Eur Spine J (2002) 11 : 258–266
DOI 10.1007/s005860100335

ORIGINAL ARTICLE

J. P. Kool
P. R. Oesch
R. A. de Bie

Predictive tests for non-return to work in patients with chronic low back pain

Received: 11 July 2000
Revised: 20 July 2001
Accepted: 21 July 2001
Published online: 5 October 2001

This study was supported by a grant from Klinik Valens.

J.P. Kool (✉) · P.R. Oesch
Research Department,
Rehabilitation Centre, Klinik Valens,
7317 Valens, Switzerland
e-mail: j.kool@klinik-valens.ch,
Tel.: +41-81-3031453

R.A. de Bie
Department of Epidemiology,
Maastricht University,
6200 MD Maastricht, Netherlands

Abstract Return to work (RTW) is the primary goal in the rehabilitation of patients with chronic low back pain. In spite of expensive rehabilitative efforts, many patients do not RTW. To increase cost effectiveness, predictive tests for non-RTW are needed to select patients for rehabilitation. The reliability of these tests must be high, to prevent exclusion of patients who might improve. This study evaluates the reliability and predictive validity of four tests and the following psychosocial factors for non-RTW: nationality, off-work duration, unemployment and work load. It was designed as a prospective cohort study of 99 patients with chronic low back pain. Upon entry, physical work load, time off work, unemployment and nationality were recorded. The study investigated four tests with an anticipated prognostic value for non-RTW: the Numeric Pain Rating Scale (NRS, 9–10 of a maximum of 10), the Step Test and Pseudo Strength Test (precipitous cessation) and Behavioural Signs. After 12 months, the RTW rate was obtained from the physicians responsible for sick-listing by postal survey. The response rate regarding RTW was 91% at 1 year. The RTW rate at 1 year was 20%. All investigated tests significantly correlated with non-RTW. Regression analysis showed that the best prediction of non-RTW was obtained when at least two out of the four tests were positive (positive predictive value 0.97, sensitivity 0.45). Unemployment, time off work, nationality and physical work load were less predictive. The results show that the combination of the four prognostic tests allows a very reliable prognosis of non-RTW. The cost effectiveness of rehabilitation aiming at RTW will, therefore, be increased by excluding patients with two or more positive tests.

Keywords Predictive tests · Low back pain · Return to work · Rehabilitation · Illness behaviour · Response to examination

Introduction

Epidemiology

Chronic low back pain (CLBP), with its associated disabilities and compensations, is a common problem in western societies that has reached ‘epidemic proportions’ [31]. Between 1986 and 1996, Switzerland showed a 4.7% annual increase in the number of persons receiving a disability pension due to musculoskeletal problems (personal communication, F. Donini). Regions with a higher unemployment rate showed a greater increase in disability [4]. Non-medical factors such as unemployment apparently influence medical opinion regarding an individual’s working capacity, with the consequence that the unemployed are more frequently diagnosed as being disabled [7]. Patients medically diagnosed as unable to perform their present

work have the right to receive help from the Swiss Disability Insurance Corporation. All possible vocational measures are considered before a disability pension is granted. Vocational measures are usually coordinated by a social worker and may consist of advisory consultations, adaptive training, workplace visits and assistance in finding a job adapted to the functional capacity of the patient.

Illness behaviour

CLBP disability is mostly attributed to non-specific low back pain (NSLBP) [6, 33], occurring without identifiable specific anatomical or neurophysiological causative factors. Known causes for specific back pain are vertebral fracture, tumour, infection, inflammatory diseases, nerve root compression, spondylolisthesis, spinal stenosis and definite instability [6]. Evidence suggests that fewer than 15% of individuals with back pain can be assigned to one of these specific back pain categories [25]. The patient with NSLBP cannot be viewed as suffering solely from a nociceptive stimulation, but rather as having a problem involving biological, psychological and social aspects, all of which must be analysed from the perspective of a biopsychosocial model [6, 32].

Increased behavioural responses during the clinical examination are often observed in patients with CLBP [33]. Pain drawings may show a spreading in the painful areas, and patients may report a very high pain intensity, even at rest. Many patients tend to catastrophise, they are afraid that activity might lead to further spinal damage. Fear avoidance behaviour is the consequence, causing a decrease in activity. Positive Behavioural Signs [34] should be understood as a response to examination affected by fear in the context of recovery from injury and the development of chronic incapacity [17].

This study investigates the performance of the Step Test and the Pseudo Strength Test (see Methods; Fig. 1, Fig. 2). The inability to perform these tests cannot be explained by musculoskeletal pathology and may be regarded as behavioural signs. The reduction of physical activity resulting from illness behaviour causes a reduction in physical capacity due to disuse.

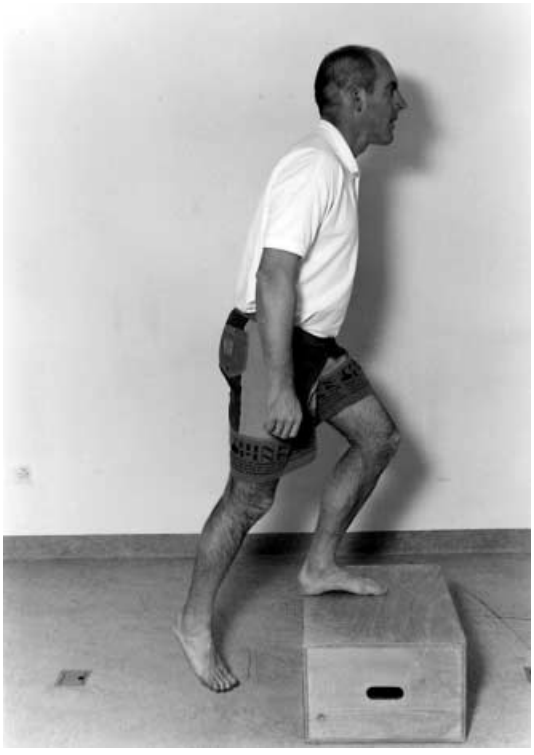

Fig. 1 The Step Test. The patient is asked to step up and down a 30-cm-high step for 3 min. Precipitous cessation of the test is counted as a positive test

Fig. 2 Pseudo Strength Test. The patient is in a supine position and holds two 3-kg weights against gravity with straight elbows and a 90° shoulder flexion for 2 min. Precipitous cessation of the test is counted as a positive test

Return to work: the primary goal of treatment

In terms of lost work days and the failure of modern medicine to diagnose the causative anatomical structure and to relieve the pain and disability in chronic pain patients, non-specific disability constitutes a continuing challenge. These problems have occasioned a change in the approach to treatment over the past few decades. Return to work (RTW), and not pain relief, has become the primary goal in the treatment of patients with CLBP [25]. To achieve this goal, an interdisciplinary evaluation and an intensive treatment, including work assessment, ergonomic intervention, postural information and strengthening exercises, is recommended [1, 25]. Research results are inconclusive as to the effectiveness of rehabilitation programmes to improve RTW in patients with CLBP [27]. RTW rates vary between 23 and 85% [19, 23]. Different selection procedures, unknown predictive factors for RTW and varying definitions of RTW may have influenced the results. For example, in a randomized clinical trial on 542 patients,

Mitchell reported an impressive rate of 79% 'full return to previous work'. If, however, the average of more than 400 sick days during the 2-year follow-up period [20] is taken into account, the investigated patients appear to be considerably more disabled than the classification as 'full return to work' might imply. In our study, RTW is defined as an improvement in actual work activity, and patients on vocational measures were regarded as non-RTW.

Prediction of non-return to work

This study investigates factors predicting non-RTW. Positive Behavioural Signs [34] are a reliable predictor for non-RTW after an intensive rehabilitation programme [12, 35]. A pilot study has shown that the Step Test, the Pseudo Strength Test and the reported pain intensity at initial examination also have the potential of being predictive for non-RTW [22]. Known psychosocial predictors for non-RTW after conservative treatment in patients with CLBP are: time off work, low social status, poor job security and nationality [2, 10, 18, 28, 32]. The outcome of a somatic-orientated rehabilitation programme for patients from the former Yugoslavia was significantly poorer than for that of Swiss patients [28]. We find it important to offer non-discriminating, objective tests to select patients for rehabilitation, and will evaluate the influence of nationality by comparing patients from the region of former Yugoslavia with the other patients mainly coming from Switzerland and its neighbouring countries.

Purpose

The purpose of this study is to determine the predictive validity of a modified version of the Step Test, the Pseudo Strength Test and a pain intensity of 9 or 10 (Numeric Rating Scale) for non-RTW in patients with CLBP. The second aim is to determine whether the prediction of non-RTW can be improved by combining these tests with the Behavioural Signs. We also evaluated the predictive value of work load, nationality, time off work and employment status.

Materials and methods

Design

The investigated predictive tests were analysed in a longitudinal prospective cohort study. Assessment and treatment of the patients were not altered for this study. Informed consent was obtained from all patients. The study design was approved by the ethical committee of the clinic.

The patients were referred to the clinic by their treating physicians without pre-selection by an insurance company or the clinic. Patients generally had suffered a long history of LBP. In most of these 'end of the line' patients, a variety of diagnostic procedures and many treatments had been performed.

All patients fulfilling the following admission requirements were consecutively included in the study.

Table 1 Measurements (*NRS* Numeric Rating Scale)

Measurements	Entry	12 months
Behavioural Signs	x	
Pain Intensity NRS 0–10	x	x
Step Test	x	
Pseudo Strength Test	x	
Disability (Roland and Morris Disability Questionnaire)	x	x
Vocational measures	x	x
Work activity	x	x

1. To enable RTW at follow-up, the age range was limited to 20–60 years.
2. Only patients willing to go back to full time work were included. Therefore, housewives and part-time employees were excluded.

And all included patients had:

3. Taken more than 6 weeks off work during the preceding 6 months because of CLBP
4. No comorbidity contributing to disability and sickness leave, and
5. Sufficient understanding of either German, Serbo-Croat, Spanish or Italian needed to fill out the questionnaires

Follow-up questionnaires were sent to the patients and the treating physicians 12 months after treatment (Table 1). Non-responders received a reminder and a phone call. The physicians were not informed about the results of the predictive tests.

Population, treatment and work status

The diagnosis was described according to the guidelines of the Quebec Task Force Report [25] (see Results and Table 4). According to the definition of Fordyce [6], the following four diagnostic codes are regarded as representative of a specific diagnosis:

1. Pain with radiation to the lower limb in combination with neurological signs
2. Confirmed compression of a spinal nerve root
3. Spinal stenosis
4. Low back pain less than 6 months following surgery

Work activity at the time of 1-year follow-up was recorded as a percentage of full-time activity. RTW was defined as an improvement in work activity. Patients were considered as non-RTW if their work activity was unimproved. Patients on vocational measures were regarded as non-RTW. Unemployed patients were also considered as non-RTW, because most patients lose their job as a consequence of LBP.

The physical work load was registered according to the classification of the US Department of Labour (max. load that has to be lifted at work: 1=0–5 kg, 2=5–10 kg, 3=10–25 kg, 4=25–45 kg, 5=>45 kg) [29]. Disability was measured with the Roland and Morris Disability Questionnaire [24] in order to describe the research population at entry and to investigate whether an improvement in disability correlates with an increased work activity.

Tests with an expected prognostic value for outcome

Pain intensity

Before the physical examination, the patient rated the current pain intensity on a Numeric Rating Scale (NRS 0–10). A pain intensity

Table 2 Behavioural Signs

Type of test	Clinical signs
Tenderness	
Superficial	The skin is tender to light pinch over a wide lumbar area. A localised band in a posterior primary ramus distribution may be caused by nerve irritation and should be discounted.
Deep	Tenderness is felt over a wide area. It is not localised to one structure, and often extends to the thoracic spine, sacrum, or pelvis.
Simulation tests	
Axial loading	Low-back pain is reported on vertical loading over the standing subject's skull by the examiner's hands. Neck pain is common and should be discounted.
Rotation	Back pain is reported when shoulders and pelvis are passively rotated in the same plane, as the subject stands relaxed with the feet together (see Fig. 3). In the presence of root irritation, leg pain may be produced and should be discounted.
Distraction Test	
Straight leg raising	Straight leg raising is the most useful distraction test. The subject whose back pain has a non-organic component shows marked improvement in straight leg raising on distraction as compared with formal testing.
Regional disturbances	
Sensory	Sensory disturbances include diminished sensation to light touch, pinprick, and sometimes other modalities, fitting a 'stocking' rather than a dermatomal pattern.
Weakness	Weakness is demonstrated on formal testing by a partial cogwheel 'giving way' of many muscle groups that cannot be explained on a localised neurological basis.
Overreaction	
	Overreaction during examination may take the form of disproportionate verbalisation, facial expressions, muscle tension and tremor, collapsing, or sweating. The response to procedures such as venipuncture or myelography provides additional information. Judgements should, however, be made with caution, minimising the examiner's own emotional reaction; there are considerable cultural variations, and it is very easy to introduce observer bias or to provoke this type of response unconsciously.

of 9 or 10 was counted as a positive test. The guidelines for the assessment of psychosocial 'Yellow Flags' [13] describe risk factors for the early identification of chronicity in patients with CLBP, and interpret a pain intensity of 9 or 10 as 'pain behaviour'.

Step Test

The Step Test was originally designed to evaluate aerobic capacity [30]. The patient is asked to step up and down a 30-cm-high step for 3 min, which requires four steps per cycle (Fig. 1). The initial stepping frequency is 96 (24 cycles per min). Patients were allowed to reduce speed to a comfortable level. The test does not increase the stress on the lumbar spine more then going up stairs. Precipitous cessation is counted as a positive test.

Pseudo Strength Test

In the Pseudo Strength Test, the patient lies in a supine position and is asked to hold two 3-kg weights with straight arms against gravity for 2 min (Fig. 2) [21]. This test does not give any significant load to the lumbar spine or the spinal muscles, and can be performed by patients with acute radicular pain. Precipitous cessation is counted as a positive test.

Behavioural Signs

Waddell et al. first described behavioural or 'nonorganic' signs as clinical signs that have a 'predominantly nonorganic basis' in patients with LBP [34]. They described eight items and grouped them into five categories (Table 2). Waddell's definition was used to interpret the test as positive if at least three out of the five categories scored positive.

'Two out of Four'

Single tests may have a greater risk of giving false-positive results, causing a decreased positive predictive value and specificity. Sensitivity of one test may also be limited. Therefore the Behavioural Signs, the Step Test, the Pseudo Strength Test and a pain score of 9 or 10 were combined into the 'Two out of Four' test, which was considered positive when two or more of these four tests were positive.

Psychosocial factors with an expected prognostic value for outcome

The main purpose of this study was the investigation of the previously described predictive tests. Because psychosocial factors are known to be important predictors for RTW, we also investigated the influence of work load, off-work duration, unemployment and nationality for RTW.

Table 3 Predictive tests for non-return to work (non-RTW)

	Non-RTW	RTW	
Positive predictive test	a	b false positive	a+b
Negative predictive test	c false negative	d	c+d
	a+c	b+d	a+b+c+d

PP = Prior Probability (non-RTW)=(a+c)/(a+b+c+d)
ARI = Absolute Risk Increase=a/(a+b)–(a+c)/(a+b+c+d)
PPV = Positive Predictive Value=a/(a+b)
NPV = Negative Predictive Value=d/(c+d)
Se = Sensitivity=a/(a+c)
Sp = Specificity=d/(d+b)

Statistical analysis

Logistic regression analyses were performed to evaluate which of the previously mentioned tests and psychosocial factors were significantly associated with non-RTW. The prior probability (PP) of RTW and non-RTW was determined. The absolute risk increase (ARI) measures the gain in predictive knowledge obtained from the test result. The maximum value for the ARI depends on the PP (ARI≤1–PP).

For each factor showing a significant association with non-RTW we determined the positive predictive value (PPV), the probability of non-RTW in patients with a positive factor. In a perfect test the PPV=1.

A positive predictive test for non-RTW indicates an increased probability of non-RTW. The negative predictive value expresses the ability of the test to predict RTW. Predictive tests must at least be able to predict either RTW or non-RTW in a reliable manner. Only a very good test fulfils both aims.

To select prognostic tests for the future exclusion of patients from treatment, a limit for false-positive results was set at 0.05. Therefore, the first requirement for the test is that the positive predictive value be >0.95. Optimal sensitivity is the second criterion for test selection. The significance level of the non-parametric correlation is displayed by the Chi-squared value.

Table 3 shows the formulas for the computation of the prior probability of non-RTW, the absolute risk increase, the positive and negative predictive value, the sensitivity and the specificity. Statistical analysis was performed using SPSS software.

Results

Subjects

Patients entered the study between January 1996 and December 1998. The average age of the 99 patients (84 male) was 41.8 years (SD 8.9 years). The median time off work because of LBP before rehabilitation was 26 weeks (range 6–349 weeks), including sick leave, time on a disability pension because of LBP and time out of work. Twenty-six percent of the patients were unemployed, 5% were unemployed and received a disability pension. The following nationalities were represented: region of former Yugoslavia 41%, Switzerland 35%, Italy 10%, Portugal 5% and other countries 9%. The physical work load was rated as 5 for 45%, 4 for 25%, 3 for 21% and 1 for 8% of the patients. Fifteen percent of the test population had been given a specific diagnosis. Table 4 shows the frequency of the different diagnostic categories. The 17 patients with symptomatic post-surgical status were classified as non-specific because they had shown negative results on specific imaging techniques concerning compression of a spinal nerve root, instability or spinal stenosis.

Table 4 The incidence of the different diagnostic categories

No. of patients (*N*=99)	Diagnosis according to the Quebec Task Force
7	Pain without radiation
6	Pain + radiation to extremity, proximally
30	Pain + radiation to extremity, distally
7	Pain + radiation to lower limb, neurological signs
0	Presumptive compression of a spinal nerve root
6	Compression of a spinal nerve root
0	Spinal stenosis
2	Post-surgical status, 1–6 months postoperative
17	Symptomatic post-surgical status, >6 months postoperative
24	Chronic pain syndrome

Return to work

The response rate at follow-up regarding RTW was 91% at 12 months. We obtained the data regarding RTW from the treating physicians who, in Switzerland, are responsible for the determination of work activity. Among physicians, the reason for non-response at 12 months was the withdrawal of informed consent by one patient and an untraceable new address in eight patients.

Before rehabilitation, 96% of the patients had been off work. Four percent of the patients were on a partially reduced work activity because of low back pain. The rate of persons at work increased from 4% at entry to 26% at 3 months ($P<0.05$), and then slightly decreased to 24% after 1 year. The RTW rate, the difference between the work rate before rehabilitation and after 1 year, was 20%. Thirteen patients returned to full work activity, five worked 50% and three worked 25%.

At entry, 25% of the patients were in contact with the Swiss Disability Insurance Corporation for vocational measures. This percentage increased to 47% at 3 months and 53% at 1 year (Fig. 3). The number of patients receiving a disability pension increased from 10% at entry to 28% after 1 year. Approximately 50% of the patients with a disability pension received a full pension. The 25–75% pensions of the other patients were supplemented either by working at a reduced level (time or work load), or being sick-listed or receiving unemployment benefit.

The correlation of improvement in the Roland Disability Questionnaire with RTW was positive, weak and not significant (Spearman $\rho=0.20$, $P=0.124$).

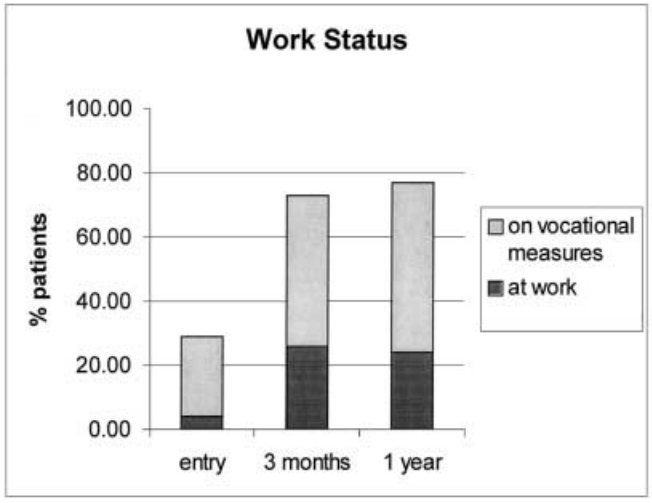

Fig. 3 Work status

Predictive tests and psychosocial factors

Table 5 shows the main results of this study: the reliability of the investigated tests in predicting non-RTW after 1 year. The results of all predictive tests were available from all patients. None of the patients showed comorbidity limiting the performance and interpretation of the Step Test and the Pseudo Strength Test. In all patients with precipitous cessation of the Pseudo Strength Test or the Step Test, the reason for discontinuation of the activity was an increase in back pain. None of the patients mentioned tiredness as the reason for stopping. The outcome concerning RTW at 1 year was obtained for 91% of the patients.

Logistic regression analysis showed that the Behavioural Signs, the Step Test, pain (NRS 9 or 10), the Pseudo Strength Test and the combination of these four tests 'Two out of Four' were significantly associated with non-RTW. The contribution of the psychosocial factors: unemployment, off-work duration, nationality and workload was not significant. Using a stepwise regression, the only factor taken into the model was 'Two out of Four'.

The negative predictive value was below 0.50 for all tests (Table 5), which means that none of the investigated tests is clinically useful for the identification of a patient who will RTW.

The prior probability for the outcome non-RTW was 0.80. The maximum absolute risk increase (ARI) for non-RTW was 0.20. The ARI was greater than 0.10 in seven out of nine investigated factors. Four tests ('Two out of Four', Step Test, Behavioural Signs and 'pain 9 or 10') had a positive predictive value greater than 0.95 and reached the first criterion for screening tests used to exclude patients from treatment. The second criterion, an optimal sensitivity, was best met by the Step Test (0.49) followed by 'Two out of Four' (0.45).

The Behavioural Signs, a recognised test to assess the behavioural response to examination, with a well-known predictive value for non-RTW, showed good results regarding sensitivity, specificity and positive predictive value. A pain intensity of 9 or 10 had the best positive predictive value and specificity for non-RTW after 1 year. There were no false-positive results. The sensitivity of a pain intensity of 9 or 10 proved relatively poor. Of the four investigated tests, the Step Test had the best sensitivity, but the specificity was relatively poor.

The frequency of positive predictive tests was the same in patients with (91%) and without (9%) 1-year RTW data.

Pain and disability

The following description of the results regarding pain and disability are meant to further describe the population and are based on the 63% response rate of patients at 12 months. Pain was initially rated at 6.2 and increased to 7.0 (P=0.015) over the following year. The score on the Roland Disability Questionnaire deteriorated insignificantly, from 16.2 to 16.5 out of a maximum score of

Table 5 Positive results in the Behavioural Signs and the three investigated tests significantly correlate with non-RTW. When two or more of these four tests ('Two out of Four') are positive, the prognosis of non-RTW shows optimal sensitivity and a PPV greater than 0.95 (*positive* patients with positive result/patients followed up, *ARI* Absolute Risk Increase, *Se* Sensitivity, *Sp* Specificity, *PPV* Positive Predictive Value, *NPV* Negative Predictive Value, χ^2 *P*-value for χ^2 test of non-parametric correlation)

	Positive	ARI	Se	Sp	PPV	NPV	χ^2
Two out of Four positive	31/90	0.19	0.45	0.95	0.97	0.31	0.0015
Step Test positive	36/90	0.14	0.49	0.88	0.95	0.30	0.0050
Behavioural Signs positive	36/90	0.18	0.38	0.94	0.96	0.30	0.0073
Pain NRS 9 or 10	18/90	0.20	0.26	1.00	1.00	0.28	0.0159
Pseudo Strength Test positive	26/90	0.13	0.37	0.88	0.93	0.27	0.0470
Unemployed	27/90	0.17	0.35	0.88	0.88	0.35	0.0597 ns
Off work >52 weeks	25/90	0.12	0.32	0.89	0.92	0.25	0.0727 ns
Nationality (former Yugoslavia)	39/90	0.06	0.48	0.72	0.87	0.26	0.1246 ns
Work load 4–5	63/90	0.00	0.71	0.30	0.77	0.24	0.9474 ns

24 points. In individual patients, a change is considered significant and clinically relevant when the difference exceeds four points [26]. Using this criterion, disability remained unchanged over the follow-up year in 61% of the patients. Twenty-five percent of the patients showed an improvement, whereas 14% had a deterioration of four or more points.

Treatment

This study does not investigate the effectiveness of treatment. Nevertheless the treatment may influence the prognosis of RTW and will therefore be briefly described here.

The average length of stay in the rehabilitation centre was 28.2 days (SD 7.4 days). The treatment duration was 16.2 h/week, with an emphasis on active treatment (11.2 h/week) such as strength and endurance training (4 h/week), exercise therapy (2.3 h/week), back school (1.9 h/week) and swimming with fins to increase endurance. Psychological interventions were offered to approximately 10% of the patients, depending on their personal needs. Relaxation techniques and passive treatment completed the interdisciplinary rehabilitation.

Discussion

To increase cost effectiveness in the rehabilitation of patients with CLBP, screening tests are needed to either include patients with a good prognosis for RTW or exclude patients who will almost certainly not RTW. The results of this study indicate that it is much easier to predict non-RTW than RTW. This may be because many prerequisites must be fulfilled for successful RTW. It is impossible to detect all relevant psychosocial factors and to judge their relevance and interactions. Moreover RTW depends to an important degree on factors that are unknown in advance, such as the availability of lighter work and the state of the job market. Some people ready to RTW might therefore not be at work after 1 year. Non-RTW is easier to predict, as negative factors apparently constitute significant barriers to RTW independent of other factors. Therefore, tests can be developed to predict non-RTW and exclude those patients from rehabilitation who have an extremely small probability of RTW. Cost effectiveness can be increased using these tests.

An important question is whether the results of this study can be generalized to other populations of patients with CLBP. It is important that, in this study, no pre-selection by the clinic or insurance companies took place. Patients were referred to the clinic by their treating physicians. Therefore, we think that the results of this study can be applied to other patients with CLBP.

The investigated group consisted mainly of men (84%) doing heavy or very heavy work. Sex did not correlate with RTW or with the number of predictive tests. The results remained the same when women were excluded from the analysis. Patients from other countries than Switzerland (65%) were over-represented in this study. This can be explained by the fact that heavy work is frequently done by persons from other countries. The off-work duration was long (median 26 weeks) and the unemployment rate was high (31%). Most patients reported that they had lost their job and had found it difficult to find a new job because of their LBP.

The incidence of a specific diagnosis may seem small, raising the question of whether diagnostic procedures are sufficient. Non-specific LBP was diagnosed in 85% of the patients. The medical records of the patients showed that diagnostic procedures had been exhaustively applied and that the rate of specific diagnosis correlates very well with the findings of other studies [8, 25].

The validity of the translated versions of the Roland and Morris Disability Questionnaire requires some analysis. The risk of a change in meaning after translation is considered small, as the items concern straightforward activities of daily life, such as climbing stairs, sleeping, walking and sitting. It is, however, possible that cultural differences lead to variations in scoring. Comparisons between groups of different nationalities might therefore not be valid. In this study, a comparison within a group of persons between entry and follow-up was performed. As expected [25], there was no change in symptoms or disability.

Loss to follow-up among patients was 37% at 12 months. As the follow-up data were not used for the prediction model, the incomplete follow-up did not bias the results. Responders and non-responders were the same at entry with regard to pain, disability, nationality and ‘Two out of Four’ positive tests. The improvement by more than four points on the disability level in 25% of the patients can be viewed as an illustration of the fact that LBP can get better. Further efforts to increase this percentage should be made.

The 91% response rate among the physicians at 12 months concerning RTW enables a reliable analysis of the predictive tests. Information about RTW was obtained from the physicians.

The 20% RTW rate after 12 months in the present study may seem very low compared to other studies. A possible explanation is that RTW is defined differently in different studies, contributing to the wide variations in results, which range from 23% [23] to 81% [9] and 85% [19]. Some authors use the first day of RTW [16]; in other studies patients ready to work [3, 11] or looking for work [5] are included in the group of persons who have RTW. In this study RTW was defined as being at work at the time of follow-up. Patients ready to work or on vocational measures were not considered as RTW. If patients on vocational measures (53%) had been included in the present study, the RTW rate would increase to 73%. There is also

some doubt as to the internal validity of the studies of Mayer et al. [19] and Hazard et al. [9], reporting RTW rates of 81% and 85%. In both studies, the experimental group was preselected by a medical insurance corporation, which agreed to finance the treatment. In addition, by including the drop-outs in Mayer's study (17%), the RTW is reduced to 70%. Lack of compliance with treatment was not clearly defined, and may have been one of the reasons for the drop-outs in Mayer's study [27]. In the present study, 37% of the patients can be regarded as showing a lack of compliance, as they were unable to perform the Step Test and 28% were unable to hold two weights of 3 kg from a supine position against gravity. The inability to perform these tests cannot be explained by a physical problem, and suggest that behavioural or psychosocial factors may play an important role. The RTW rate in our study has also been negatively influenced by the 31% unemployment rate. The fact, however, that two unemployed patients were working at the follow-up, indicates that unemployment alone does not justify exclusion from treatment. If patients with 'Two out of Four' positive predictive tests had been excluded from treatment, the RTW rate would have been 32% instead of 20%, comparable with Oland and Tveiten [23], who reported a RTW rate of 32% at 6 months and 23% after 18 months.

Non-RTW can be predicted with excellent reliability by means of the Step Test, the Pseudo Strength Test, and a Pain Rating of 9 or 10 (positive predictive value= 0.93–1.00, specificity 0.88–1.00). A positive Step Test and Pseudo Strength Test cannot be explained by mechanical pathology and suggest illness behaviour. This hypothesis is supported by the fact that 80% of the patients with a specific diagnosis were able to perform these two tasks. A pain intensity of 9 or 10 is interpreted as an exaggerated pain behaviour and not primarily related to a somatic disorder. This explains the comparable predictive value of these tests with the Behavioural Signs (0.96). Maximal sensitivity with a concomitant positive predictive value of more than 0.95 is achieved by the combination of four tests, interpreted positive when two or more of the four tests are positive. Kummel improved prediction of RTW by adding two 'new non-organic physical signs' to the Behavioural Signs [15]. In the present study, the evaluation of activity-associated illness behaviour in combination with the Behavioural Signs improves the prediction of RTW.

The decision to compare non-RTW in persons from former Yugoslavia with all other persons was based on the specific psychosocial factors among these patients. The constitution of groups can be debated and the results of this study argue strongly against the use of nationality as a predictive factor. In an analysis of non-RTW limited to persons with less than 'Two out of Four' positive tests, nationality no longer influenced RTW.

An unexpected result in this study is that work load is not predictive for non-RTW. This finding can be explained by the small number of patients with a low work load, and the fact that persons with a low work load are not white-collar workers, but employees with a low education, low responsibility and a very limited possibility of finding more attractive work.

Although nationality, unemployment and off-work duration are related to RTW (Table 5), using these factors to select patients for treatment would result in an unacceptable number of wrong decisions [14]. It is not ethical to apply nationality as a reason to exclude patients from treatment.

Conclusion

The validity to predict non-RTW in patients with CLBP was established for three of the four investigated tests (Step Test, Pseudo Strength Test, Pain Rating of 9 or 10). Positive results in these tests reliably indicate that the patient will not RTW after an extensive active rehabilitation programme.

A very good positive predictive value of 0.97 with a sensitivity of 0.45 is achieved by the combination of these three tests with the Behavioural Signs. These are interpreted as positive when two or more of these four tests are positive.

The investigated tests are based on objective criteria. Using these criteria, patients are evaluated based on their behaviour and not on potentially discriminating and less predictive factors such as nationality, work load, unemployment or time off work.

Cost effectiveness in rehabilitation of patients with CLBP can be improved by selecting patients accordingly.

Acknowledgements The authors wish to thank all physicians and physical therapists involved in the acquisition of data at entry and follow-up, and Lynne Mathews-Bietenhalder for her help with the preparation of the manuscript.

References

1. Abenhaim L, Rossignol M, Valat JP, et al (2000) The role of activity in the therapeutic management of back pain. Report of the International Paris Task Force on Back Pain. Spine 25 [Suppl4]: 1S-33S
2. Bendix A, Bendix T, Haestrup C (1998) Can it be predicted which patients with chronic low back pain should be offered tertiary rehabilitation in a functional restoration program? Spine 23:1775–1784
3. Bendix AF, Bendix T, Lund C, Kirkbak S, Ostenfeld S (1997) Comparison of three intensive programs for chronic low back pain patients: a prospective, randomized, observer-blinded study with one-year follow-up. Scand J Rehabil Med 29:81–89
4. Breitenmoser B, Foffa D, Rouiller C, Donini F, Nydegger B (1999) 'Eingliederung vor Rente' – realisierbares Ziel oder bloss wohltönender Slogan? Soz Sicherheit 6:288–294
5. Corey DT, Koepfler LE, Etlin D, Day HI (1996) A limited functional restoration program for injured workers: a randomized trial. J Occup Rehabil 6: 239–249
6. Fordyce W (1995) Back pain in the workplace: management of disability in nonspecific conditions. International Association for the Study of Pain, Seattle
7. Frey D, Kippele Emara C, Tyndall A (1997) Berentung unspezifischer Rückenschmerzen trotz multidisziplinärem Therapieprogramm - ein Systemversagen? Schweiz Med Wochenschr 127 [Suppl 90]:127
8. Frymoyer JW (1988) Back pain and sciatica. N Engl J Med 318:291–299
9. Hazard R, Fenwick J, Kalish S (1989) Early prediction of chronic disability after occupational low back injury. Spine 14:157–161
10. Hildebrandt J, Pfingsten M, Saur P, Jansen J (1997) Prediction of success of a multidisciplinary treatment program for chronic low back pain. Spine 22:990–1001
11. Indahl A, Velund L, Reikeraas O (1995) Good prognosis for low back pain when left untampered. A randomized clinical trial. Spine 20:473–477
12. Karas R, McIntosh G, Hall H, Wilson L, Melles T (1997) The relationship between nonorganic signs and centralization of symptoms in the prediction of return to work for patients with low back pain. Phys Ther 77:354–360; discussion: 61–69
13. Kendall N, Linton S, Main C (1997) Guide to assessing psychosocial yellow flags in acute low back pain: risk factors for long-term disability and work loss. Accident Rehabilitation and Compensation Insurance Corporation of New Zealand and the National Health Committee, Wellington
14. Kool J, Oesch P, Bachmann S (2000) Ist die Nationalität prädiktiv für die Beurteilung der körperlichen Leistungsfähigkeit und für das Rehabilitationsresultat? Schweiz Arztez 81:2656–2659
15. Kummel BM (1996) Nonorganic signs of significance in low back pain. Spine 21:1077–1081
16. Loisel P, Abenhaim L, Durand P, et al (1997) A population-based, randomized clinical trial on back pain management. Spine 22:2911–2918
17. Main CJ, Waddell G (1998) Behavioural responses to examination. A reappraisal of the interpretation of 'nonorganic signs'. Spine 23:2367–2371
18. Main CJ, Wood PL, Hollis S, Spanswick CC, Waddell G (1992) The Distress and Risk Assessment Method (DRAM). Spine 17:42–52
19. Mayer TG, Gatchell RI, Mayer H (1987) A prospective two year study of functional restoration in industrial low back injuries: an objective assessment procedure. JAMA 258:1763–1767
20. Mitchell RI, Carmen GM (1994) The functional restoration approach to the treatment of chronic pain in patients with soft tissue and back injuries. Spine 19:633–642
21. Oesch P, Kool JP, Sloksnath U, Hasegawa A (1994) Die Zuverlässigkeit und Empfindlichkeit von 4 isometrischen Muskeltests. Physiotherapie 6:4–12
22. Oesch P, Kool J, Wunderlin B, Knüsel O (1997) Rehabilitation von Patienten mit chronischen Rückenbeschwerden: Assessment, Ergebnis und prädiktive Faktoren. Phys Med Kurort Med 7:224–230
23. Oland G, Tveiten G (1991) A trial of modern rehabilitation for chronic low back pain and disability: vocational outcome and effect of pain modulation. Spine 16:457–459
24. Roland M, Morris R (1983) A study of the natural history of back pain. Spine 8:141–150
25. Spitzer W (1987) Quebec Task Force Report. Spine 12:S10-S53
26. Stratford P, Binkley J, Solomon P, Finch E, Gill C, Moreland J (1996) Defining the minimum level of detectable change for the Roland-Morris Questionnaire. Phys Ther 76:359–365
27. Teasell RW (1996) Functional restoration. Returning patients with chronic low back pain to work – revolution or fad? Spine 21:844–847
28. Thali A, Stern S, Rothenbühler B, Kraan K (1994) The role of psychosocial factors in patients with chronic low back pain. Z Unfallchir Versicherungsmed 87:31–44
29. United States Department of Labor (1991) Dictionary of occupational titles, 4th edn. United States Department of Labor, Washington, DC
30. Villiger B, Egger K, Lerch R, Probst HP, Schneider W (1995) The 3-Minute Step Test. In: Villiger B (ed) Endurance. Thieme, Stuttgart
31. Waddell G (1991) Low back pain disability, a syndrome of western civilisation. Neurosurg Clin N Am 2: 719–738
32. Waddell G (1992) Biopsychosocial analysis of low back pain. Baillieres Clin Rheumatol 6:523–557
33. Waddell G (1998) The back pain revolution. Churchill Livingstone, Edinburgh
34. Waddell G, McCulloch JA, Kummel E, Venner RM (1980) Non-organic physical signs in low back pain. Spine 5: 117–125
35. Werneke MW, Harris DE, Lichter RL (1993) Clinical effectiveness of Behavioural Signs for screening chronic low-back pain patients in a work-oriented physical rehabilitation program. Spine 18:2412–2418