ORIGINAL ARTICLE

Elisabeth Hansson
Tommy Hansson
Robert Jonsson

Predictors for work ability and disability in men and women with low-back or neck problems

E. Hansson (✉) · T. Hansson
Department of Orthopaedics,
Sahlgrenska University Hospital,
Bruna stråket 11, 413 45 Göteborg,
Sweden
E-mail: elisabeth.hansson@orthop.gu.se
Tel.: +46-31-3422399
Fax: +46-31-825599

R. Jonsson
Department of Statistics,
Gothenburg University,
Gothenburg, Sweden

Abstract In many countries back and neck problems are the most common reason for work incapacity. Numerous studies have looked for reliable predictors of return to work (RTW) or not return to work (NRTW) for those incapacitated, but with few exceptions their predictive and detective ability has been limited. We aimed to perform a comparative study of various commonly used health measures ability to predict RTW or NRTW in a cohort of men and women sick-listed for more than 28 days due to low-back pain (LBP) or neck pain (NP). A cohort of 1,575 men and women sick-listed more than 28 days due to back or neck problems was asked to answer ten different health measures (scales) about various aspects of health status (health related quality of life, pain, back function and depressivity) at four occasions during a 2-year period. A statistical diagnostic test was used to study how well the subject's values on these scales would serve as a predictor for work resumption. There was an improvement from day 28 up until 2 years, reflected by the different scales, more pronounced for LBP than for NP patients and for men with LBP compared with women with LBP. For all separately tested scales EuroQol (EQ-5D) had the highest overall ability to predict RTW or NRTW irrespective of gender, diagnoses or duration of the problems. When RTW or NRTW were predicted in a cohort of sick-listed low-back or neck patients, EQ-5D had outstanding properties in this respect irrespective of gender, diagnosis or elapsed time during this 2-year study.

Keywords Back pain · Neck pain · Predictors · Return to work · Disability

Introduction

When all the costs, direct and indirect, were individually determined during a 2-year period among nearly 2,000 employed Swedish men and women sick-listed 28 days or more due to low-back or neck problems, it was found that the direct costs, the costs for all medical interventions, including surgery, were only 6.9% of the total costs [20]. Although higher in that study (93.1%) than earlier reported, the indirect costs are known to be the major costs for back problems and increasing with the duration of the work disability [18, 43, 48].

Strikingly low return to work rates within 2 years (63% in the US and Sweden, 40% in Denmark, 35% in Germany, 72% in the Netherlands and 60% in Israel) were recently reported in a multinational study of subjects sick-listed more than three months for back problems [8]. In that multinational study, which aimed at reflecting "the everyday" treatment and rehabilitation of those chronic patients, there was no clear evidence of any effect on pain or back function, or on the

return to work rate, of any of all the non-surgical medical treatments or vocational rehabilitation tried [22]. Those "negative findings" confirmed what others have found, that reliable scientific evidence is missing for an effective treatment of long-lasting work-related back or neck problems [3, 17, 22, 35, 40]. For secondary prevention, many treatment efforts and modalities, and their combinations, have been tried without definite and reproduced success in bringing people back to work, however, especially those with long-lasting disabling spine problems back to work [22, 24, 37, 40, 44].

Awareness of the multifaceted character of the problem has generated trials emphasising exercise, education, ergonomic intervention or behavioural therapy alone or combined. In spite of such efforts, scientific evidence for a generally working concept is weak [17, 44]. Several studies have reported that modified work programs increase the cumulative return to work (RTW) rate and decrease the disability time considerably [29]. Whether such programs have a general effect or influence upon certain groups of sick-listed people has not been clarified, however.

For the back and neck patients as a group the work ability/inability is a multifaceted problem, to a great extent consisting of factors other than strictly medical ones [21, 40, 46]. Numerous high quality studies have looked for predictors not only of the medical outcome but work capacity/incapacity as well [1, 26, 36]. With few exceptions, the predictive and detective ability of the reported predictors has been limited [34]. Many of the frequently reported predictive factors have been quite general as higher age, earlier episodes of spine problems, sciatica or leg pain, while other more specific factors have had uncertain validity and frequently have consisted of single items from specific questionnaires alone or combined [19, 21, 42].

Presently, and at large, it seems as if the sick-listed sub-chronic or chronic back or neck patient, as a consequence of continuous problems, receives one usually ineffective "new" treatment/examination after the other until "everything" has been tried [22].

It is reasonable to assume that an early, versatile and reliable predictor of, for example, prognosis for RTW or non return to work (NRTW) would at least hold back some of the apparently ineffective interventions that not only add to costs but also probably prolong the already protracted course of the problems.

The object of this study was to perform a comparative study of the ability of various commonly used health measures to predict RTW or NRTW in a cohort of men and women sick-listed for more than 28 days due to low-back pain (LBP) or neck pain (NP).

Material and methods

Design

The study was a prospective cohort questionnaire and register study with a 2-year follow-up. The participants received mailed self-administered questionnaires after 28 days, 90 days, 1 year and 2 years.

Participants

A cohort of 1,822 employed Swedish men and women between 18 and 59 years sick-listed for 28 days due to either LBP or NP (sick listing certified by a physician) were selected consecutively from five socio-demographically representative regional Swedish General Insurance Offices [5]. Of the 1,822 subjects, 247 persons were not included in the present study since they reported both LBP and NP. A higher proportion of the remaining 1,575 subjects were women (56%), more than 70% of all subjects suffered from LBP. Among the persons who were sick-listed due to neck pain 70% were women. Because of these differences the participants in the study were presented in four groups: (1) men with LBP, (2) men with NP, (3) women with LBP and (4) women with NP. Self-employed and unemployed people and those with generalised arthritis or a fracture, tumour or infection as well as women suffering from back pain in connection with pregnancy, were excluded in the study.

Questionnaires

The questionnaires included a large set of baseline characteristics, like working and family conditions, education, economy, treatments and rehabilitations measures etc. and also ten different commonly used instruments reflecting various aspects of present health status, such as health-related quality of life, pain, disability, back function and depressivity (Table 1).

EuroQol

EuroQol (EQ-5D) is a generic health-related quality of life measure. It provides a single index. The individuals classify their own health status into five dimensions (5D)—mobility, self-care, usual activity, pain/discomfort and anxiety/depression within three levels: no problems, moderate problems and severe problems. The instrument yields a total of 243 possible states, the Time Trade Off method is used to rate the different states of health. The value 0.00 indicated dead and 1.00 indicated full health. The EuroQol thermometer is a visual analogue scale, on which the respondents is asked to mark his or her health between 0 and 100 [7–10, 13, 14].

Table 1 List of the instruments/scales used in the study, and the times at which the data were collected (*X* time when the different instruments were applied)

Scales	28 days	90 days	1 year	2 years
EQ-5D, health-related quality of life index	X	–	–	X
EuroQol, thermometer, quality of life index	X	–	–	X
Hannover ADL, back function (HADL)	X	X	X	X
SF36VI, vital health	X	X	X	X
SF36MH, mental health	X	X	X	X
SF36SF, social function	X	X	X	X
SF36GH, general health	–	X	X	X
ZUNG, depressivity	X	–	–	–
Von Korff's pain, pain	X	–	–	X
Von Korff's score, disability	X	–	–	X

Hannover ADL

This is a questionnaire developed for measuring back-pain-related disability. It is a self-administered questionnaire of 12 questions for the assessment of functional limitations in activities of daily living among patients with musculoskeletal disorders. The 12 items have to be scored, summed and transformed on to a scale from 0 (worst back function) to 100 (best back function) [25].

Thirty-six-item short form health survey

The questionnaire is a standardized, generic self-administered instrument. The thirty-six-item short form health survey (SF-36) describes eight domains of health with each scored from 0 (poor health) to 100 (optimal health) and results in a health profile. It is not designed to generate a single index. In the present study, only the subscales for General Health, Vitality, Social Functioning and Mental Health were used [45, 47].

Von Korff's pain and disability score

This questionnaire was made as a simple method of grading the severity of chronic pain for use in general population surveys and studies of primary care pain patients. It included seven questions which measure pain experience and functional restrictions during the last 6 months. Low values on the scale mean less pain or disability [27].

Zung's depressivity scale

It is a self-rating depression scale. The most commonly found characteristics of depression were used and divided into 20 items, ten were worded symptomatically positive and ten negative. The scale is constructed so that the less depressed patients will have a low score on the scale and the more depressed patient will have a higher score [49].

Higher values on the scale indicate better health or function, with exceptions for Von Korff's pain and disability score and Zung's depressivity scale, where better health is indicated by low values. For most scales, data were obtained at 28 days and 2 years, and therefore only the values obtained at the latter times were used in the comparative study.

Response rates

The response rates at 28 days and 2 years differed considerably according to gender and diagnoses (back or neck diagnoses) (Table 2). After 28 days, women and those with NP were more apt to respond than men and those with LBP. The number of respondents decreased substantially between 28 days and 2 years, especially among men with NP, where the non-response rate exceeded 50%. Each subject's sick-listing status (i.e. RTW or NRTW) during the 2-year study was obtained to 100% from each participating General Health Insurance office. From those data a rather comprehensive non-response analysis was made and showed that the refusals revealed a higher proportion of young people, males, individuals with LBP and people sick-listed for a shorter period [20, 22].

Table 2 The response rates after 28 days and after 2 years among men and women, and among those with LBP and NP diagnoses. The percentage of respondents varied slightly between the different scales

Original number of participants			Percentage of respondents after	
			28 days (%)	2 years (%)
Men	LBP	545	64–69	38–40
	NP	147	69–74	33–35
Women	LBP	552	72–74	50–51
	NP	331	74–79	51–53
Total		1,575		

Prevalence of work resumption

The ability to predict work resumption based on the value of the scale will be dependent on the *prevalence*, i.e. the proportion of subjects who have not returned to work after a certain time. The prevalence decreased steadily from 28 days to 2 years with different patterns depending on gender and diagnosis, which is reported elsewhere [5]. The prevalence for respondents and non-respondents obtained for the scale 1. EQ-5D is presented in Table 3. A similar pattern was obtained for all the other scales. No significant difference in prevalence was detected between respondents and non-respondents, in spite of the relatively large sample sizes. It thus seems reasonable to assume that the respondent data were not biased in this respect.

Determination of a cut-off value c representing RTW or NRTW for each scale

In order to study how well the subject's value on a scale may serve as a predictor of work resumption, an approach based on a statistical diagnostic test was used [11]. Here, the ability of a certain scale to serve as a predictor is expressed in terms of probability (proportion) of correct predictions. To this end, one first has to determine a cut-off value c on the scale such that a subject is defined as RTW on scale for values in one direction of c and NRTW on scale in the opposite direction. Since the direction of the scales used in the present study differed, NRTW on scale was used for values below c on the scales 1–7 in Table 1 and NRTW on scale for values above c on the scales 8–10 in Table 1. One proper way to determine the numerical value of c is considered below.

The true states to be predicted were termed RTW if the subject had returned to work and NRTW if not. Consider the following definitions: *predictive value of RTW on scale* (PRTW) = proportion of subjects which are RTW given they are RTW on scale, and *predictive value of NRTW on scale* (PNRTW) = proportion of subjects which are NRTW given they are NRTW on scale. These predictive values are dependent on the cut-off value c and also on the prevalence, so that with decreasing prevalence PRTW increases, while PNRTW decreases. This is illustrated in Fig. 1, where PRTW and PNRTW are plotted against c on scale 1 (EQ-5D score) for men with LBP diagnosis, as an example. Notice that PNRTW approaches the value of the prevalence as c increases, while PRTW tends to 1. It is also seen that the PRTW-PNRTW difference increases as the prevalence increases. Since the predictive values are dependent on the prevalence, they are less useful in a comparative study. Instead the following relative measures were used:

$$\text{Re1PRTW} = \frac{\text{PRTW} - (1 - \text{Prevalence})}{1 - \text{Prevalence}} \qquad (1)$$

$$\text{Re1PNRTW} = \frac{\text{PNRTW} - \text{Prevalence}}{\text{Prevalence}} \qquad (2)$$

Here, RelPRTW and RelPNRTW can be interpreted as the relative gain obtained by using the scale as a predictor rather than merely using the relative frequency of RTW and NRTW, respectively. RelPNRTW = 0.70 thus means that 70% more NRTW subjects can be classified correctly by using the scale as a predictor, than a prediction based on the relative frequency of NRTW.

Sensitivity, specificity and some further diagnostic measures

Besides the predictive measures (1) and (2) above, the scales were compared regarding the detective concepts *Sensitivity* = proportion of NRTW which are classified as NRTW on scale and *Specificity* = proportion of RTW which are classified as RTW on scale. Both these measures depend on the cut-off point c. By plotting Sensitivity on the Y-axis against 1-Specificity on the X-axis for various values of c, one obtains the receiver operating characteristic (ROC) curve, (Fig. 2). The latter is used to compare the detective ability of alternative diagnostic tests in such a way that the diagnostic test with the highest values on the ROC curve is preferred [4]. In this study there were a large number of ROC curves to be compared. In order to simplify the comparisons, only one point on the ROC curve was used and the following measure was found to be useful:

$$\text{ROClevel} = \sqrt{(\text{Sensitivity})^2 + (\text{Specificity})^2}. \qquad (3)$$

The measure (3) is simply the distance from the point on the curve to point 1.0. The measure can take values

Table 3 Prevalence (proportion NRTW) among those who participated in the study. The corresponding prevalence for the non-respondents is shown *in parentheses*

Values on scale after	True state after	Men		Women	
		LBP	NP	LBP	NP
28 days	90 days	0.56 (0.54)	0.58 (0.60)	0.65 (0.63)	0.63 (0.63)
28 days	1 year	0.19 (0.17)	0.27 (0.30)	0.27 (0.24)	0.27 (0.25)
28 days	2 years	0.12 (0.11)	0.20 (0.21)	0.17 (0.15)	0.20 (0.18)
2 years	2 years	0.10 (0.11)	0.25 (0.21)	0.19 (0.15)	0.21 (0.18)

Fig. 1 a The EQ-5D values after 28 days and RTW within 90 days for men with LBP diagnoses. Prevalence (NRTW) at 90 days 0.56. **b** The EQ-5D values after 28 days and RTW within 1 year for men with LBP diagnoses. Prevalence (NRTW) at 1 year 0.19. **c** The EQ-5D values after 28 days and RTW within 2 years for men with LBP diagnoses. Prevalence (NRTW) at 2 years 0.12
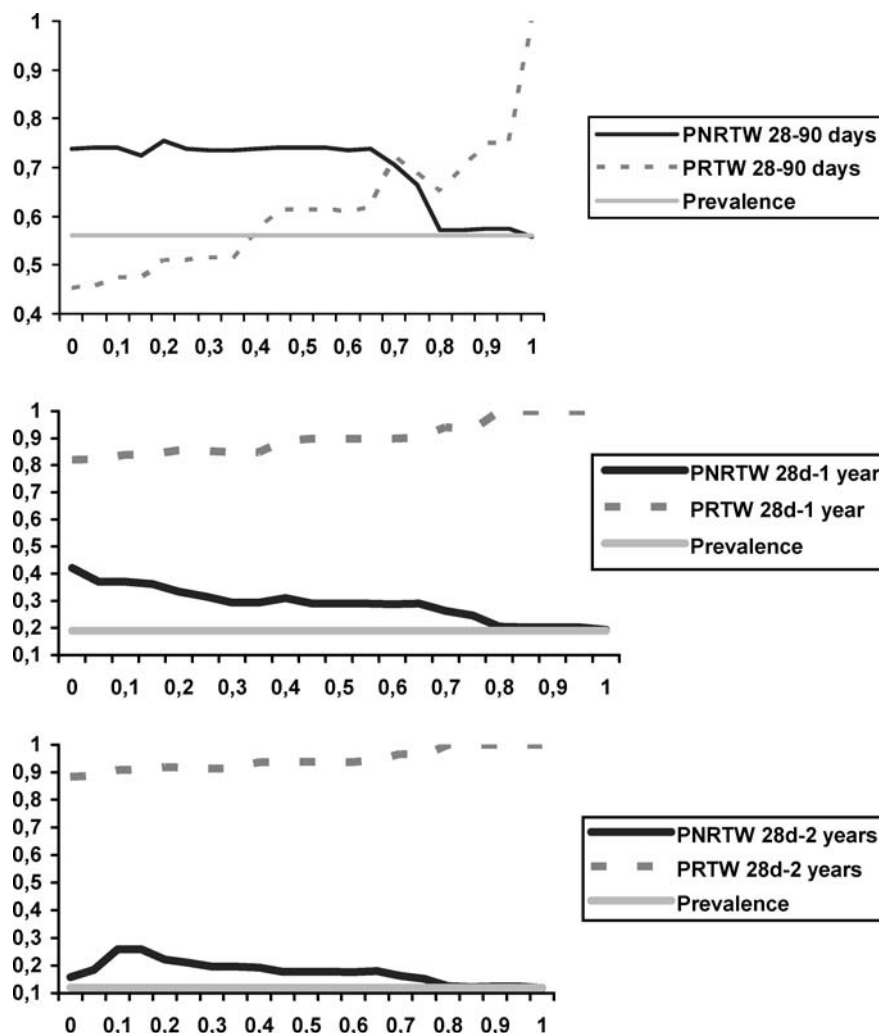
within the interval $(0, \sqrt{2} \approx 1.41)$, large values indicating high detecting ability of the diagnostic test. In practice one should require that (3) is larger than $1/\sqrt{2} \approx 0.71$ since the latter limit is obtained for a diagnostic test which produces false classifications at the same rate as true classifications.

Predictive measures are used for predicting the outcome on the basis of a diagnostic test. Detective measures are typically used when one wants to replace a comprehensive health examination by a simpler one. Here the emphasis will be on prediction.

The problem of determining a proper cut-off value c is complicated by the fact that a large value of c on the scales 1–7 increases PRTW but decreases PNRTW (Fig. 1), with a similar reversed relation for Sensitivity and Specificity. In the present study, a good compromise was to chose c as the largest value on the scales 1–7 (and the smallest value on the scales 8–10) for which PNRTW−Prevalence > PRTW−(1−Prevalence). For

all three curves in Fig. 1 this rule yielded $c = 0.65$. In some cases, the PRTW and PNRTW curves were unstable with heavy fluctuations. They sometimes also behaved in an unexpected way, e.g. the PNRTW increased with increasing c on the scales 1–7. In these cases no cut-off value was determined. Such anomalies may to some extent be explained by sampling fluctuations if the sample size is small. However, the most plausible explanation is that the scale has very little to do with the predicted outcome.

The ability of the scales in Table 1 to predict work resumption, as measured by rel PRTW and rel PNRTW, and to detect those subjects who returned to work, as measured by ROC level, was ranked. This was made separately for men with LBP, men with NP, women with LBP and women with NP. The measures (1, 2, 3) were calculated for scale values after 28 day and work resumption after 90 days, 1 year and 2 years.
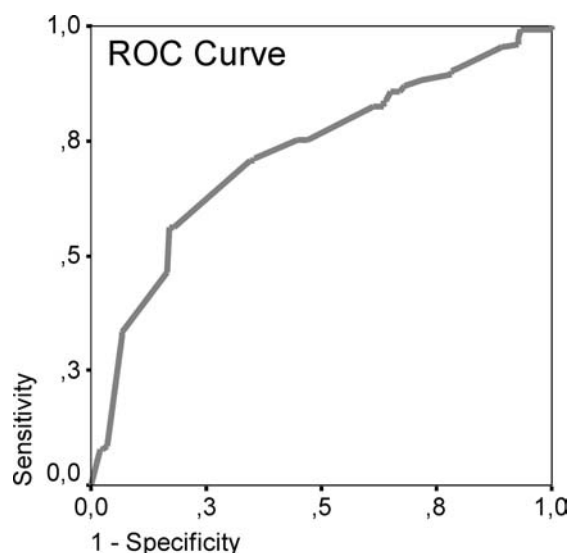
Fig. 2 Sensitivity (proportion of NRTW which are classified as NRTW on EQ-5D scale) and 1 minus specificity (proportion of RTW which are classified as RTW on EQ-5D scale) at 90 days for men with LBP. Each point on the ROC curve has a corresponding cut-off value

One may use other measures of diagnostic ability, such as *rate of true positive accuracy* (TP) = PNRTW/total sample size, *rate of true negative accuracy* (TN) = PRTW/total sample size, and *overall diagnostic accuracy* = TP + TN. These measures (TP and TN) are also presented, although they may be less informative in a forecast situation. To help clarifying the diagnostic tests used in this study a general representation of diagnostic test is presented below in Table 4 [2].

Table 4 General representation of diagnostic test.

	Disease status		
	Positive	Negative	Total
Test			
Positive	*a*	*b*	*a* + *b*
Negative	*c*	*d*	*c* + *d*
Total	*a* + *c*	*b* + *d*	*n*

Test	Disease status	Name
+	+	True positive (*a*)
+	−	False positive (*b*)
−	+	False negative (*c*)
−	−	True negative (*d*)

Sensitivity = *a*/(*a* + *c*)
Specificity = *d*/(*b* + *d*)
Positive predictive value = *a*/(*a* + *b*)
Negative predictive value = *d*/(*c* + *d*)
Prevalence = (*a* + *c*)/*n*

Results

Summary characteristics of the scales

The mean scores on the various scales are presented in Table 5 for gender and type of diagnoses. With few exceptions, there was an improvement in the different aspects (quality of life, pain, function and depressivity) the scales were meant to measure from day 28 up until 2 years. Generally, the improvement—as reflected by the different scales—was more pronounced for LBP than for NP, and for men with LBP compared with women having LBP. There was also an overall tendency that men scored "better" than women. This pattern was most obvious for patients with LBP, where the largest differences between men and women were obtained for the scales 3 (HADL), 6 (SF36 Social function) and 9 (VKP). Here, the men scored 7–15% better and the differences were highly significant ($P < 0.001$).

Table 6 shows the correlations between the scores on the different scales after 28 days according to gender and type of diagnoses. The negative signs for scales 8–10 are simply due to the fact that an improvement (better) is indicated by lower values on these scales. When combining scales, one shall look for scales with low correlations in absolute value. Such correlations are found between the groups of scales (5, 6) and (1–3), and also between the groups of scales (9, 10) and (1–8), respectively. It must be noted that the correlations between the scales within the different groups were high. The implication of this is that, if one wants to combine values from several scales, then one shall choose some of (1–3), some of (5, 6) and some of (9, 10). These combinations were valid in all the four gender and diagnoses groups.

Predictive and detective ability of the scales

Tables 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 reflect the predictive and detective ability of the scales 1–10 (except for 7) when the scores were obtained after 28 days and the outcome (RTW or NRTW) after 90 days, 1 and 2 years. In the tables, the Sensitivity, Specificity, *TP* and *TN* are also presented. Some of the scales could not be used, due to the anomalous behaviour of the PRTW and PNRTW curves described in the Materials and methods. This was especially notable for outcomes far ahead in time and for women.

On the whole, RelPNRTW for the highest ranked scales was larger for men than for women and for subjects with NP than for those with LBP. Men and those with NP were thus easier to predict as NRTW. Scale 1 (EQ-5D), followed by 3 (HADL) and 2 (EuroQol thermometer), performed best on the whole as judged by RelPNRTW. It is notable that the scale 6 (SF-36 Social Function) served very well for predicting NRTW after

Table 5 Mean scores for the scales 1–10 in Table 1 after 28 days and 2 years. RTW is indicated by large values on scales 1–7 and by small values on scales 8–10. (Scales: *1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score)

Scale	After 28 days				After 2 years			
	Men		Women		Men		Women	
	LBP	NP	LBP	NP	LBP	NP	LBP	NP
1	0.49	0.53	0.49	0.50	0.68	0.58	0.63	0.59
2	59	59	55	54	65	58	62	59
3	60	73	52	67	71	68	62	66
4	44	44	38	35	59	50	49	46
5	64	63	62	62	73	67	68	69
6	60	63	53	60	77	65	65	68
7	_a	_a	_a	_a	61	53	56	54
8	0.49	0.49	0.52	0.52	_b	_b	_b	_b
9	67	67	72	73	48	50	53	60
10	71	65	73	67	41	51	46	50

[a]Not included at 28 days
[b]Not included at 2 years

Table 6 Correlations (%) between the scores on the scales after 28 days according to gender and diagnosis. No values from the scale 7 SF-36GH were obtained. *Scales: 1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score

Scale\scale	1	2	3	4	5	6	8	9
Men LBP								
1								
2	54							
3	58	50						
4	48	53	46					
5	38	39	34	65				
6	30	24	33	48	52			
8	−42	−48	−41	−62	−76	−50		
9	−49	−42	−46	−40	−25	−14	−32	
10	−31	−29	−39	−35	−20	−25	−26	−50
Men NP								
1								
2	67							
3	56	36						
4	44	37	47					
5	23	19	37	70				
6	33	43	32	42	47			
8	−45	−42	−56	−67	−77	−52		
9	−42	−34	−31	−24	−14	−23	−26	
10	−30	−32	−41	−30	−27	−40	−30	−41
Women LBP								
1								
2	57							
3	51	46						
4	41	46	40					
5	31	36	19	57				
6	40	42	38	48	42			
8	−40	−40	−28	−53	−76	−41		
9	−40	−38	−35	−28	−21	−24	22	
10	−33	−27	−40	−31	−19	−35	16	42
Women NP								
1								
2	44							
3	39	49						
4	34	47	41					
5	28	38	22	61				
6	28	29	31	52	49			
8	−30	−42	−32	−61	−75	−52		
9	−36	−31	−35	−32	−17	−18	23	
10	−29	−39	−50	−42	−31	−44	37	40

Table 7 Outcome after 90 days, men with LBP and NP. The relative predictive value of not return to work (*Rel PNRTW*) and the relative predictive value of return to work (*Rel PRTW*) are shown *in parentheses* (Prevalence LBP 0.56, NP 0.58). *Scale: 1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score

Men

LBP						NP					
Scale	(Rel) PNRTW	Scale	(Rel) PRTW	Scale	ROC level	Scale	(Rel) PNRTW	Scale	(Rel) PRTW	Scale	ROC level
6	(0.33) 74%	9	(0.57) 69%	1	0.96	1	(0.59) 92%	9	(0.71) 72%	1	1.17
1	(0.32) 74%	1	(0.40) 62%	9	0.94	3	(0.55) 87%	1	(0.69) 71%	2	1.08
2	(0.24) 69%	2	(0.38) 61%	2	0.93	2	(0.50) 87%	2	(0.54) 65%	3	1.02
3	(0.22) 69%	4	(0.30) 57%	6	0.93	6	(0.44) 84%	4	(0.50) 63%	4	0.98
10	(0.19) 67%	3	(0.24) 55%	3	0.88	10	(0.43) 83%	3	(0.28) 54%	6	0.96
4	(0.17) 65%	5	(0.23) 54%	4	0.87	4	(0.29) 75%	5	(0.20) 50%	10	0.96
5	(0.16) 65%	10	(0.22) 54%	10	0.85	8	(0.23) 71%	8	(0.17) 49%	9	0.96
8	(0.16) 65%	8	(0.14) 50%	5	0.85	5	(0.19) 69%	6	(0.12) 47%	8	0.88
9	(0.14) 64%	6	(0.12) 49%	8	0.83	9	(0.09) 63%	10	(0.11) 47%	5	0.85

Table 8 Outcome after 90 days, men with LBP and NP. The Sensitivity (*Sens*), Specificity (*Spec*), true positive (*TP*) and true negative (*TN*) values for the different scales. *Scale: 1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score

Men

LBP								NP							
Scale	Sens	Scale	Spec	Scale	TP	Scale	TN	Scale	Sens	Scale	Spec	Scale	TP	Scale	TN
9	0.86	6	0.88	9	0.48	6	0.38	9	0.92	6	0.93	9	0.54	6	0.39
2	0.71	1	0.71	2	0.41	1	0.31	1	0.72	1	0.93	1	0.42	1	0.39
4	0.67	8	0.69	4	0.38	8	0.30	4	0.71	10	0.93	4	0.42	10	0.39
1	0.65	3	0.68	1	0.37	3	0.29	2	0.66	3	0.93	2	0.38	3	0.39
7	0.64	10	0.64	7	0.36	10	0.28	7	0.60	2	0.86	7	0.38	2	0.36
5	0.61	7	0.59	5	0.34	7	0.26	5	0.51	8	0.77	5	0.25	8	0.32
10	0.56	5	0.59	10	0.32	5	0.26	3	0.42	7	0.73	3	0.24	7	0.27
3	0.56	2	0.59	3	0.31	2	0.25	8	0.42	5	0.69	8	0.24	5	0.27
8	0.45	4	0.56	8	0.26	4	0.24	6	0.24	4	0.67	6	0.14	4	0.27
6	0.29	9	0.39	6	0.16	9	0.17	10	0.23	9	0.27	10	0.13	9	0.11

Table 9 Outcome after 90 days, women with LBP and NP. The relative predictive value of not return to work (*Rel PNRTW*) and the relative predictive value of return to work (*Rel PRTW*) are shown *in parentheses*) (Prevalence LBP 0.65, NP 0.63). *Scale: 1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score

Women

LBP						NP					
Scale	(Rel) PNRTW	Scale	(Rel) PRTW	Scale	ROC level	Scale	(Rel) PNRTW	Scale	(Rel) PRTW	Scale	ROC level
1	(0.27) 83%	9	(0.76) 62%	1	0.97	1	(0.30) 82%	8	(0.64) 61%	1	0.97
6	(0.17) 76%	2	(0.49) 52%	9	0.95	3	(0.29) 81%	1	(0.37) 51%	8	0.97
8	(0.15) 75%	1	(0.43) 50%	2	0.90	2	(0.20) 76%	5	(0.19) 44%	3	0.94
2	(0.13) 73%	3	(0.33) 47%	6	0.87	10	(0.11) 70%	2	(0.19) 44%	2	0.88
3	(0.13) 73%	8	(0.22) 43%	3	0.86	5	(0.07) 67%	3	(0.18) 44%	10	0.81
10	(0.11) 72%	10	(0.15) 40%	8	0.85	8	(0.03) 65%	10	(0.13) 42%	5	0.80
4	(0.11) 72%	6	(0.15) 40%	4	0.82	4	–	4	–		
9	(0.06) 69%	4	(0.10) 39%	10	0.81	6	–	6	–		

788

Table 10 Outcome after 90 days, women with LBP and NP. The Sensitivity (*Sens*), Specificity (*Spec*), true positive (*TP*) and true negative (*TN*) values for the different scales. *Scale: 1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score

Women

LPB								NP							
Scale	Sens	Scale	Spec	Scale	TP	Scale	TN	Scale	Sens	Scale	Spec	Scale	TP	Scale	TN
9	0.92	6	0.78	9	0.60	6	0.27	8	0.95	3	0.89	8	0.61	3	.31
2	0.75	1	0.77	2	0.49	1	0.27	5	0.65	1	0.80	5	0.42	1	0.29
3	0.65	4	0.74	3	0.42	4	0.26	1	0.54	2	0.78	1	0.34	2	0.28
1	0.59	8	0.70	1	0.38	8	0.24	7	0.47	10	0.69	7	0.31	10	0.25
8	0.50	10	0.67	8	0.32	10	0.23	10	0.43	7	0.68	10	0.28	7	0.23
10	0.47	3	0.56	10	0.31	3	0.20	2	0.42	5	0.47	2	0.27	5	0.17
6	0.37	2	0.51	6	0.24	2	0.19	3	0.33	8	0.12	3	0.21	8	0.04
4	0.37	9	0.23	4	0.24	9	0.10								

Table 11 Outcome after 1 year, men with LBP and NP. The relative predictive value of not return to work (*Rel PNRTW*) and the relative predictive value of return to work (*Rel PRTW*) are shown *in parentheses* (Prevalence LBP 0.19, NP 0.27). *Scale: 1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score

Men

LBP						NP					
Scale	(Rel) PNRTW	Scale	(Rel) PRTW	Scale	ROC level	Scale	(Rel) PNRTW	Scale	(Rel) PRTW	Scale	ROC level
3	(0.53) 29%	1	(0.12) 91%	1	0.94	2	(0.67) 45%	2	(0.21) 88%	2	1
1	(0.51) 29%	4	(0.12) 91%	3	0.92	3	(0.57) 42%	4	(0.14) 83%	1	0.92
8	(0.47) 28%	3	(0.11) 90%	4	0.91	1	(0.56) 42%	1	(0.13) 82%	4	0.90
2	(0.44) 27%	2	(0.08) 87%	2	0.87	6	(0.44) 39%	8	(0.12) 82%	3	0.90
6	(0.37) 26%	8	(0.06) 86%	8	0.86	8	(0.34) 36%	3	(0.09) 79%	6	0.89
4	(0.36) 26%	9	(0.05) 85%	6	0.86	5	(0.25) 34%	5	(0.07) 78%	8	0.87
5	(0.33) 25%	5	(0.03) 83%	5	0.84	4	(0.23) 33%	6	(0.03) 75%	5	0.81
9	(0.32) 25%	6	(0.03) 83%	9	0.82	9	–	9	–	9	–

Table 12 Outcome after 1 year, men with LBP and NP. The Sensitivity (*Sens*), Specificity (*Spec*), true positive (*TP*) and true negative (*TN*) values for the different scales. *Scale: 1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score

Men

LPB								NP							
Scale	Sens	Scale	Spec	Scale	TP	Scale	TN	Scale	Sens	Scale	Spec	Scale	TP	Scale	TN
4	0.77	6	0.81	4	0.15	6	0.64	4	0.80	6	0.86	4	0.22	6	0.63
1	0.76	5	0.76	1	0.14	5	0.61	2	0.73	3	0.75	2	0.21	3	0.55
3	0.69	8	0.69	3	0.13	8	0.55	8	0.66	2	0.68	8	0.18	2	0.49
2	0.62	9	0.64	2	0.12	9	0.51	1	0.63	1	0.67	1	0.17	1	0.49
9	0.52	2	0.62	9	0.10	2	0.49	5	0.53	5	0.61	5	0.15	5	0.44
8	0.51	3	0.60	8	0.10	3	0.48	3	0.50	8	0.57	3	0.13	8	0.42
5	0.35	1	0.56	5	0.07	1	0.46	6	0.24	4	0.41	6	0.07	4	0.30
6	0.29	4	0.48	6	0.06	4	0.38								

90 days, but lost some relative predicting ability after 1 and 2 years.

A somewhat different pattern was seen for the prediction of RTW, as measured by rel PRTW. Also, here the scales 1–3 performed well on the whole, but scale 9 (Von Korff's pain) was outstanding as a short-time (90 days) predictor for all subjects except for women with NP. Here, scale 9 performed so badly that it could

Table 13 Outcome after 1 year, women with LBP and NP. The relative predictive value of not return to work (*Rel PNRTW*) and the relative predictive value of return to work (*Rel PRTW*) are shown *in parentheses* (Prevalence LBP 0.27, NP 0.27). *Scales: 1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score

Women

LBP						NP					
Scale	(Rel) PNRTW	Scale	(Rel) PRTW	Scale	ROC level	Scale	(Rel) PNRTW	Scale	(Rel) PRTW	Scale	ROC level
1	(0.48) 40%	3	(0.25) 91%	3	0.97	3	(0.58) 43%	1	(0.14) 83%	1	0.92
9	(0.30) 35%	1	(0.15) 84%	1	0.92	1	(0.53) 41%	2	(0.13) 82%	2	0.87
2	(0.30) 35%	2	(0.13) 82%	2	0.87	2	(0.24) 33%	3	(0.10) 80%	4	0.81
5	(0.28) 34%	9	(0.08) 79%	9	0.83	4	(0.20) 32%	4	(0.08) 70%	3	0.78
6	(0.27) 34%	4	(0.08) 79%	5	0.83	5	–	5	–	5	–
4	(0.24) 33%	6	(0.05) 77%	6	0.82	6	–	6	–	6	–
3	(0.24) 33%	5	(0.04) 76%	4	0.82	8	–	8	–	8	–

Table 14 Outcome after 1 year, women with LBP and NP. The Sensitivity (*Sens*), Specificity (*Spec*), true positive (*TP*) and true negative (*TN*) values for the different scales. *Scale: 1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score

Women

LBP								NP							
Scale	Sens	Scale	Spec	Scale	TP	Scale	TN	Scale	Sens	Scale	Spec	Scale	TP	Scale	TN
3	0.92	5	0.76	3	0.25	5	0.55	2	0.74	3	0.76	2	0.20	3	0.55
2	0.70	6	0.71	2	0.19	6	0.52	1	0.63	1	0.67	1	0.17	1	0.49
1	0.69	1	0.62	1	0.18	1	0.45	4	0.61	9	0.54	4	0.16	9	0.39
4	0.59	9	0.62	9	0.16	9	0.45	9	0.56	4	0.52	9	0.15	4	0.38
9	0.56	4	0.57	4	0.15	4	0.41			2	0.46			2	0.34
6	0.40	2	0.52	6	0.10	2	0.38								
5	0.35	3	0.32	5	0.09	3	0.23								

Table 15 Outcome after 2 years, men with LBP and NP. The relative predictive value of not return to work (*Rel PNRTW*) and the relative predictive value of return to work (*Rel PRTW*) are shown *in parentheses* (Prevalence LBP 0.12, NP 0.20). *Scales: 1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score

Men

LBP						NP					
Scale	(Rel) PNRTW	Scale	(Rel) PRTW	Scale	ROC level	Scale	(Rel) PNRTW	Scale	(Rel) PRTW	Scale	ROC level
4	(0.65) 20%	1	(0.07) 94%	4	0.97	1	(0.77) 35%	1	(0.16) 93%	1	1.02
2	(0.55) 19%	4	(0.07) 94%	1	0.93	2	(0.69) 34%	2	(0.15) 92%	2	0.98
1	(0.52) 18%	2	(0.06) 93%	2	0.91	3	(0.64) 33%	8	(0.11) 89%	3	0.90
8	(0.39) 17%	8	(0.03) 91%	8	0.83	6	(0.62) 32%	3	(0.08) 86%	6	0.90
3	(0.32) 16%	3	(0.02) 90%	3	0.81	8	(0.42) 28%	4	(0.07) 85%	8	0.90
5	–	5	–	5	–	5	(0.33) 27%	5	(.07) 85%	5	0.83
6	–	6	–	6	–	4	(0.26) 25%	6	(0.03) 82%	4	0.82

not be used and instead the scale 8 (Zung), was the outstanding one.

These findings suggest that if the prediction of return to work or not is to be based on a single scale, then scale 1 (EQ-5D) should be used. A prediction based on the latter scale for men with NP and the outcome after 90 days, yields that 92% of the NRTW and 72% of the RTW cases can be correctly predicted (Table 6) [the latter figures were obtained by inserting the prevalence 0.58 from Table 3 and (2)]. However, it is evident that one can do better by combining those scales that are good in predicting NRTW, with those which are good in

Table 16 Outcome after 2 years, men with LBP and NP. The Sensitivity (*Sens*), Specificity (*Spec*), true positive (*TP*) and true negative (*TN*) values for the different scales. *Scale: 1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score

Men

LPB								NP							
Scale	Sens	Scale	Spec	Scale	TP	Scale	TN	Scale	Sens	Scale	Spec	Scale	TP	Scale	TN
1	0.76	3	0.69	1	0.09	3	0.61	1	0.80	6	0.86	1	0.16	6	0.68
4	0.70	8	0.67	4	0.08	8	0.59	2	0.74	3	0.74	2	0.16	3	0.58
2	0.69	4	0.62	2	0.08	4	0.54	8	0.69	2	0.65	8	0.15	2	0.50
8	0.48	2	0.59	8	0.06	2	0.52	4	0.61	1	0.63	4	0.13	1	0.50
3	0.42	1	0.54	3	0.05	1	0.48	5	0.56	5	0.61	5	0.12	5	0.48
								3	0.52	8	0.57	3	0.11	8	0.44
								6	0.27	4	0.55	6	0.06	4	0.43

Table 17 Outcome after 2 years, women with LBP and NP. The relative predictive value of not return to work(*Rel PNRTW*) and the relative predictive value of return to work (*Rel PRTW*) are shown *in parentheses* (Prevalence LBP 0.17, NP 0.20) *Scale: 1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score

Women

LBP						NP					
Scale	(Rel) PNRTW	Scale	(Rel) PRTW	Scale	ROC level	Scale	(Rel) PNRTW	Scale	(Rel) PRTW	Scale	ROC level
2	(0.57) 27%	1	(0.10) 91%	1	0.93	3	(0.67) 33%	1	(0.11) 89%	1	0.94
1	(0.56) 27%	2	(0.07) 86%	2	0.93	1	(0.65) 33%	3	(0.07) 86%	3	0.90
6	(0.36) 23%	3	(0.06) 88%	3	0.83	2	(0.31) 26%	2	(0.04) 83%	2	0.81
3	(0.32) 22%	6	(0.03) 85%	6	0.83	8	(0.18) 24%	8	(0.03) 82%	8	0.78

Table 18 Outcome after 2 years, women with LBP and NP. The Sensitivity (*Sens*), Specificity (*Spec*), true positive (*TP*) and true negative (*TN*) values for the different scales. *Scale: 1* EQ-5D; *2* EuroQol thermometer; *3* HADL; *4* SF-36VI; *5* SF-36MH; *6* SF-36SF; *7* SF-36GH; *8* Zung; *9* Von Korff's pain and *10* Von Korff's score

Women

LPB								NP							
Scale	Sens	Scale	Spec	Scale	TP	Scale	TN	Scale	Sens	Scale	Spec	Scale	TP	Scale	TN
1	0.72	6	0.71	1	0.12	6	0.59	1	0.68	3	0.74	1	0.13	3	0.60
3	0.61	2	0.68	3	0.11	2	0.56	3	0.52	2	0.67	3	0.10	2	0.55
2	0.57	1	0.59	2	0.10	1	0.49	8	0.46	1	0.65	8	0.09	1	0.52
6	0.43	3	0.56	6	0.07	3	0.47	2	0.45	8	0.62	2	0.08	8	0.50

predicting RTW. Account shall also be taken of the length of the forecast interval, gender and diagnosis. Accordingly, the scales which have been found useful for predicting the outcome after 90 days are 1–3, 6 and 9, with the exception of women with NP, for whom scale 8 should be replaced by scale 9.

Discussion

The results from this study suggest that RTW or NRTW among sick-listed lower back or neck pain patients can be predicted to a high extent, especially through the use of the quality of life measure EQ-5D. The study revealed some particularities; e.g. that men generally scored "better", i.e. higher QoL, less pain, less functional impairment and less depression than women at all four measuring occasions during the 2-year study. That men scored "better" was especially pronounced for those sick-listed with LBP problems [22]. It was also evident that the response patterns underwent changes with time during the 2-year follow-up. For those reasons it was obvious that the most accurate prediction could be made after grouping of the sick-listed according to gender, diagnoses and elapsed time off work.

It was obvious that of all the separately tested scales EQ-5D had the highest overall ability to predict RTW or NRTW in this cohort of initially (28 days or more) sick-listed LBP or NP subjects. That ability remained irrespective of gender, diagnoses or duration of the problems. EQ-5D's prediction ability of, for example, 92% of NRTW and 72% of RTW within 90 days is the highest or among the highest seen in the literature. It was also evident that the predictive ability could be even more complete by combining selected scales. To obtain the best result when different scales were combined, scales best at predicting RTW should be combined with those best at predicting NRTW. Although EQ-5D had the best overall capacity, the predictive ability of the different instruments varied somewhat with time, as well as with diagnoses and gender. EQ-5D and von Korff's pain instrument had the advantage of retaining a high predictive ability irrespective of gender and diagnoses over the 2-year study period.

Von Korff's pain questionnaire was for example best in predicting early RTW (< 90 days), except for women with NP. That pain is of particular importance during the early phases of back or neck problems is not surprising and seemed to be confirmed recently in both acute and sub-acute LBP [28]. Why pain was of less importance specifically for women with NP is one of many results which could not be easily explained. The pain level has been found to be of great importance for the RTW/NRTW also in chronic LBP patients, [15], [16, 26]. When several job and clinical factors that all predicted RTW/NRTW were analysed and compared, it was found that the level of pain was of greatest importance. In one of those studies it was even found that without utilising the pain level RTW/NRTW could not be predicted. That several of the instruments used in the present study, EQ-5D, von Korff's pain and disability scales, as well as Zung and Hannover ADL, involve dimensions or consequences of pain directly or indirectly might thus be one explanation for their high predictive ability. Several other studies have also found that there are different predictors depending on the duration of time off work (compensated disability) [32]. The predictive ability of EQ-5D revealed in the present study is especially interesting, since EQ-5D was one of the QoL instruments, SF-36 the other, that was recommended for standard use in spine research of a group of international researchers [12]. In their proposal of useful instruments they discuss properties that ought to characterise such instruments [12]. To be recommended for use in a standard battery the instrument must have validity, especially construct validity, but also responsiveness and be practical to use. As far as we know EQ-5D's predictive ability, as shown here, has not been tried before. Since not only its validity, but also responsiveness, specificity and sensitivity were tested here, it apparently is a useful instrument for prediction as well.

The simplicity of the EQ-5D instrument in relation to its ability is a little bit astonishing. Even when compared with the most recently tested predictive instrument covering a relatively wide range of psychological variables, EQ-5D's only (five questions with three alternatives) seem quite adequate [33]. The complexity of prediction was shown by Kool et al. [26], who reported that when "two out of four" of a step test positive, behavioural signs positive, a high pain level (nine or ten), or pseudo strength test positive, NRTW could be predicted with high probability.

The focus on psycho-social factors for the LBP patient's work ability, particularly job satisfaction, was set by the findings in the so-called Boeing study during the early 1980s [6]. From that time it was more or less accepted that the disability related to neck and back problems must be understood as having a multifactor background where psychological and/or psycho-social factors are of considerable importance [8, 17, 22, 30, 32, 38, 40, 41]. That is especially true for disabilities lasting beyond the acute (medical) phase of the problems [21, 39, 40]. Several different physical, psycho-social, social and psychological factors repeatedly have been identified as risk indicators of disability accompanying those problems [8, 22, 23, 31, 40]. In a recent scientific evidence review of risk factors for long-lasting compensated work absence (> 3 months) due to back or neck problems, it was concluded that they were very similar for those seen in non-specific pain syndromes [21].

As frequently seen in prospective questionnaire surveys, the non-response rate was rather high, especially after 1 and 2 years. Since in this particular study, the work status of each participant, responder and non-responder, i.e. RTW or NRTW to 100%, was obtained from the registers of the compulsory National Health Insurance the effects of missed subjects in those respects were limited and also allowed a rather comprehensive non-response analysis [20]. How can the information from this study be used? Shaw et al. [42] concluded in their wide-ranging review of prognostic factors for low-back disability that the understanding of the importance of many factors in those respects are still missing or even contradictory. Although in no way clarifying the specific reason(s), neither individually nor on a group basis for a sick-listed subject's decision or ability for RTW or NRTW, some of the tested predictors in the current study, alone or combined, can serve the purpose of detecting early who, with sub-chronic back or neck pain, will or can return early, late or not at all to work. An early accurate prediction can help in directing more adequately, for example, the rehabilitation efforts of bringing someone back to work. By avoiding, for example, "unnecessary interventions" among those where the chances for RTW are good and instead concentrating vocational and medical

rehabilitation resources on those where the risks for an extended disability period are great ought to improve the outcome. How applicable are the findings? Although this study was undertaken on a Swedish cohort, it is quite likely that the results are applicable for similar patients in other countries. The reason for this is that the Swedish cohort was only one part of a study with identical core protocols that was also performed simultaneously in the Netherlands, Germany, United States (California and New Jersey), Denmark and Israel as well [8]. The national results from this multinational study, particularly in relation to the main outcome, RTW/NRTW were in many aspects so similar that the predictors found in the present study might work even in the other countries [22].

The present paper has been focused on finding those scales that might be useful for predicting return to work or not. It turned out that EQ-5D had the highest overall ability in this respect. However, as can be seen from the Tables 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 there were other scales that occasionally performed better. This suggests that predictions can be improved by combining the values from several scales. For this purpose other statistical techniques are available, for example discriminant analyses or logistic regression. But, this is beyond the scope of this study.

Conclusions

When RTW or NRTW were predicted in a cohort of sick-listed low-back or neck patients, EQ-5D had outstanding properties in this respect irrespective of gender, diagnosis or elapsed time during a 2-year study.

References

1. Abenhaim L, Rossignol M, Gobeille D et al (1995) The prognostic consequences in the making of the initial medical diagnosis of work-related back injuries. Spine 20:791–795
2. Altman DG (1995) Practical statistics for medical research. Chapman & Hall, London, p 415
3. Aronoff G (2000) Preventing disability from chronic pain: a review and reappraisal. Int Rev Psychiatry 12:157–169
4. Beck SE Jr (1986) The use of relative operating characteristics (ROC) curves in test performance evaluation. Arch Lathol Lab Med 110:13–20
5. Bergendorff S, Hansson E, Hansson T (1997) Projektbeskrivning & undersökningsgrupp. Stockholm: Rikförsäkringsverket, Sahlgrenska Universitetssjukhuset; Report No. 1
6. Bigos S, Battie M, Fisher L (1991) Methodology for evaluating predictive factors for the report of back injury. Spine 16:669–670
7. Björk S, Norinder A (1999) The weighting exercise for the Swedish version of the EuroQol. Health Econ 8:117–126
8. Bloch F, Prins R (eds) (2000) Who returns to work & why? A six country study on work incapacity & reintegration. Transaction Publisher, New Brunswick
9. Brooks R (1996) EuroQol: the current state of play. Health Policy 37:53–72
10. Burström K (2003) Population Health and Inequalities in Health—measurement of health-related quality of life and changes in QALY over time in Sweden. Karolinska Institutet, Stockholm
11. Campbell M (1990) Medical statistics. A commonsense approach. Wiley, West Sussex
12. Deyo R, Battie M, Beurskens A et al (1998) Outcome measures for low back pain research. A proposal for standardized use. Spine 23:2003–2013
13. Dolan P (1990) A social tariff for EuroQol: results from a UK general population survey. In: York Center for Health Economics Discussion Paper No 138. University of York, York
14. Dolan P (1997) Modeling valuations for EuroQol health states. Med Care 35:1095–1108
15. Fishbain D, Abdel-Moty E, Cutler R et al (1994) Measuring residual functional capacity in chronic low back pain patients based on the dictionary of occupational titles. Spine 19:872–880
16. Fishbain D, Cutler R, Rosomoff H et al (1999) Validity of the dictionary of occupational titles residual functional capacity battery. Clin J Pain 15:102–110
17. Frank J, Brooker A, DeMaio 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
18. Franklin G, Fulton-Kehoe D (1996) Outcomes research in Washington state workers' compensation. Am J Ind Med 29:642–648
19. Gallagher R, Rauh V, Haugh L et al (1989) Determinants of return-to-work among low back pain patients. Pain 39:55–67
20. Hansson E, Hansson T (2004) The cost for subjects sick listed more than one month because of low back or neck problems. Eur Spine J (in press)
21. Hansson T, Jensen J (2003) Sjukskrivning - orsaker och praxis. En systematisk litteraturöversikt. SBU-rapport, Stockholm
22. Hansson T, Hansson E (2000) The effects of common medical interventions on pain, back function, and work resumption in patients with chronic low back pain: a prospective 2-year cohort study in six countries. Spine 25:3055–3064
23. Hogg-Johnson S, Cole D (2003) Early prognostic factors for duration on temporary total benefits in the first year among workers with compensated occupational soft tissue injuries. Occup Environ Med 60:244–253
24. Indahl A, Velund L, Reikeraas O (1995) Good prognosis for low back pain when left untampered. A randomized clinical trial. Spine 20:473–477
25. Kohlmann T, Raspe H (1994) Grading of backache. Ther Umsch 51:375–380

26. Kool J, Oesch P, de Bie R (2002) Predictive tests for non-return to work in patients with chronic low back pain. Eur Spine J 11:258–266
27. Korff von M, Ormel J, Keefe F et al (1992) Grading the severity of chronic pain. Pain 50:133–149
28. Kovacs F, Abraira V, Zamora J et al (2004) Correlation between pain, disability, and quality of life in patients with common low back pain. Spine 29:206–210
29. Krause N, Frank J, Dasinger L et al (2001) Determinants of duration of disability and return-to-work after work-related injury and illness: challenges for future research. Am J Ind Med 40:464–484
30. Krause N, Ragland D (1994) Occupational disability due to low back pain: a new interdisciplinary classification based on a phase model of disability. Spine 19:1011–1020
31. Krause N, Ragland D, Greiner B et al (1997) Psychosocial job factors associated with back and neck pain in public transit operators. Scand J Work Environ Health 23:179–186
32. Lancourt J, Kettelhut M (1992) Predicting return to work for lower back pain patients receiving worker's compensation. Spine 17:629–640
33. Linton S, Boersma K (2003) Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Orebro Musculoskeletal Pain Questionnaire. Clin J Pain 19:80–86
34. Linton S, Hallden K (1998) Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. Clin J Pain 14:209–215
35. Maher C (2000) A systematic review of workplace interventions to prevent low back pain. Aust J Physiother 46:259–269
36. Mahmud M, Webster B, Courtney T et al (2000) Clinical management and the duration of disability for work-related low back pain. J Occup Environ Med 42:1178–1187
37. Malmivaara A, Hakkinen U, Aro T et al (1995) The treatment of acute low back pain–bed rest, exercises, or ordinary activity?. N Engl J Med 332:351–355
38. Mannion A, Dolan P, Adams M (1996) Psychological questionnaires: do "abnormal" scores precede or follow first-time low back pain? Spine 21:2603–2611
39. Nachemson A, Jonsson E (2000) Neck and back pain. The scientific evidence of causes, diagnosis and treatment. Lippincott Williams & Wilkins, Philadelphia
40. Nachemson A, Jonsson E (eds) (2000) Ont I ryggen, ont I nacken. SBU Statens beredning för medicinsk utvärdering. Report No: 145. Stockholm
41. Sewitch M, Rossignol M, Bellavance F et al (2000) First lifetime back pain and psychiatry treatment: psychological factors and recovery in compensated workers. Aaohn J 48:234–242
42. Shaw W, Pransky G, Fitzgerald T (2001) Early prognosis for low back disability: intervention strategies for health care providers. Disabil Rehabil 23:815–828
43. Spengler D, Bigos S, Martin N et al (1986) Back injuries in industry: a retrospective study I overview and cost analysis. Spine 11:241–245
44. Staal J, Hlobil H, van Tulder M et al (2002) Return-to-work interventions for low back pain: a descriptive review of contents and concepts of working mechanisms. Sports Med 32:251–267
45. Sullivan M, Karlsson J, Ware J (1994) The Swedish SF-36 Health Survey 1. Evaluation of data quality, scaling assumptions, reliability, and construct validity across general populations in Sweden. Göteborg
46. Waddell G, Sawney P (2002) Back pain, incapacity for work and social security benefits: an international literature review and analysis. London
47. Ware J, Kosisnki M, Gandek B (1993) SF-36 Health survey Manual and Interpretation Guide. New England Medical Center, The Health Institute, Boston
48. Williams D, Feuerstein M, Durbin D et al (1998) Health care and indemnity costs across the natural history of disability in occupational low back pain. Spine 23:2329–2336
49. Zung W (1965) A self-rating depression scale. Arch Gen Psychiatry 12:63–70