause, N., Frank, J. W., Dasinger, L. K., Sullivan, T. J., & Sinclair, S. J. (2001). Determinants of duration of disability and return-to-work after work-related injury and illness: Challenges for future research. *American Journal of Industrial Medicine*, *40*, 464–484. Review.
23. Fairbank, J. C., Couper, J., Davies, J. B., & O'Brien, J. P. (1980). The Oswestry low back pain disability questionnaire. *Physiotherapy*, *66*(8), 271–273.
24. Waddell, G., Newton, M., Henderson, I., Somerville, D., & Main, C. J. (1993). A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain*, *52*, 157–168.
25. Andersson, G. B., Brown, M. D., Dvorak, J., Herzog, R. J., Kambin, P., Malter, A., McCulloch, J. A., Saal, J. A., Spratt, K. F., & Weinstein, J. N. (1996). Consensus summary of the diagnosis and treatment of lumbar disc herniation. *Spine*, *21*(24 Suppl), 75S–78S. Review.
26. Clare, H. A., Adams, R., & Maher, C. G. (2005). Reliability of McKenzie classification of patients with cervical or lumbar pain. *Journal of Manipulative and Physiological Therapeutics*, *28*(2), 122–127.
27. Shiraishi, T., & Crock, H. V. (1995). Re-exploration of the lumbar spine following simple discectomy: A review of 23 cases. *European Spine Journal*, *4*(2):84–87.
28. Waddell, G., McCulloch, J., Kummel, E., & Venner, R. (1980). Nonorganic physical signs in low back pain. *Spine*, *5*, 117–125.
29. Werneke, M. W., Harris, D. E., & Lichter, R. L. (1993). Clinical effectiveness of behavioral signs for screening chronic low-back pain patients in a work-oriented physical rehabilitation program. *Spine*, *18*(16), 2412–2418.
30. Foldes, K., Balint, P., Gaal, M., Buchanan, W. W., & Balint, G. P. (1992). Nocturnal pain correlates with effusions in diseased hips. *Journal of Rheumatology*, *19*(11), 1756–1758.
31. Willams, M., & McKenzie, R. A. (1993). Patient classification. *Physical Therapeutics*, *73*(8), 550–551.
32. Harrell, F. E. Jr. (2001). *Regression modeling strategies*. New York: Springer-Verlag.
33. Van Buuren, S. (2000). The Mice Library. <http://www.multiple-imputation.com>. Accessed.
34. Holländer, N., Augustin, N. H., & Sauerbrei, W. (2006). Investigation on the improvement of prediction by bootstrap model averaging. *Methods of Information in Medicine*, *45*(1), 44–50.
35. Sauerbrei, W., & Schumacher, M. (1992). A bootstrap resampling procedure for model building: Application to the Cox regression model. *Statistics in Medicine*, *11*(16), 2093–2109.
36. Steyerberg, E. W., Eijkemans, M. J., & Habbema, J. D. (1999). Stepwise selection in small data sets: A simulation study of bias in logistic regression analysis. *Journal of Clinical Epidemiology*, *52*(10), 935–942.
37. Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, John Wiley & Sons.
38. Peduzzi, P., Concato, J., Feinstein, A. R., & Holford, T. R. (1995). Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *Journal of Clinical Epidemiology*, *48*(12), 1503–1510.
39. Harrell, F. E. Jr, Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, *15*(4), 361–387. Review.
40. Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*, 691–692.
41. Vergouwe, Y., Steyerberg, E. W., Eijkemans, M. J., & Habbema, J. D. (2002) Validity of prognostic models: When is a model clinically useful? *Seminars in Urologic Oncology*, *20*, 96–107.
42. Harrell, F. E. (1997). Design: S-plus functions for biostatistical / epidemiological modeling, testing, estimation, validation, graphics, prediction and typesetting by storing enhanced model design attributes in the fit, Programs available at: <http://www.biostat.mc.vanderbilt.edu/twiki/bin/view/Main/FrankHarrell>.
43. Gheldof, E. L., Vinck, J., Vlaeyen, J. W., Hidding, A., & Crombez, G. (2005). The differential role of pain, work characteristics and pain-related fear in explaining back pain and sick leave in occupational settings. *Pain*, *113*(1–2), 71–81.
44. Fritz, J. M., & George, S. Z. (2002). Identifying psychosocial variables in patients with acute work-related low back pain: The importance of fear-avoidance beliefs. *Physical Therapy*, *82*, 973–983.
45. Steenstra, I. A., Verbeek, J. H. A. M, Heymans, M. W., & Bongers, P. M. (2005). Prognostic factors for return to work in patients sick listed with acute low-back pain, a systematic review of the literature. *OEM*, *62*, 851–860. Review.
46. Crook, J., Milner, R., Schultz, I. Z., & Stringer, B. (2002). Determinants of occupational disability following a low back injury: A critical review of the literature. *Journal of Occupational Rehabilitation*, *12*, 277–295. Review.
47. Von Korff, M., Ormel, J., Keefe, F. J., & Dworkin, S. F. (1992). Grading the severity of chronic pain. *Pain*, *50*(2), 133–149.

48. Truchon, M., & Fillion, L. (2000). Biopsychosocial determinants of chronic disability and low-back pain: A review. *Journal of Occupational Rehabilitation*, *10*, 117–142.
49. Bogduk, N. (1997). *Clinical anatomy of the lumbar spine and sacrum*. New York, Churchill Livingstone.
50. DiFabio, R. P., & Boissonault, W. (1998). Physical therapy and health-related outcomes for patients with common orthopaedic diagnoses. *Journal of Orthopaedic and Sports Physical Therapy*, *27*(3), 219–230.
51. Saal, J. S. (1995). The role of inflammation in lumbar pain. *Spine*, *20*(16), 1821–1827. Review.
52. Augustin, N. H., Sauerbrei, W., & Schumacher, M. (2005). The practical utility of incorporating model selection uncertainty into prognostic models for survival data. *Statistical Modelling*, *5*(2), 95–118.
53. Austin P. C., & Tu, J. V. (2004). Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *Journal of Clinical Epidemiology*, *57*(11), 1138–1146.