Work-related evaluation and rehabilitation of patients with non-acute nonspecific low back pain

Peter Oesch

Dissertation for the degree Philosophiae Doctor (PhD)

Department of Health Sciences
Institute of Health and Society
Faculty of Medicine
University of Oslo

December 2011

Acknowledgments

First of all I would like to thank Professor Kare Birger Hagen, my supervisor from the University of Oslo, for guiding me into the world of research, for his continuing support during my university years and, most importantly, for his prompt responses whenever I had any questions. I also would like to express my gratitude to Professor Nina K. Vollestad, my co-supervisor, for sharing her knowledge in research and her comprehensive and constructive inputs.

This work would not have been possible without the help of my colleagues interested in work-related evaluation and rehabilitation. I owe special thanks to Dr. Michael Oliveri, who made me aware for the need of work specific assessments and rehabilitation interventions. I also want to thank PD Dr. Stefan Bachmann and Dr. Jan Kool for all these years of scientific work in this field. I may not forget my colleague Roger Hilfiker giving me statistical advice and Kathrin Meyer and Beatrice Jansen for their highly valued contributions as well as recruiting patients.

This work received tremendous support from my colleagues at the Klinik Valens. I am deeply grateful to Urs Gamper for his continuous support and encouragement during my whole professional career and to Heike Daniel, Kurt Luyckx, Beate Schnurr and Volker Kirchberg for taking an extra amount of daily work so that I could focus on this thesis. I also want to thank the management board of the Klinik Valens supporting me to conduct this doctoral thesis.

Additional thanks go to all my colleagues who provided invaluable help in the assessment of patients with non-acute nonspecific low back pain. Not least, I am indebted to the patients who participated in the examinations.

Many thanks also to Ariane Knüsel and Daryl Snell for their assistance in preparing the manuscripts.

Finally, and most important my deepest and sincerest thanks go to my wife Cornelia and our children Lisa and Simon for their patience and for tolerating – over and over again – that I disappeared for hours behind the computer to complete this work. Without their understanding and patience this thesis would not have been completed.

Content

List of papers

This doctoral thesis is based on the following publications:

Paper I

Perceived functional ability assessed with the spinal function sort: is it valid for European rehabilitation settings in patients with nonspecific non-acute low back pain?
Oesch P, Hilfiker R, Kool JP, Bachmann S, Hagen KB. Eur Spine J2010;2010:21.

Paper II

What is the Role of 'Nonorganic-Somatic-Components' in Functional Capacity Evaluations in Patients with Chronic Non-Specific Low Back Pain Undergoing Fitness for Work Evaluation?
Oesch P, Meyer K, Janssen B, Mowinckel P, Bachmann S, Hagen KB. Spine (Phila Pa 1976)2011 Aug 18.

Paper III

Comparison of two methods for interpreting lifting performance during Functional Capacity Evaluation
Oesch P, Meyer K, Bachmann S, Hagen KB, Vollestad NK. Submitted for Publication, 2011.

Paper IV

Effectiveness of exercise on work disability in patients with non-acute nonspecific low back pain: Systematic review and meta-analysis of randomised controlled trials.
Oesch P, Kool J, Hagen KB, Bachmann S. J Rehabil Med2010 Mar;42(3):193-205.

Abbreviations

FAB	Fear avoidance belief
FABQ	Fear Avoidance Belief Questionnaire
FCE	Functional Capacity Evaluation
LBP	Low back pain
NSLBP	Nonspecific low back pain
PFA	Perceived functional ability
ROC	Receiver Operating Characteristic
RTW	Return to work
SFS	Spinal Function Sort
SRM	Standardised Response Mean

Background of this thesis at a glance

Perceived functional ability for work tasks assessed with the Spinal Function Sort
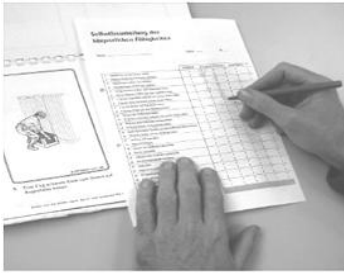
Assessment of perceived functional ability depends on the use of questionnaires. However, self-reported measures require an adequate literacy level and depend on linguistic abilities. Text-based questionnaires are often impossible to administer in European rehabilitation settings. A possible approach to overcome this problem is the use of picture-based questionnaires such as the Spinal Function Sort (SFS). The clinical utility of the SFS has so far only been reported in patient samples from the USA. No studies have been performed investigating the validity of the SFS in European patients.

Functional Capacity Evaluation – The role of 'nonorganic-somatic-components'

Functional Capacity Evaluation (FCE) has been shown to reflect physical capacity to some degree but is also influenced by perceived disability and pain intensity. It was therefore proposed that FCE should be considered as behavioural tests influenced by multiple factors, including physical ability and psychosocial factors. However, there must have been some missing determinants of physical performance as these variables were unable to explain large amounts of the variation in FCE performance. A possible confounding factor related to FCE performance might have been the presence of 'nonorganic-somatic-components'.

Functional Capacity Evaluation – Two methods for interpretation of lifting performance

Physical effort determination is attempted during FCE in order to interpret lifting performance. The Isernhagen FCE uses observational criteria for effort level determination during lifting tests. 'Nonorganic-somatic-components' were not intended to determine physical effort but have been used as a mean for effort determination. The questions arise whether determination of physical effort by observational criteria and 'nonorganic-somatic-components' can be interchangeably used to interpret lifting performance during FCE.

Exercise to reduce work-related disability
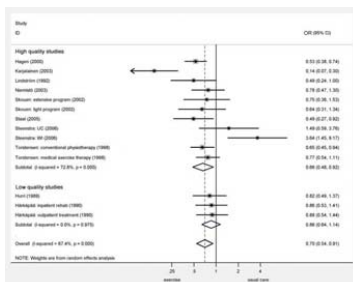
Previous reviews found strong evidence that exercise reduces work disability in patients with nonspecific low back pain. These reviews were based on studies published before 2004 and did not evaluate the effectiveness of different exercise characteristics. The effect of specific exercise characteristics on work disability is still unclear; a more up to date review is required.

Definitions

Assessment and evaluation‡	Assessment and evaluation are the preferred terms to be used in Functional Capacity Evaluation (FCE). One can use either one of them, depending on the purpose.
	Evaluation: A systematic approach including observation, reasoning and conclusion. Going beyond monitoring and recording, the evaluation process implies an outcome statement that is explanatory, as well as an objective measurement.
	Assessment: A systematic approach including observation, reasoning and conclusion.
Capacity‡	The highest probable level of functioning that a person can reach in a domain at a given moment in a standardised environment.
Fitness for work evaluation	The medical determination of whether the employee can perform the job or task under the working conditions.
Functional Capacity Evaluation†	An evaluation of the physical capacity to perform work tasks on a safe basis that is used to make recommendations for participation in work.
'Nonorganic somatic components'	A patient does not have a straightforward physical problem, but that illness behaviour and psychological factors also need to be considered
Perceived functional ability	People's beliefs about their capabilities to perform work tasks.
Psychosocial	The interaction between the person and his or her social environment, and the influences on his or her behaviour [1].
Safety‡	Safety is a situation in which, given the known characteristics of the person, the procedure should not be expected to lead to injury.
Submaximal effort	A patient stops a manual handling test before the criteria indicative of a maximum weight are observed.
Test‡	A standardised procedure of measurement.
Performance‡	Performance is 'what a person does in the current environment'.
Work-related rehabilitation	A treatment program aimed at enabling injured or disabled patients to return to work.

‡Inconsistent terminology has hindered research on Functional Capacity Evaluation [2]. A delphi round among FCE experts [3] found consensus in 10 out of 19 definitions. These were used throughout this thesis.

† Soer et al. [3] recommend that researchers state how they define FCE as only 63% agreement was found in the Delphi round on the following definition: "*FCE is an evaluation of capacity of activities that is used to make recommendations for participation in work while considering the person's body functions and structures, environmental factors, personal factors and health status*".

Abstract

Background

Low back pain (LBP) continues to be a major health problem causing personal suffering [4] and enormous socioeconomic costs [5-7]. Most of the patients suffer from nonspecific LBP (NSLBP) [8-10], defined as not attributable to a recognisable known specific pathology [10]. NSLBP is classified according to the duration and localisation of symptoms. Pain lasting longer than 4 weeks is non-acute [11].

Management of LBP should include a medical evaluation screening for specific pathology and for psychosocial [12] and work-related factors [9,12]. Traditional impairment-based medical measures defining fitness for work are criticised for lacking predictive validity as only few objective physical or biomechanical measures are associated with return to work (RTW) [13-16]. Currently, there are reforms in progress in many countries that move away from an 'essentialist' diagnostic approach [17] in disability determination towards an evaluation of functional capacity [7]. Functional tests purporting to measure a patient's physical capacity to perform work tasks are employed within Functional Capacity Evaluation (FCE). However, there is increasing evidence that not only physical but also psychosocial factors, such as perceived functional ability (PFA) or work tasks and 'nonorganic-somatic-components', influence FCE results and consequently decisions on work ability.

Modern treatment guidelines for non-acute NSLBP recommend multidisciplinary treatment, cognitive behavioural therapy, supervised exercises and early RTW. No evidence has been found that one form of exercise is superior to another [12]. Although new studies have been published in the meantime, the effect of specific exercise characteristics on work disability is still unclear.

Aims

1) To test the validity of the Spinal Function Sort (SFS) assessing PFA for work-related activities by evaluating internal consistency, unidimensionality, concurrent and predictive validity, and responsiveness in a European rehabilitation setting for patients with non-acute NSLPB.

2) To investigate the influences of '"nonorganic-somatic-components'', together with physical and other psychosocial factors, on the results of an FCE in patients with chronic NSLBP undergoing physical fitness for work evaluation.

3) To assess the contribution of 'nonorganic somatic components' and 'submaximal effort' to lifting performance and to determine the concurrent validity of the 'nonorganic somatic components with 'submaximal effort' during FCE in patients with chronic NSLBP.

4) To determine whether exercise is more effective than usual care to reduce work disability in patients with non-acute NSLBP, and if so, to explore which type of exercise is most effective.

Material and Methods

The different aims of the thesis required different research designs. Patients with non-acute NSLBP within working age, presenting with at least 6 weeks of sick leave, and with sufficient understanding of German, French or Italian were included in papers I, II, and III.

For paper I a prospective cohort study was used. This was embedded within a randomised controlled trial performed during inpatient rehabilitation investigating the effectiveness of a function-centred treatment compared with a pain-centred treatment with 3 and 12 month follow-up for working status

[18,19]. A total of 170 out of the 174 patients who participated in the randomised controlled trial were included in paper I. All measurements were taken by a blinded research assistant; work status was assessed with questionnaires sent to employers and the patients' primary physicians, who were blinded to the patients' group assignments. Internal consistency of the SFS was assessed with Cronbach's alpha. A principal component analysis was performed to investigate unidimensionality. SFS scores were correlated with fear avoidance beliefs (FABs), pain intensity, physical factors, and FCE. Predictive validity for work status at 3 and 12 month follow-up was investigated with Receiver Operating Characteristic (ROC) curve analysis and responsiveness by calculating Standardised Response Mean (SRM).

For papers II and III an analytical cross-sectional study was performed in three rehabilitation clinics in Switzerland. Included were 130 patients referred for physical fitness for work evaluation, with chronic NSLBP as their primary complaint. Evaluations were performed by two independent assessors. These were blinded to each other's results. The first assessors assessed 'grip strength' and psychosocial factors including 'nonorganic-somatic-components'. The second assessors undertook FCE tests and determined physical effort level during the lifting tests by applying observational criteria. Paper II analysed the influence of psychosocial and physical factors on FCE performance with robust regression analysis. Paper III investigated the contribution of 'nonorganic-somatic-components' and 'submaximal effort' to lifting performance with linear regression analysis and the concurrent validity of 'nonorganic-somatic-components' with 'submaximal effort' by calculating sensitivity and specificity.

Paper IV consists of a systematic review and meta-analysis. Studies were included if randomised controlled trials were performed, the primary diagnosis in all patients was non-acute NSLBP with a duration of at least four weeks, the experimental treatments used exercise alone or as part of a multidisciplinary treatment, work disability was the primary outcome, and if at least 90% of the patients under treatment were available for the job market in that they were either employed or unemployed but seeking work. Data sources were MEDLINE, EMBASE, PEDro, Cochrane Library databases, NIOSHTIC-2, and PsycINFO until August 2008. Work disability data were converted to odds ratios. Random effects meta-analyses were conducted.

Results

Paper I revealed a high internal consistency of the SFS (Cronbach's alpha = 0.98) and reasonable evidence for unidimensionality. Correlations of the SFS with work activities were high (Spearman's rho > 0.6). ROC curve analysis revealed discriminating power for work status at 3 and 12 month by (area under curve =0.760, 95%CI: 0.689-0.822 resp. 0.787, 95%CI: 0.712-0.851). SRM within the two treatment groups was 0.18 and -0.31.

Paper II showed that between 42% and 58% of the variation in the FCE tests was explained in the final multivariate regression models. 'Nonorganic-somatic-components' were consistent independent predictors for all tests. Their influence was most important on forward bend standing and walking distance, and less on grip strength and lifting performance. PFA for work tasks was the most important predictor for lifting performance and also contributed significantly to grip strength and forward bend standing. In paper III 'nonorganic-somatic-components' and 'submaximal effort' were found to be independent contributors to lifting performance during FCE. The contribution of 'submaximal effort' was higher than that of 'nonorganic-somatic-components', shown by a higher change of coefficients ranging between 42–58% when 'submaximal effort' was added to the model compared to 14–17%

when 'nonorganic-somatic-components' was added. Between 53%-63% of the patients with 'nonorganic-somatic-components' were classified as showing 'submaximal effort'.

23 trials met the inclusion criteria for paper IV. Three studies were excluded from meta-analysis as the presented data did not allow pooling. The remaining 20 studies allowed 17 comparisons of exercise interventions with usual care and 11 comparisons of two different exercise interventions. A statistically significant effect in favour of exercise on work disability was found in a long term follow-up (OR = 0.66, 95% CI 0.48 – 0.92) while this was not the case in the short term (OR = 0.80, 95% CI 0.51 – 1.25) and in the intermediate term (OR = 0.78, 95% CI 0.45 – 1.34). Meta-regression indicated no significant effect of specific exercise characteristics.

Conclusions

PFA for work tasks can be validly assessed with the SFS in a European rehabilitation setting in patients with NSLBP and is predictive for future work status. PFA together with 'nonorganic-somatic-components' should be considered for interpretation of FCE. 'Nonorganic-somatic-signs' testing and determination of physical effort by observational criteria should not be interchangeably used for interpreting lifting performance during FCE. Exercise interventions have a significant effect on work disability in patients with non-acute NSLBP in the long term. No conclusions can be drawn regarding exercise types.

Key words: Low back pain, fitness-for-work, work-related rehabilitation, exercise, systematic review, functional capacity evaluation, physical effort, lifting performance, behaviour, nonorganic-somatic-components, physical performance, perceived functional ability

1.   Introduction

1.1   My background and motivation for writing this thesis

I have been working for many years in a rehabilitation centre confronted with work disability due to nonspecific diseases. According to the general aims of rehabilitation, our interventions strive to maximise function and minimise the limitation of activity and the restriction of participation in these patients of which RTW is of major importance. We have been able to show that clinical tests allow a prediction of RTW of patients with chronic LBP after inpatient rehabilitation [20,21]. We developed a specific function-centred rehabilitation program [22] which showed its effectiveness compared to a pain-centred rehabilitation program in returning patients with non-acute NSLBP to work [18,19]. However, the total costs of these two treatment approaches were similar over the whole 3-year follow-up [23]. We found that FCE improves quality and information regarding working capacity of medical fitness for work certificates in these patients [24]. Our clinical research showed the implications of self PFA for work tasks on assessment and rehabilitation interventions [18] but also the difficulties in assessing it in a rehabilitation setting where patients from various European nations are treated [25]. We are increasingly asked to evaluate fitness for work in patients referred for inpatient rehabilitation to plan RTW but also for disability determination. Patients with nonspecific disease such as LBP, in particular, require an evaluation of functional capacity as medical findings alone are insufficient to determine fitness for work. Such an approach is in line with Swiss legal requirements to judge inability to work by the extent of the functional loss regarding the demands of the previous work. However, FCE's results are in some cases minimal and can hardly be explained by pathological findings. In such cases, external findings are needed to interpret the test results. Exercise is a major intervention of our work-related rehabilitation approach. The question arises whether there are specific exercise characteristics we should use. The above outlined issues motivated me to perform further research about work-related evaluation and rehabilitation of patients with non-acute NSLBP.

1.2   The magnitude of the problem

1.2.1   Prevalence of low back pain

Low back pain (LBP) is known to be a very common problem in western societies[4,9] and increasingly in developing nations[26-29]. However, there is a wide range of prevalence rates found. A systematic review of population prevalence studies of LBP published between 1966 and 1998 reports point prevalence ranging from 12% to 33%, 1-year prevalence ranging from 22% to 65%, and lifetime prevalence ranging from 11% to 84% [30]. A systematic review on the prevalence of LBP among adolescents reports increasing lifetime prevalence with age approximating adult levels by around the age of 18 years [31]. Methodological differences among epidemiological studies such as differences in study design, mode of data collection, duration of LBP, lack of agreement on a clear definition of LBP, and patient age are repeatedly reported to be the cause of these differing prevalence rates [30,32,33]. Markedly

different prevalence rates of LBP are also reported between different western countries. Intercultural differences in perceiving or reporting back pain were hypothesised as the most likely explanation of these differences [34].

Despite these uncertainties on the true prevalence rate of LBP, recently published data continues to confirm that LBP is a common disorder. A review and analysis of data from two national U.S. surveys in 2002 found that LBP lasting at least a whole day in the previous 3 months was reported by 26.4% of respondents [35]. A cross-sectional, telephone survey of a representative sample of North Carolina households conducted in 1992 and repeated in 2006 revealed a significant increase in the prevalence of chronic LBP from 3.9% to 10.2% [36]. In Switzerland, back pain is currently the most prevalent health problem. 43% - 50% of the surveyed population reported various back problems in the preceding 4 weeks [6,37]. Women were more frequently affected then men [37]. This is comparable to prevalence rates found in Norway, where 45% of women and 38% of men reported LBP within a 14 day period [38].

1.2.2 Sickness, disability and work

An epidemiological study on LBP published in 1991 in the US already reported an increasing rate of disability due to LBP [39]. The 5% of people who became temporarily or permanently disabled from back pain caused 75% or more of the resulting costs, according to the authors, a phenomenon that seems more rooted in psychosocial rather than disease determinants. In Switzerland the number of pensioners from 1993 until 2003 has grown annually by an average of 3.5% [40]. Musculoskeletal disorders including LBP were one of the two causes for disability with the biggest annual growth [41] and are, at 31%, the second largest reason to receive a disability pension [42]. In 1995 and 1996, the estimated overall one year incidence of LBP in patients from the general working population in Norway who take at least 2 weeks of compensated absence from work was 2.27%. It was significantly higher for women (2.72%) than for men (1.91%). Approximately 42% of those still off work after 6 months had not returned to work after 12 months, and were switched to permanent disability pension or other compensation forms [43]. In 2006 about a third of sickness absences in Norway were due to musculoskeletal conditions amounting to 41% of days lost due to sickness absence [5].

1.2.3 Costs of low back pain

Spending on disability benefits has become a significant burden to public finances in most OECD countries. Public spending on disability benefits totals 2% of GDP on average across the OECD, rising to as much as 4%-5% in countries such as Norway, the Netherlands and Sweden [7]. Numerous studies report on the enormous costs of back pain. A systematic review on the costs of LBP identified 153 studies published in English up to 2007, of which 27 were deemed as relevant [44]. Estimates of the economic costs worldwide varied greatly depending on study methodology. 9 studies reported total costs which were for Australia AUD 9.2 billion, Belgium €1.2 billion, Japan Yen 6 billion, Netherlands €6.4, Sweden €1.9 - €3.3 billion, and the United Kingdom £12.3 billion. Proportions of indirect costs ranged between 34%-97% of the total costs [44]. The most recent study on the costs of LBP in Switzerland [6] estimated direct costs of LBP at €2.6 billion and direct medical costs at 6.1% of the total healthcare

expenditure. Productivity losses were estimated at €4.1 billion with the human capital approach and €2.2 billion with the friction cost approach. Presenteeism (reduction of productivity whilst still being at work) was the single most prominent cost category. The total economic burden of LBP to Swiss society was estimated between 1.6% and 2.3% of the GDP. Thus, it has been concluded that back pain is not only a major medical problem but also a major economical problem [44,45]. Concerns about the high public spending on disability has led to policies focusing on increased employment opportunities for individuals with disabilities. The health care provider is addressed within these policies by the demand to assess people's work capacity rigorously and refocus sick workers on early RTW [7].

1.3 Focus of this thesis

The first major review on the assessment and management of back pain had already stated in 1987 that not pain relief but improvement of function, including RTW, was the primary goal in the treatment of LBP [46]. This goal has become a consistent feature of modern treatment guidelines for subacute and chronic NSLBP recommending staying active, the use of exercise and early RTW [12]. RTW requires a thorough medical assessment including physical, psychosocial and work-related factors [9,12,47] including an evaluation of fitness for work [7].

The overall aim of this thesis was to investigate work-related assessments in patients with non-acute NSLBP. In addition, the use of exercise to achieve RTW in these patients was investigated.

1.4 Assessment of patients with low back pain

1.4.1 Medical evaluation

Two independent comparisons of national clinical guidelines for the management of LBP published in 2001 and 2003 [48,49] found consistent recommendations that diagnostic procedures should focus on the identification of red flags indicative of specific spinal pathology (see Table 1) and consider psychosocial factors as risk factors for the development of chronic disability. Koes et al. [12] updated their review in 2010 and included national clinical guidelines from 13 countries as well as 2 international clinical guidelines from Europe published from 2000 until 2008. The authors concluded that all guidelines continued to recommend diagnostic triage and screening for psychosocial factors.

1.4.1.1 Diagnostic triage

The types of physical examination and physical tests that are recommended within national guidelines for the management for LBP show some variation. History taking, neurologic examination, and straight leg raise testing are consistently recommended and X-ray examination as routine use at the initial visit discouraged [12]. Some guidelines (i.e. Australian and European) recommend imaging at the initial visit only for cases of suspected specific spinal pathology (see Table 1).

Table 1: Red flags indicative of specific spinal pathology [50]

- Age of onset <20 or >55 yr
- Violent trauma
- Constant progressive, non mechanical pain (no relief with bed rest)
- Thoracic pain
- Past medical history of malignant tumour, systematic steroids or drug abuse, HIV
- Prolonged use of corticosteroids
- Drug abuse, immunosuppression, HIV
- Systematically unwell
- Unexplained weight loss
- Widespread neurology (including cauda equina syndrome)
- Structural deformity

Known causes of specific LBP are infection, tumour, osteoporosis, fracture, structural deformity, inflammatory disorder, radicular syndrome or cauda equina syndrome [8-10].

Evidence suggests that fewer than 15% of individuals with LBP can be assigned to one of these specific LBP categories [39]. Most patients suffer from NSLBP [8-10]. NSLBP is defined as LBP not attributable to a recognisable, known specific pathology [10]. Due to these diagnostic difficulties the recommendation is to classify LBP according to the duration and localisation of symptoms [11] using the following criteria: Acute LBP (< 4 weeks); Subacute LBP (4–12 weeks), and Chronic LBP (CLBP) (> 12 weeks). 4 diagnostic groups are differentiated: 1) LBP with no radiation; 2) LBP radiating no further than the knee; 3) LBP radiating beyond the knee, with no neurological signs; 4) LBP radiating to a precise and entire leg dermatome, with or without neurological signs. In diagnostic group 4, if neurological signs are prevalent, nerve root compression must be assumed.

1.4.1.2 Assessment of psychosocial factors

There is considerable variation among national guidelines for the management of LBP in the amount of detail given on the assessment of psychosocial factors or the optimal timing of the assessment [12,48]. Several guidelines for the management of LBP [51-53], including the European Guidelines for the management of chronic NSLBP [9], specifically recommend the assessment of psychosocial Yellow Flags. Psychosocial Yellow Flags in patients with acute LBP indicate psychosocial barriers to recovery and, therefore, need to be addressed so that the risk of developing long-term disability and work loss can be reduced [54].

The term 'psychosocial' as used within the 'psychosocial Yellow Flags' refers to the interaction between the person and his or her social environment, and the influences on his or her behaviour [1]. Family members, co-workers, employers, the compensation system and health professionals form the social environment. Any of these people have the potential to affect a person with back pain and may influence behaviour, levels of distress, attitudes and beliefs, and experience of pain. The biopsychosocial model of back pain and disability emphasises the interaction between these multiple factors (see Figure 1).

Figure 1: The biopsychosocial model of back pain and disability [50]
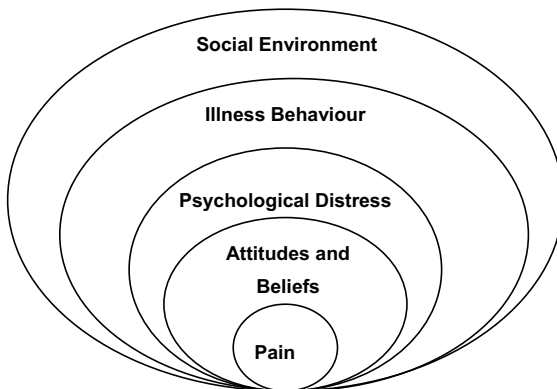
According to Kendall et al. [1,55] psychosocial Yellow Flags can be identified with a structured acute LBP screening questionnaire or a guide for the clinical assessment based on the following 7 headings: Attitudes and Beliefs about Back Pain, Behaviours, Compensation Issues, Diagnosis and Treatment, Emotions, Family, and Work. The authors do not intend this guide for clinical assessment to be a rigid prescription: "It is thought to allow flexibility and choice, thereby allowing the exercise of good clinical judgement according to the particular circumstances of the patient" [1]. A recently published reappraisal of the identification and management of psychosocial yellow flags emphasises that targeting yellow flags, particularly when they are at high levels, does seem to lead to more consistently positive results. However, questions remain about which factors are the most important, and how they can be identified clinically in relation to timing, necessary skills, and context [56].

1.4.2 Perceived functional ability for work tasks

Perceived functional self-efficacy is a relevant psychosocial factor contributing to the outcome in patients with chronic musculoskeletal pain [57]. According to Bandura, perceived self-efficacy affects how people behave in difficult situations. People who doubt their capabilities shy away from tasks which they view as personal threats [58]. Patients with back pain tend to have experienced difficulties with manual material handling [59], which may affect their perceived functional ability (PFA) for work tasks. The assessment of PFA for work tasks in patients with NSLBP plays an important role during rehabilitation and is proposed to be a predictor for RTW [60].

Assessment of PFA depends on the use of questionnaires. However, self-reported measures require an adequate literacy level [61] and depend on linguistic abilities. Text-based questionnaires are often impossible to administer in European rehabilitation settings for the treatment of patients with different mother tongues. A possible approach to overcome this problem is the use of picture-based questionnaires such as the Spinal Function Sort (SFS) [60]. The SFS has shown a high practicability in rehabilitation settings where patients from pan-European origin are treated, and is recommended for work-related rehabilitation [62].

The clinical utility of the SFS has so far only been reported in patient samples from the USA [63,64]. The reliability and validity of the SFS was investigated in English speaking patients with back pain from the United States [60] and Australia [65] reporting good psychometric properties of the SFS. No studies have been performed investigating the validity of the SFS in European patients.

1.4.3 Fitness for work evaluation and disability determination

Self-certified sickness absence is possible in some countries while others require a certificate from the treating doctor from the first day of illness [7]. Long term sickness absence always has to be certified by a medical doctor. Such certificates are frequently based on a medical examination and the resulting medical diagnosis [66]. However, fitness-for-work certificates based on diagnosis are criticised on the grounds that few objective physical or biomechanical measures are associated with RTW [13,14,16] and a person's potential work ability is not explored [17]. Medically driven judgements of fitness for work are also accused of being the major cause for the significant increase in the number of disability beneficiaries in the past two decades [7].

In many countries, there are now reforms in progress to move disability determination away from an 'essentialist' diagnostic approach [17] towards an evaluation of functional capacity [7]. Following comprehensive reform in 2003, disability assessment in Denmark now focuses on the person's remaining functions and the possible jobs the person can still perform. Similarly, in the Netherlands disability assessment is based on the person's functional abilities which are matched to job requirements [7]. With the introduction of the new Social Insurance Act in 2004 in Norway, a GP has to evaluate functional capacity [67] if a worker is off work for more than 8 weeks. Swiss law requires a physician to judge 'inability to work' by the extent of the functional loss regarding the demands of the previous work and not by the medical diagnosis. In the case of long-term inability to work (> 3 months), a judgement concerning a reasonable occupation in another profession or field has to be made [68]. The Department of Work and Pensions in the United Kingdom introduced the work capability assessment [69], which employs functional tests to determine 'fitness for work'.

Information from functional tests is regarded by insurance physicians as being of complementary value to their assessment of claimants with musculoskeletal disorders [70] and positively influencing quality and information of medical fitness for work certificates in patients with chronic LBP [24].

1.4.4 Functional Capacity Evaluation

Functional tests purporting to measure a patient's physical capacity to perform work tasks are employed within FCE. Many FCE systems are available but criticised for not having been rigorously analysed according to their psychometric properties [71]. While an early review published in 1999 found only limited scientific evidence for the reliability and validity of FCE [72,73], there is now an increasing body of knowledge on their clinimetric properties. A systematic review comparing 4 FCE methods concluded that the interrater reliability and

predictive validity of the Isernhagen FCE [74] is good while the procedure used in the test-retest studies was not rigorous enough to allow any conclusion [75]. In an update of their systematic review published in 1999, the authors' state that the Isernhagen FCE had the most comprehensive coverage of all aspects of reliability and validity [76].

In the FCE as described by Isernhagen [74], 28 physical tests are administered over two days. The 28 tests can be categorised in manual handling capacity tests, work postures, hand capacity and ambulation. Physical capacity determined by FCE is compared with the required physical job demands of the patient's occupation and recommendations for participation in work are made [77]. Critical job demands are assessed by a job analysis involving collecting relevant information by direct observation, an interview with employer or employee, or existing job descriptions.

The FCE approach is in line with the guidelines of the International Labour Organization [78] that refer to the assessment of fitness for work of diseased or disabled persons. These guidelines state that two major risks in the assessment of fitness for work must be avoided. The first is to overestimate functional disability by failing to allow for any adaptation of the job to the worker, while the second is to underestimate an intelligent and determined person's ability to overcome a disability and produce satisfactory results in a job that might be beyond such determination.

Evidence regarding the predictive validity for RTW of the Isernhagen FCE is contradictory. A study found in 650 adults of working age that the more weight they lifted from floor to waist during an FCE, the more likely was RTW [79]. Contrary to this finding, another study revealed that a better performance during FCE was only weakly associated with faster recovery, defined by shorter time until suspension of total temporary disability benefits and claim closure[80]. In a different study, the authors were unable to confirm the hypothesis that FCE is able to determine readiness or ability for safe RTW following musculoskeletal injury. A lower number of failed FCE tasks was consistently associated with higher risk of recurrence defined upon restarting temporary disability benefits [81]. However, the use of total temporary disability suspension and claim closure as an accurate substitution parameter for RTW and the use of restarting total temporary disability benefits in the year following FCE as an adequate indicator for 'recurrence' were questioned [82]. A recent study evaluating the quality of FCE information in predicting RTW found a significant relation between these two variables but a poor predictive efficiency [83].

1.4.4.1 The role of 'nonorganic-somatic-components' in Functional Capacity Evaluation

Physical factors such as age and gender have shown their association with lifting performance during FCE in manual material handling tests [84,85] as well as in non-manual material handling tests [85-88]. However, there is increasing evidence that not only physical but also psychosocial factors influence FCE results. Performance during FCE is associated with pain intensity [89], perceived disability [89-91], and functional self-efficacy [92].

A previous study [93] investigating factors influencing results of FCE in Workers' Compensation Claimants with LBP found that FCE reflects physical capacity to some degree but also found influences of perceived disability and pain intensity. It was therefore proposed that FCE should be considered as behavioural tests influenced by multiple factors, including physical ability and psychosocial factors. However, the authors state that there must have been some important determinants of physical performance that were not measured as they were unable to explain large amounts of the variation in FCE performance. A possible confounding factor related to FCE performance might have been the presence of 'nonorganic-somatic-components' within the physical examination.

Waddell et al. 1980 [94] described eight 'nonorganic-somatic-signs' that are distinguishable from the standard clinical signs of physical pathology in patients with LBP. Multiple signs suggest that a patient does not have a straightforward physical problem, but that illness behaviour and psychological factors also need to be considered. 'Nonorganic-somatic-components' correlate with illness behaviour and distress [94,95], as well as with increased disability [96-100] and a poorer rate of RTW [21,101].

The identification of 'nonorganic-somatic-components' within the physical examination has a long standing history in medical examination but is debated. A major criticism is that such findings are frequently interpreted as evidence of malingering [102-104]. In a later published reappraisal of the interpretation of their 'nonorganic-somatic-signs', the authors emphasise that these should not be used as evidence of simulation for the purpose of financial gain, but should be viewed as a form of communication between patient and examiner [105]. Such a behavioural response to examination is influenced by expectations and must be understood in the context of the patient's history.

To our knowledge, the influences of 'nonorganic-somatic-components', together with physical and other psychosocial factors, on the results of an FCE have not yet been investigated. We hypothesised that that the inclusion of 'nonorganic-somatic-components' as an explanatory variable will substantially increase the explained variation in FCE performance.

1.4.4.2 Effort determination during lifting tests

Effort determination has been widely used to validate the findings of physical performance tests [102,106-110]. Results from physical performance tests that are biased by 'submaximal-effort' may lead to false classifications of disability and consequently incorrect care as well as unwarranted disability compensation [106,111].

Various research has been performed within the fields of effort determination during muscle testing. Two literature reviews found a total of 61 studies investigating a wide variety of methods used to determine 'submaximal-effort' [111,112]. Robinson et al. concluded that, despite some promising aspects of methods examining motion variability, radial/ulnar force output ratios, difference scores of eccentric-concentric ratios, and electromyography, there is not sufficient empirical evidence to support the clinical application of muscle testing for this purpose [111].

Physical effort determination is also attempted during FCE [77,106-108,113] in order to interpret the performance results. Maximum effort of the client is required to obtain valid results in these physical performance tests [77]. The Isernhagen FCE uses observational criteria for physical effort level determination during manual handling tests to judge the physical demands and consequently the weight load as 'light – to moderate', 'heavy' or 'maximal' [77]. 'Submaximal-effort' is assumed if a patient stops the manual handling test before the criteria indicative of a maximum weight are observed [77,114]. 'Submaximal-effort' [107,111,112] classification has shown to be associated with decreased functional performance.

The eight 'nonorganic-somatic-signs' were not intended to determine physical effort [105] but have been used as a means of effort determination [102,108,113]. To our knowledge, it has not been investigated whether 'nonorganic-somatic-components' identified by 'nonorganic-somatic-signs' testing and 'submaximal-effort' determined by observational criteria during FCE contribute independently to lifting performance. It is also not known whether patients identified by 'nonorganic-somatic-signs' testing as presenting with 'nonorganic-somatic-components' will be classified by observational criteria as showing 'submaximal-effort'.

1.5 Work-related rehabilitation for patients with low back pain

Work-related rehabilitation is multidisciplinary consisting of a combination of physical, vocational, and behavioural components, and the modification of medication use. Many different terms such as multidisciplinary biopsychosocial rehabilitation, behavioural programmes, back schools, functional restoration programmes, work hardening or work conditioning are used to describe such a treatment approach [9]. The content of such rehabilitation programs corresponds to the recommendations found in the updated review of national and international guidelines on the management of NSLBP [12]. The authors concluded that for chronic LBP, consistent features included multidisciplinary treatment, cognitive behavioural therapy and supervised exercises. There were discrepancies for recommendations regarding spinal manipulation and drug treatment for acute and chronic LBP.

1.5.1 Multidisciplinary treatment

A systematic review [115] on the effectiveness of multidisciplinary rehabilitation for subacute LBP among working age adults found only two relevant studies that satisfied the criteria on subacute LBP. Based on these two studies of low methodological quality, the authors concluded that there was moderate evidence of positive effectiveness of multidisciplinary rehabilitation helping patients with subacute LBP to RTW faster, resulting in fewer sick leaves and alleviating subjective disability, and that a workplace visit increased the effectiveness. However, there was an obvious need for high quality trials in this field [115].

Two systematic reviews investigated the effect of multidisciplinary biopsychosocial rehabilitation on clinically relevant outcomes [116,117] and work status [117] in patients with chronic LBP [116]. Both found evidence for the effectiveness of such an approach compared to less intensive interventions [116] and to management strategies that do not include physical

conditioning programs [117]. However, the European guidelines for the management of chronic NSLBP require further research to define the optimal content of multidisciplinary treatment programmes [9].

1.5.2 Behavioural therapy

Behavioural therapy involves procedures where changes in the cognitions and behaviours are attempted in order to reduce disability [9]. Three behavioural approaches are generally distinguished: operant, cognitive, and respondent; but they are often combined as a treatment package [118]. The consistent recommendation of behavioural therapy among the different national guidelines [12] is only sparsely supported by scientific evidence. A systematic review on behavioural treatment for chronic low-back pain (CLBP) [119] found that combined respondent-cognitive therapy and progressive relaxation therapy are more effective than waiting list control on short-term pain relief. No significant differences could be detected between behavioural treatment and exercise therapy. Whether clinicians should refer patients with CLBP to behavioural treatment programs or to active conservative treatment could not be concluded from this review. An update of this review published in 2010 [118] came to similar findings. There was evidence of moderate quality that in the short-term, operant therapy is more effective than waiting list and that behavioural therapy is more effective than usual care for pain relief, but no specific type of behavioural therapy is more effective than another. In the intermediate- to long-term, there was little or no difference between behavioural therapy and group exercises for pain or depressive symptoms.

1.5.3 Exercise therapy

Koes et al. [12] found consensus among the guidelines considering subacute LBP and CLBP to use exercise but note that there is no evidence that one form of exercise is superior to another. This is in line with a previous review published in 2000 [11] which found no evidence for the effectiveness of specific exercises in the management of CLBP. The authors stated: 'it appears that the key to success is physical activity itself—i.e. activity of any form—rather than any specific activity'.

Exercises applied in the treatment of patients with NSLBP encompass a wide variety of interventions and are applied with different rationales. The sports medicine approach applies exercise based on the principles of exercise physiology, and is used in functional restoration programs with the goal of restoring physical function and thereby enabling patients to RTW [120]. Behavioural treatment programs use exercise with the aim of modifying pain behaviour. Patients learn that it is safe to move while restoring function by receiving continuous feedback and positive reinforcement [121].

While until 2000 no evidence had been found for the effectiveness of specific exercises in the management of CLBP [11], a later review revealed that individually designed stretching or strengthening exercises delivered with supervision may improve pain and function in chronic NSLBP. The authors recommended further testing with this multivariable model and further assessment with specific patient-level characteristics and exercise types [122].

A systematic review of trials with positive outcomes on work disability [117] revealed that all had significant cognitive behavioural components combined with intensive physical training. The authors, however, advised caution when interpreting this post hoc analysis and recommended further investigation into the contribution of these exercise characteristics.

Whereas additional reviews found limited evidence for the effectiveness of behavioural graded activity in improving absenteeism outcomes [119,123], strong evidence has been found that exercise reduces work disability in patients with NSLBP [117,124,125]. These reviews were based on studies published before 2004 that did not evaluate the effectiveness of different exercise characteristics.

Although new studies have been published in the meantime, the effect of specific exercise characteristics on work disability is still unclear; a more up to date review is required. The questions arise as to whether exercise is more effective than usual care to reduce work disability in patients with non-acute NSLBP, and if so, to explore which type of exercise is most effective.

2. Aims of the thesis

The overall aim of this thesis was to investigate work-related assessments in patients with non-acute NSLBP. In addition, the use of exercise to achieve RTW in these patients was investigated. The specific aims were:

- To test the validity of the Spinal Function Sort assessing self-perceived functional ability for work tasks by evaluating internal consistency, unidimensionality, concurrent and predictive validity, and responsiveness in a European rehabilitation setting for patients with non-acute nonspecific low back pain.   **Paper I**

- To investigate the influences of ''nonorganic-somatic-components'', together with physical and other psychosocial factors, on the results of a Functional Capacity Evaluation in patients with chronic nonspecific low back pain undergoing physical fitness for work evaluation.   **Paper II**

- To assess the contribution of 'nonorganic-somatic-components' and 'submaximal-effort' to lifting performance and to determine the concurrent validity of the 'nonorganic-somatic-components with 'submaximal-effort' during FCE in patients with chronic nonspecific low back pain.   **Paper III**

- To determine whether exercise is more effective than usual care in reducing work disability in patients with non-acute nonspecific low back pain, and if so, to explore which type of exercise is most effective.   **Paper IV**

3. Materials and methods

3.1 Designs

The different aims of the thesis required different research designs and materials. For paper I, a prospective cohort study was used. This was embedded within a randomised controlled trial performed during inpatient rehabilitation investigating the effectiveness of a function-centred treatment compared with a pain-centred treatment with 3 and 12 month follow-up for working status [18,19]. For papers II and III an analytical cross-sectional study was performed in three rehabilitation clinics in Switzerland. Paper IV consists of a systematic review and meta-analysis.

3.2 Patients and materials

Patients with non-acute NSLBP within working age, presenting with at least 6 weeks of sick leave, and with sufficient understanding of German, French or Italian were included in all studies. The following Table 2 gives an overview on the patient characteristics.

Table 2: Study populations of the different patient samples
Papers II and III investigated a common patient sample of 130 patients referred for fitness for work evaluation. Paper II included only 126 out of 130 due to missing work data.

	Paper I	Paper II	Paper III
	n=170	n=126	n=130
Gender			
Male n (%)	133 (78%)	94 (75%)	97 (75%)
Female n (%)	37 (22%)	32 (25%)	33 (25%)
Age (years)	42 (8)	44 (10)	44 (10)
Days out of work	200 (152.7)	670 (1031)	670 (1031)
Self-rated ability for work tasks (SFS)	104.9 (46.1)	96.3 (50.9)	95.6 (51.2)
Fear avoidance belief – work (FABQ)	32.1 (9.6)	32.7 (9.0)	32.8 (9.0)
Momentary pain intensity (NRS 0–10)	5.6 (2.3)	5.1 (2.2)	5.1 (2.2)

Values are mean (SD) for continuous variables and numbers (%) for categorical variables

3.2.1 Patients referred for inpatient work-related rehabilitation (paper I)

Eligible were patients who were referred to the Rehabilitationsklinik Valens (Switzerland) for in-patient rehabilitation. Patients included in paper I participated in a randomised controlled trial investigating the effectiveness of a function-centred treatment compared with a pain-centred treatment with 3 and 12 month follow-up for working status [18,19]. A total of 170 out of

the 174 subjects who participated in the randomized controlled trial fully completed the SFS at discharge and were included in this study.

3.2.2   Patients referred for fitness for work evaluation (papers II and III)

Included were patients referred for physical fitness for work evaluation, presenting with chronic NSLBP as their primary complaint.

From March 2009 until August 2010 678 patients were referred to three rehabilitation clinics in Switzerland for FCE. All clinics are national competence centres for FCEs, each performing more then 100 FCEs per year. Physicians identified 203 subjects with LBP of which 26 suffered from specific back pain, and 16 of relevant comorbidity affecting work ability. 5 were excluded because of language problems, 11 subjects did not sign the informed consent, 2 were older than 60 years and 13 were missed for inclusion leaving 130 subjects with chronic NSLBP. Mean time off work was over two years. 52% of the included patients had performed heavy or very heavy work in their previous job.

3.2.3   Studies included in the systematic review (paper IV)

Studies were included if randomised controlled trials were performed, the primary diagnosis in all patients was non-acute NSLBP with a duration of at least 4 weeks, the experimental treatments used exercise alone or as a part of a multidisciplinary treatment, work disability was the primary outcome, and if at least 90% of the patients under treatment were available for the job market in that they were either employed or unemployed but seeking work.

3.3   Ethical approval

The prospective cohort study and the cross-sectional study involving patient assessments were approved by the three regional ethical committees where the rehabilitation clinics are located (Request numbers: EKSG 03/35, EKSG 08/029/2B, SPUK N°. 784, EKAG 08/058).

3.4   Procedures and Measurements

3.4.1   Procedures

3.4.1.1 Prospective cohort study (paper I)

All measurements were taken by a research assistant blinded to the treatment received at entry and discharge after inpatient rehabilitation. Work status at 3 and 12 month follow-up was assessed with a questionnaire sent to employers and the patients' primary physicians, who were blinded to the patients' group assignments.

3.4.1.2 Cross-sectional study (papers II and III)

Evaluations were performed by two independent assessors. These were blinded to each other's results. The first assessor administered the questionnaires and assessed the independent variable 'nonorganic-somatic-signs' and 'grip strength'. The second assessor

undertook FCE tests, which consisted of forward bend standing, walking, lifting from 'floor to waist', 'waist to crown', and 'horizontal' and determined physical effort level during the lifting tests by applying observational criteria.

11 physiotherapists having a mean professional experience of 12.1 years (SD 6.9) acted as first assessors while 17 physiotherapists performed FCE. All FCE assessors completed a two day course on FCE and had extensive experience in FCE testing, having performed on average 36.5 FCEs (SD 10.0) in the 2 years prior to this study. In addition, for the purpose of this study a teaching person for FCE discussed 8 FCE video sequences with the FCE assessors regarding observational criteria for effort level during manual handling tasks and trained the assessment of the eight 'nonorganic-somatic' signs according to the description of Waddell et al. [94] with the first assessors.

3.4.1.3 Systematic review (paper IV)

The search strategy was based on the recommendations of the Cochrane Back Review Group [126]. We searched MEDLINE, EMBASE, PEDro, the Cochrane Library, PsycINFO (2002 – August 2008) and NIOSHTIC-2 (until August 2008). This search was combined with a previous search performed in December 2002 covering MEDLINE (1966-Oct. 2002), EMBASE (1988 - Oct. 2002), PEDro (until Dec. 2002), the Cochrane Library (2002, Issue 4) and PsycLIT (1984 -Dec. 2002) [124]. References were checked for further trials.

Studies were included if randomised controlled trials were performed, the primary diagnosis in all patients was non-acute NSLBP with a duration of at least four weeks, the experimental treatments used exercise alone or as a part of a multidisciplinary treatment, work disability was the primary outcome, and if at least 90% of the patients under treatment were available for the job market in that they were either employed or unemployed but seeking work. Excluded were studies that did not report work disability, investigated the effect of treatments that did not contain any form of exercise such as respondent psychological interventions, included patients with thoracic pain, cervical pain or specific LBP such as nerve root compression, vertebral fracture, tumour, infection, inflammatory diseases, spondylolysthesis, spinal stenosis and definite instability, and studies that included pregnant women with LBP.

Two authors (Oesch, Kool) independently applied the admission criteria for the studies and assessed risk of bias. Disagreements were solved through discussion involving a third researcher (Bachmann). Authors were contacted if the information regarding the eligibility of a trial, quality criteria, or work disability was unclear.

Study quality was assessed according to Juni et al. [127], who stated that the internal validity of an randomised controlled trial can be threatened by detection bias, attrition bias, selection bias, and performance bias. Thus, the following three criteria were rated as 'met', 'unclear' or 'not met': Concealed allocation, blinding of the outcome assessor, and intention to treat analysis. Performance bias was not assessed as it is not strictly possible to blind the treatment provider and recipient in clinical trials investigating the effect of exercise to treatment allocation. The internal validity of the included studies were then evaluated on methodological overall assessment: Studies were classified as high quality studies if two or

three of the criteria were met, while studies were classified as being of low quality if one or none of the criteria were met.

For each study, two of the authors (Oesch, Bachmann) independently extracted data from all included studies and defined exercise characteristics. Four criteria designed by Hayden et al. [122] were used, namely program design, delivery type, dose, and type. Additionally, two criteria proposed by Schonstein et al. [128] were used, namely work context and exercise administration within a cognitive behavioural approach. A further criterion was the setting in which exercise was applied.

3.4.2 Measurements

Patients were assessed with questionnaires, underwent a physical examination and an FCE. Work-related data were obtained from the clinical database, by interviews with patients, employers and primary physicians. An overview of the measurements is given in Table 3.

3.4.2.1 Work status

Days at work (paper I): As in Switzerland no national database exists to obtain work data information we had to resort to questionnaires sent to both the primary physicians and the employers at 3 and 12 month follow-up. Both were blinded to the treatment group. Each calendar day within a period at work was counted. This method is insensitive to the fact that patients work on different days of the week. Inconsistencies were resolved through additional phone calls to the people involved.

3.4.2.2 Physical factors

Back strength (paper I): Back strength was assessed with the Biering-Sorensen test [129]. The patient was lying prone on a treatment table supported up to the iliac crest. Their legs were strapped to the treatment table. Patients were instructed to hold the unsupported upper body horizontally with the arms crossed behind their back for as long as possible. During the test, no extra motivation was given. Maximum holding time was measured with a stopwatch. Reliability of the Biering-Sorensen test has shown to be high in patients with chronic LBP [130].

Finger to floor distance (FFD) (paper I): The patients were asked to reach with the fingertips of both hands as far as possible down towards their toes. The knees had to be kept straight. The distance between the finger tips and the floor was measured with a tape measure in centimetres. The Fingertip-to-floor test has shown excellent reliability in patients with LBP and is recommended for use in clinical practice and therapeutic trials [131].

Straight leg raise tests (SLR) (paper I): SLR was carried out with the patients in supine position; the head remained relaxed on the treatment table. The ankle was held with one hand making sure that the hip was in neutral rotation. A bubble inclinometer was positioned on the tibia crest with the other hand and set at zero. The leg was then raised passively by the examiner, whose other hand continued to hold the inclinometer in position and also held the patient's knee fully extended with his elbow. The leg was raised slowly to the maximum tolerated SLR (not the onset of pain). The maximum reading in degrees against a horizontal

line was recorded. The reliability of gravity dependent inclinometer in measuring leg movement against a horizontal line have been shown to be high [132].

Heart rate (HR) (paper III): HR was measured by Polar watch FT 1 in resting position and at the end of the lifting cycle.

Table 3: Overview of the measurements taken in papers I - III

Measurements		Paper I	Paper II		Paper III		
		Obs.	Dep.	Indep.	Obs.	Dep.	Indep.
Work status	Work days	▪					
Physical factors	Age	▪		▪			▪
	Gender	▪		▪			▪
	Back strength	▪					
	Finger to floor distance	▪					
	Straight leg raise	▪					
	Heart rate				▪		
Psychosocial factors	'Nonorganic-somatic-components'			▪			▪
	Perceived ability for work tasks	▪		▪			
	Fear avoidance belief – work	▪		▪			
	Fear avoidance belief – activity	▪					
	Pain intensity	▪		▪			
	Salary in the previous job			▪			
	Days out of work			▪			
Functional Capacity Evaluation	Floor to waist lifting	▪	▪			▪	
	Waist to overhead lifting	▪				▪	
	Horizontal lifting	▪				▪	
	Single handed carry right	▪					
	Single handed carry left	▪					
	Forward bend standing		▪				
	Grip strength		▪				
	6-Minute walking distance		▪				
	Submaximal effort						▪

Obs. indicates observational variable, Dep, dependent variable; Indep, independent variable

3.4.2.3 Psychosocial factors

'Nonorganic-somatic-components' (papers II, III): The eight 'nonorganic-somatic-signs' were used to assess 'nonorganic-somatic-components' (see Table 4). The eight 'nonorganic-somatic-signs' are grouped into five categories. A category is positive if at least one 'nonorganic-somatic-sign' in that category is positive. Three categories are required to be positive to indicate that a patient with LBP does not have a straightforward physical problem [94].

Table 4: The eight 'nonorganic-somatic-signs' [94]

Tenderness

1. 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.
2. 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

3. Axial Loading: LBP is reported on vertical loading over the standing subject's skull by the examiner's hands. Neck pain is common and should be discounted.
4. Rotation: LBP is reported when shoulders and pelvis are passively rotated in the same plane as the subject stands relaxed with the feet together. In the presence of root irritation, leg pain may be produced and should be discounted.

Distraction Test

5. Straight Leg Raising: Straight leg raising is the most useful distraction test. The subject whose back pain has a nonorganic component shows marked improvement in straight leg raising on distraction as compared with formal testing.

Regional disturbances

6. Sensory: Sensory disturbances include diminished sensation to light touch, pinprick, and sometimes other modalities fitting a "stocking" rather than a dermatomal pattern.
7. 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

8. Overreaction during examination may take the form of disproportionate verbalisation, facial expressions, muscle tension and tremor, collapsing or sweating. 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.

Perceived functional ability for work tasks (papers I, II): The SFS assesses PFA to perform work tasks that involve the spine in various ways. It was developed in the United States by Matheson et al. and consists of 50 graphically depicted tasks with simple descriptions [133]. The patient is instructed to look at each drawing and rate each task on a separate evaluation sheet on a 5 point scale from 1=able, 2-4= increasingly restricted, and 5=unable. The SFS is scored by the assessor and yields a single rating of PFA ranging from 0 to 200. Patients with an SFS score of less than 100 are categorised as perceiving minimal working capacity [60].

Fear avoidance belief for activity (paper I) and work (papers I, II): The Fear Avoidance Belief Questionnaire (FABQ) was used to assess how patients were affected by fear and avoidance beliefs for work activities (FAB). The questionnaire provides a score for fear of physical activity ranging from 0 (no fear) to 24 (maximum fear) and one for work activities ranging from 0 to 42 [134]. FAB's for work [135] have shown to be among the best predictors for RTW. Translated versions of the FABQ were available in German, Italian, Serbo-Croatian, Albanian, Turkish, French, Spanish and Portuguese.

Pain intensity (papers I, II): Pain intensity at the moment of examination was measured with numeric rating scale (NRS) (0 = no pain, 10 = worst pain possible). NRS has shown its reliability in literate and illiterate patients [136].

Compensation issues (papers II): Economic aspects have an impact on RTW [67,137]. Low salary in the previous job and days out of work were assumed to have a negative influence on FCE performance. We collected information about patients' salaries in their previous job and days off work from the clinical database.

3.4.2.4 Functional Capacity Evaluation

Manual handling tests were performed using a kinesiophysical approach as proposed by Isernhagen [77]. Within this approach, lifting and carrying performance is determined by using observational criteria indicative of physical efforts at 'light – to moderate', 'heavy' or 'maximum' levels of demand (see Table 5). Patients were instructed that the objective was to perform five lifts or carries with as much weight as safely possible. For safety reasons all tests were commenced with a light weight. Weights were then increased by the FCE therapist and five lifts repeated until maximum safe weight load was reached. The maximum safe weight load in kilograms was determined by the FCE assessor when the criteria indicating maximum level of demands were observed [77]. Heart rate was measured with a Polar watch. If the patient stopped the manual handling test before the criteria indicative of a maximum level of demands were observed the highest weight in kilogram that a patient was willing to handle five times was recorded.

The reliability of these observational criteria to judge lifting performance has been established in several studies based on video observation [138,139] or on direct observation by one [140-142] or more observers [143]. In this study, five experienced FCE assessors tested patients simultaneously, but independently. Their agreement in the application of the Isernhagen observational criteria to determine the maximal safe lifting performance during kinesiophysical FCE was excellent as shown by ICC values ranging from 0.95 to 0.98 [143]. One study using the observational criteria during video recordings of lifting from 'floor to waist' in patients with chronic LBP and healthy subjects serving as a control group concluded that the effort level can be determined validly by means of visual observation [107].

The following manual handling tests were performed:

Floor to waist lifting (papers I – III): Participants had to lift the weight receptacle from a shelf at waist height, make a 90 degree turn, lower the weight to the floor, stand up again lifting the weight, and then place it back on the shelf

Waist to crown lifting (papers I, III): Participants had to lift the weight receptacle from a shelf at waist height to a second shelf on shoulder height, and place it back on the shelf at waist height.

Horizontal lifting (paper I, III): Participants had to lift the weight receptacle from a shelf at waist height, walk 1.5m with the weight receptacle, place it on a second shelf at waist height, and walk back with the weight receptacle to the first shelf.

Single handed carry right and left (paper I): Participants had to carry a weight receptacle with one hand from a shelf at knee height, walk 15m with the weight receptacle, and place it on the shelf.

FCE employs the following tests for the assessment of work postures, hand and walking capacity:

Forward bend standing (paper II): Patients were instructed to maintain a 30 degree forward bending trunk position for 5 minutes without straightening the hips or back while performing a hand activity. Test performance in seconds was measured by the observer with a stopwatch. The test showed acceptable test-retest reliability in patients with chronic LBP [141].

Grip strength (paper II): Grip strength measurements in kilograms were taken with a hand held G200 Dynanometer from Biometrics Ltd. Version 8 software according to a standardised and reliable protocol [86,144] was used. Patients were instructed to squeeze the Dynanometer as hard as possible for three tests in alternating order with the right and left hand in 5 different handle positions. The mean grip-strength measurements of each handle position were calculated and the maximum grip strength was recorded.

Six minute walking distance (paper II): The six minute walking test was performed according to the recommendations of the American Thorax Society [145]. Participants were instructed that the objective was to walk as far as possible for six minutes, without running or jogging, by walking back and forth along a distance of 30 meters marked by traffic cones. Test performance was measured by the observer in meters walked. Reliability of the six minute walking test has been proven high in various patient groups [146,147].

Submaximal effort (paper III): 'Submaximal effort' was assumed if a patient stopped the manual handling test before the criteria indicative of a maximum level of demand were observed [77,114].

Table 5: Observational criteria for effort level during manual handling tests [77,114]
Lifting performance measured in kg was determined by using observational criteria indicative of physical efforts at 'light – to moderate', 'heavy' or 'maximum' levels of demand. 'Submaximal effort' was assumed if a patient stopped the manual handling test before the criteria indicative of a maximum weight were observed.

Observational criteria	Lifting performance		
	Maximum demand	Heavy demand	Light – moderate demand
Muscle recruitment			
Prime movers	Bulging	Bulging	Normal recruitment
Accessory muscles	Bulging	Distinct recruitment	No or just slight muscle recruitment
Base of support	Very wide base	Distinctly increased	Natural stance
Posture	Substantial counter balance	Distinctly increased counter balance	No or beginning counter balance in extension
Heart rate, respiration	Substantial increase in heart rate and respiration	Distinct increase in heart rate and respiration	No or minimal increase in heart rate and respiration
Control and safety	Still safe, but unable to maintain control if any more weight is added	Increasingly controlled movement; might begin to use momentum; execution with difficulty, but not yet at the limit	Smooth movements
Pace	Very slow (increased pace would affect stability and control)	Distinctly slower. Very deliberate movements	Moderate/comfortable pace

3.5   Statistical analyses

3.5.1   Sample size calculation

Paper I: We planned to use Receiver Operating Characteristic (ROC) curve analysis [148] to evaluate the diagnostic performance of the SFS scores at discharge for work status at 3 and 12 month follow-up. We wanted to show that the area under the curve of the SFS for work status was > 0.725 and is significant from the null hypothesis value 0.5 (meaning no discriminating power). With alpha defined as 0.05 and beta as 0.9 we calculated a minimal sample size of 86 subjects necessary for ROC curve analysis [149].

Papers II and III: We used the same approach for both papers to calculate sample size but under different conditions. For paper I we relied on the study of Gross et al. [93] and assumed R2 to be at least 0.15 giving an effect size of 0.18. With alpha defined as 0.05 and beta as 0.9 and 8 independent variables we calculated a minimum sample size of 114 subjects needed for multiple regression analysis. For paper II we assumed based on paper II [150] R2 to be at least 0.30 giving an effect size of 0.43. With alpha defined as 0.05 and beta as 0.9 and 4 independent variables we calculated a minimum sample size of 41 subjects needed for multiple regression analysis.

3.5.2   Statistical methods

Descriptive data analysis, principal component analysis, Spearman's rho correlation coefficients, and linear regression analysis were performed using SPSS for windows version 18. ROC curve analysis was performed with MedCalc version 9.7.3.0 and robust regression analysis with Number Cruncher Statistical System (NCSS, version 07.1.19, Kaysville, Utah). Stata statistical software (Stata Corporation, College Station, TX V10) was used to conduct DerSimonian and Laird random effects meta-analyses. The following Table 6 gives an overview on the statistical methods used in papers I - IV.

Table 6: Statistical methods

Statistics	Paper 1	Paper 2	Paper 3	Paper 4
Descriptive statistics	▪	▪	▪	
Cronbach's alpha	▪			
Principal component analysis	▪			
Spearman's rho correlation coefficient	▪			
Receiver Operating Characteristic	▪			
Standardised Response Mean	▪			
Robust regression analysis		▪		
Linear regression analysis			▪	
Sensitivity and Specificity	▪		▪	
Positive and negative predictive value	▪			
Random effects meta-analyses				▪
Meta-regression analysis				▪

3.5.2.1 Prospective cohort study (paper I)

Internal consistency: This was assessed by item-total correlations and Cronbach's alpha. For a scale to be sufficiently reliable for use in groups of patients, an alpha value of 0.80 is considered acceptable [151].

Unidimensionality: Principal component analysis of all 50 items of the discharge SFS score was used to assess the unidimensionality of the instrument. An eigenvalue criterion of 1.0 was used. Ratios of first to second eigenvalues of 3:1 are generally considered evidence for unidimensionality [152].

Concurrent validity: Concurrent validity was assessed by comparing the entry and discharge SFS scores with FAB's, pain intensity, FCE, and physical factors with Spearman's correlation coefficient. We hypothesised that the SFS scores would correlate highly to FAB's, pain intensity and FCE (i.e. r > 0.60) and moderately to physical factors (0.30 – 0.60). Bonferroni procedures were applied to reduce type I error in the assessment of concurrent validity. Adjustment for 12 comparisons at alpha = 0.05 resulted in the use of p < 0.005 as level of significance.

Responsiveness: Since the function-centred treatment was specially targeted towards improving perceived ability, it was hypothesised that these patients would have a greater improvement in SFS than the patients in the pain-centred treatment group. Responsiveness was, therefore, assessed in the 2 treatment groups separately, with the Standardised Response Mean (SRM) which was calculated by dividing the mean change scores by the standard deviation of the change scores.

Diagnostic properties: ROC curve analysis was used [148] to evaluate the diagnostic performance of the SFS scores at discharge for work status at 3 and 12 month follow-up. The diagnostic property of the proposed SFS of < 100 indicating minimal work ability [60] was assessed by calculating sensitivity and specificity as well as positive and negative predictive value. Within the context of work-related rehabilitation of patients with back pain, the aim is to identify those patients with a probability of not returning to work. Sensitivity refers, therefore, to the proportion of patients who had not returned to work and were correctly identified by the SFS score < 100 indicating minimal work capacity. Another interesting aspect for clinicians is the positive predictive value of the perceived functional work capability for not returning to work. We hypothesised that a perception of a minimal working capacity (SFS score < 100) would have a high positive predictive value identifying those patients who would not RTW at 3 and 12 month follow-up.

3.5.2.2 Cross-sectional study (papers II & III)

Paper II: Our main goal was to assess the association of 'nonorganic-somatic-components' with the 4 test categories of an FCE as described by Isernhagen et al. [153]. We included 'manual handling from floor to waist', thereby allowing a comparison with previous research investigating the influence of non-physical factors on lifting performance [93]. In addition, we wanted to investigate whether representative tests from the other FCE categories were also influenced by 'nonorganic-somatic-components' and chose forward bend standing, grip

strength and six minute walking distance due to their reported reliability. We chose to correct for the physical factors age and gender and for the psychosocial factors: PFA, FAB for work, pain intensity, days off work, and salary in the previous job in our analysis. Model building started with bivariate models for each dependent variable. We planned to retain all independent variables with p<0.2 in the first multivariate model and then remove the variables with the highest p-values. If a slope coefficient for a variable changed by more than 30% when removed, it was retained in the analysis as a confounder. The final model included the confounders plus the variables with p≤05. An examination of the models revealed violations of the normality as well as homoscedasticity assumptions. Consequently, we resorted to Robust regression analysis [154]. The analysis is an iterative procedure that seeks to identify the outliers and down-weighs their influence on the regression coefficient estimates. This analysis was performed using Number Cruncher Statistical System (NCSS, version 07.1.19, Kaysville, Utah). All other analyses were performed with Statistical Package for Social Sciences (SPSS Version 18). Data analysis was performed by an independent statistician who was not involved in conducting the study and data collection.

Paper III: The contribution of 'nonorganic-somatic-components' and 'submaximal -effort' to lifting performance was assessed with multivariate linear regression analysis. The dependent variables were the maximum weights handled during lifting from 'floor to waist', 'waist to crown', and 'horizontal'. The explanatory variables were 'submaximal-effort' and 'nonorganic-somatic-components'. We included age and gender in the model. We then added the explanatory variables stepwise and observed the change in coefficients if one variable was added over the other. Interaction of 'nonorganic-somatic-components' with 'submaximal-effort' was tested with product terms. An examination of the full models revealed that assumptions of the normality as well as homoscedasticity were respected.

We assessed the concurrent validity of the 'nonorganic-somatic-components' with 'submaximal-effort' by calculating sensitivity and specificity. Sensitivity within this study refers to the proportion of patients who presented with 'nonorganic-somatic-components' and were classified as giving 'submaximal-effort'. Specificity was calculated as the proportion of the patients not presenting 'nonorganic-somatic-components' and classified as giving maximal effort.

3.5.2.3 Systematic review (paper IV)

Work-related outcomes were converted into odds ratios (OR) using the method described by Chinn [155] and Hasselblad & Hedges [156]. This method is based on the fact that, when assuming logistic distributions and equal variances in the two treatment groups, the log OR corresponds to a constant multiplied by the standardised difference between means. The 'metan' command for Stata statistical software (Stata Corporation, College Station, TX V10) was used to conduct DerSimonian and Laird random effects meta-analyses [157]. To use all available means, we estimated missing standard deviations from other included studies. We assessed treatment effects at three different times of follow-up (short term = closest to four weeks, intermediate term = closest to six months, long term = closest to 12 months). Between-trial heterogeneity was quantified using the I^2 statistic, which can be understood as

the proportion of the total variation in estimated ORs that is due to between-trial heterogeneity rather than chance [158]. The extent to which one or more study characteristics explained between-trial heterogeneity was explored using meta-regression. The following explanatory variables were considered according to an a priori statistical analysis plan: Exercise design (individual vs. standard care), dose (high vs. low dose exercise), delivery type (home based exercises vs. supervised exercises), type (specific vs. mixed), administration within a cognitive behavioural approach (yes/no), work context (yes/no), and setting (in- vs. outpatient) in bivariate models. In addition, we assessed the effect of methodological quality (low vs. high). For work disability we included the variables above in meta-regression models and conducted random effects meta-analyses within each subgroup. Differences between small and large trials were assessed using funnel plots [159].

4. Main results

4.1 Perceived functional ability for work tasks

We investigated the validity of the SFS used for the assessment of PFA for work tasks in patients with NSLBP. This picture-based questionnaire has potential advantages for clinical use in a European rehabilitation setting treating patients with variate literacy levels and different mother tongues.

Descriptives: Patients were all of working age. There were high fear avoidance beliefs and low PFA for work tasks (see Table 2). 48% of the patients came from countries other than Switzerland (see Figure 2).

Figure 2: Nationalities
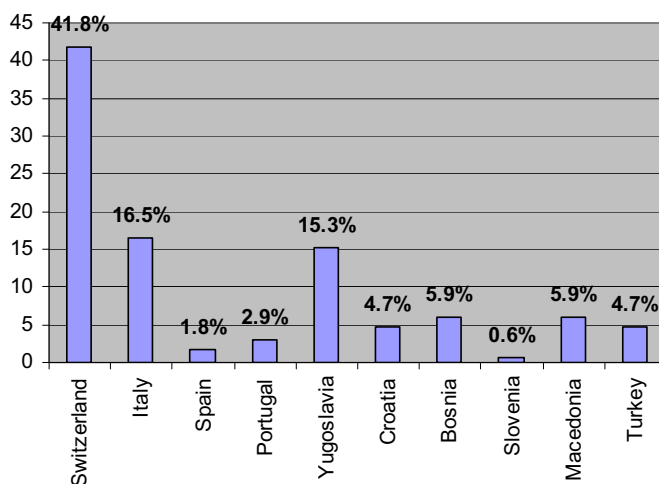
Internal consistency: We found a high internal consistency at entry and discharge shown by a Cronbach's alpha coefficient of 0.98. There was high agreement among the patients that they were unable to handle weights of 50kg as requested in 4 items (mean=4.9, SD 0.3). These four items showed a low correlation with the SFS total score (total item correlation ≤0.32). Alpha would not have been substantially changed if these items had been removed. The remaining 46 items all showed a high correlation with the SFS total score (total item correlation >0.6).

Unidimensionality: Principal component analysis revealed the presence of a mixed structure with 2 components showing a number of strong loadings. The four items asking for handling weights of 50kg loaded on different components than the other 46 items. The 2 component solution explained a total of 63.6% of the variance, with component 1 contributing 46.0% of the variance and component 2 contributing 17.6% of the variance. The ratio from first to second eigenvalue was 6.95, suggesting reasonable evidence for unidimensionality.

Concurrent validity: Complete measurements for the assessment of concurrent validity of the SFS with FAB's, pain intensity, FCE and physical factors at entry and discharge were available from 156 patients. Correlations were generally higher at discharge than at entry. Correlations between SFS and FCE were high (>0.6) except for manual handling from waist to crown. Back strength and pain intensity also showed high correlation with the SFS total score at discharge. FAB for work activities showed markedly higher correlation with the SFS score than FAB for physical activities.

Responsiveness: At discharge, patients in the function-centred treatment group (n=84) showed a mean increase of 5.9 points (SD 32.5) in perceived functional ability (SRM 0.18), while the patients in the pain-centred treatment group (n=85) showed a mean decrease of 7.4 points (SD 24.4) in perceived functional ability (SRM - 0.31).

Diagnostic properties: The area under the ROC curve (AUC) was 0.760 (95% CI: 0.689 - 0.822) for work status at 3 month follow-up and 0.787 (95% CI: 0.712 - 0.851) at 12 month follow-up. The difference in the AUC at 3 and 12 month follow-up was not statistically significant.

At the 3 month follow-up, 66 of the patients were at work and 104 were out of work, and 75 and 88 respectively at the 12 month follow-up (see Table 7). 76 patients judged their working capacity as minimal, achieving an SFS score of less than 100. 62 of these patients were not at work at 3 month follow up and 59 were not at 12 month follow-up. The 14 patients who were false positively diagnosed as not returning to work had an SFS score between 31 and 96. None of the patients with an SFS score of ≤ 30 (n=10) returned to work at 3 month follow-up, resulting in a positive predictive value of 100 and a negative predictive value of 41.2 and one patient at 12 month follow-up, resulting in positive predictive value of 90.0 and negative predictive value of 45.3.

Table 7: Sensitivity, specificity, positive and negative predictive value of an SFS score < 100 for non return-to-work at 3 and 12 month follow-up

	N-RTW (n)	RTW (n)	Total (n)	Sens (%)	Spec (%)	+PV (%)	-PV (%)
3 month follow-up							
SFS score < 100	62	14	76	59.6	78.8	81.6	55.3
SFS score ≥ 100	42	52	104				
Total (n)	104	66	170				
12 month follow-up							
SFS score < 100	59	15	74	67.1	80.0	79.7	67.4
SFS score ≥ 100	29	60	89				
Total (n)	88	75	163				

N-RTW	Not returned to work
RTW	Returned to work
Spec:	Specificity
Sens.	Sensitivity
+PV:	Positive predictive value
-PV:	Negative predictive value

4.2 Functional Capacity Evaluation

4.2.1 Influences on Functional Capacity Evaluation

The aim of this study was to investigate the influences of ''nonorganic-somatic-components'', together with physical and other psychosocial factors, on the results of a FCE in patients with chronic NSLBP undergoing physical fitness for work evaluation.

Descriptives: Mean lifting capacity (16.8, SD10.7) was comparable to earlier published samples of patients with chronic LBP from Switzerland [84] and from Canada [93]. Time for forward bend standing (mean 188.7, SD101.5) was comparable to a patient sample with CLBP achieving 141 sec (SD 101.5) [141]. Grip strength of the dominant hand (mean 33.3, SD15.3) was within the norm [86] and mean walking distance in six minutes (462.3, SD144.5) was below a reference value of 499 meters (95%CI 480–519m) established from healthy persons over 60 years of age [88]. There was a high level of FAB's for work activities (mean 32.7, SD9.0) and low PFA (mean 96.3, SD50.9).

Multivariate regression analysis: The multivariate models (see Table 8) explained between 42% and 58% of the variance in FCE performance.

Table 8: Final robust regression models (n=126)

FCE tests	Adj. R2	Final model	Unstd. Coeff.	95% CI	Sig.	Std. Coeff.
Lifting from floor to waist (kg)	0.54	Perceived functional ability	0.11	(0.08 , 0.14)	<0.001	0.57
		Gender (male)	4.73	(1.90 , 7.55)	0.001	0.20
		'Nonorganic-somatic-components'	-0.95	(-1.66 , -0.24)	0.009	-0.20
Forward bend standing (sec)	0.42	'Nonorganic-somatic-components'	-20.49	(-28.13 , -12.85)	<0.001	-0.44
		Days off work	-0.03	(-0.04 , -0.02)	<0.001	-0.30
		Perceived functional ability	0.31	(-0.02 , 0.63)	*0.065	0.16
Grip strength dominant hand (kg)	0.58	Gender (male)	15.97	(12.03 , 19.90)	<0.001	0.47
		Perceived functional ability	0.11	(0.07 , 0.15)	<0.001	0.38
		'Nonorganic-somatic-components'	-1.53	(-2.51 , -0.54)	0.003	-0.23
		Age	-0.25	(-0.41 , -0.08)	0.005	-0.17
Six minute walking distance (m)	0.52	'Nonorganic-somatic-components'	-27.13	(-37.16 , -17.11)	<0.001	-0.43
		Salary previous job	0.01	(0.00 , 0.02)	0.002	0.20
		Pain intensity	-11.65	(-21.26 , -2.04)	0.018	-0.19
		Fear avoidance belief work activities	-2.50	(-4.47 , -0.52)	0.014	-0.17
		Age	-1.90	(-3.56 , -0.24)	0.025	-0.15

* Not significant, but a confounder, Unstd. Coeff. = Unstandardised Regression Coefficient is the slope of the regression line. It is the increase in y (the dependent variable) corresponding to a unit increase in x the (independent) variable, 95%CI = 95% Confidence Interval for the unstandardised coefficient. Std. Coeff. = Standardised Coefficients are an attempt to make the regression coefficients more comparable. These were used to list the variables in order of importance, but not for interpretation.

The 'nonorganic-somatic-signs' were significant independent predictors for all FCE results. PFA was the most important predictor for lifting performance whereby a 10 point increase in SFS score was associated with a 1.1kg weight increase, followed by the 'nonorganic-somatic-signs' accounting for a 1kg decrease, and females lifting 4.7kg less than males. The 'nonorganic-somatic-signs' were the most influential predictors for forward bend standing. A one point increase was associated with a 20.5sec decrease in performance. Days off work and PFA were also retained as significant predictors in the final model. Gender was most influential on grip strength with women having 16kg less mean grip strength than men. A 10 point increase in SFS score was associated with a 1.1kg grip strength increase, and a one point increase in 'nonorganic-somatic-signs' with 1.5kg decrease in grip strength. Younger patients had more grip strength than older patients. The 'nonorganic-somatic-signs' were also the most influential predictor for walking distance whereby a one point increase was associated with a 27m decrease in distance walked. Three further psychosocial factors of similar importance were associated with walking distance. A 10 year increase in age was associated with 19m decrease in walking distance.

4.2.2   Comparison of two methods for interpretation of lifting performance

The aim of this study was to assess whether 'submaximal effort' and 'nonorganic-somatic-components' contribute independently to lifting performance during FCE and to determine their concurrent validity.

Mean lifting capacity from 'floor to waist' was 16.8kg (SD 10.6), from 'waist to crown' 13.2kg (SD 7.0), and 'horizontal' 20.0kg (SD 12.0). Among the patients classified as giving submaximal effort, the change of heart rate during lifting from 'floor to waist', 'waist to crown', and 'horizontal' was 21.2 (SD12.8), 13.5 (SD8.3), resp. 8.5 (SD7.3). This was significantly lower (p<0.001) than the change of heart rate among the patients classified as giving maximal effort 38.2 (SD17.6), 25.9 (SD12.1), 24.9 (SD11.7).

Apart from age during lifting from 'floor to waist', all variables showed significant independent contribution to lifting performance in the final multivariate models explaining between 48% and 64% of the variance. No significant interactions between 'nonorganic-somatic-components' and 'submaximal effort' was found (p= 0.11, 0.16, resp. 0.3).

'Submaximal effort' had the greatest influence on lifting performance shown by a high change in 'nonorganic-somatic-components' coefficients ranging from 42–58% if added to the model. The change in 'submaximal effort' coefficients ranging from 14–17% was considerably smaller if 'nonorganic-somatic-components' was added to the model. The higher independent contribution of 'submaximal effort' is also shown in the final models by the highest values of the standardised beta coefficients ranging between -0.49 and –0.61,

Patients classified as giving 'submaximal effort' lifted on average 10.4kg less from 'floor to waist', 8.2kg less from 'waist to crown' and 14.9kg less 'horizontal' than patients diagnosed as giving maximal effort. The influence of 'nonorganic-somatic-components' on lifting performance was substantially smaller accounting for 5.9kg less weight lifted from 'floor to waist', 3.2kg less from 'waist to crown' and 5.3kg less 'horizontal'.

The first assessor classified 33% of the patients as presenting with 'nonorganic-somatic-components'. 53%-63% of these patients were classified by the FCE therapist as making 'submaximal effort' during the three manual handling tests. Specificity for classification of maximal effort ranged from 84%-85%.

4.3   Effectiveness of exercise in reducing work disability

The aim of this study was to determine whether exercise is more effective than usual care to reduce work disability in patients with non-acute NSLBP, and if so, to explore which type of exercise is most effective.

We retrieved 838 articles in the literature search. Of these, we evaluated 87 articles in detail, of which 64 did not meet the inclusion criteria. Consequently, we included 23 studies in this review. 20 studies were included in the meta-analysis allowing 17 comparisons of exercise interventions with usual care [121,160-171] and 11 comparisons of two different exercise interventions [18,161,162,166,170,172-177]. Three studies [178-180] were excluded from meta-analysis as 'days of sick leave' were presented as median and interquartile range, thereby preventing pooling. 14 (61%) of the studies were found to be of high quality and 9 (39%) of low quality. The three studies excluded from meta-analysis were all of low quality.

Trials comparing two different exercise interventions with usual care were treated as two trials with the sample size of the usual care group equally divided between the two exercise intervention groups: inpatient rehabilitation and outpatient treatment [161]; low and high intensity back school [162]; light and extensive multidisciplinary programme [166]; conventional physiotherapy and medical exercise therapy [170]. One study [168] presented results of two patient groups defined by the previous intervention (UC, usual care; WI, workplace intervention) receiving the same exercise intervention.

Data on work disability varied between the different studies and included self-assessed work ability, days of sick leave, days at work, physician's judgement of work capability, and days of sickness compensation or numbers of workers returning to full duty work. These were obtained from insurance databases whereby national legal requirements may have influenced the recordings. The data used for pooling were the number of people who returned and did not returned to work at the time of the follow-up, or the total number of sick days within the follow-up period.

4.3.1   Exercise characteristics

35 different exercise interventions were used. Exercise design, dose and setting were reported unclear in 6% of the investigated exercise interventions. 26 (74%) of the exercise interventions were individually designed; 32 (91%) were primarily performed as supervised exercise; 28 (80%) interventions used mixed exercise types, 2 stabilisation, 3 strengthening, 1 mobilisation, and 1 stretching exercise; 27 (77%) were conducted in an outpatient setting; 10 (29%) were work-related; and 14 (40%) of the exercise interventions were administered within a cognitive behavioural approach.

Unfortunately, none of the studies using home exercise reported adherence rate or information sufficient to estimate home exercise dose. Therefore, the calculation of the exercise dose is based on the number of the supervised treatment sessions and their duration only. Such calculated exercise dose varied widely between the different exercise interventions, ranging between 1.5 and 210 hrs. The median exercise dose was 17 hours. We classified exercise interventions with 17 or more hours of contact time into high dose exercise (n = 18), and those with less than 17 hours of contact time into low-dose exercises (n = 17). A cut-off point of 14 and 20 hours resulted in a less than 10% change in exercise dose classification.

4.3.2 Comparison of exercise interventions and usual care

13 studies allowing 17 comparisons between an exercise intervention and usual care with a total of 3181 patients were available for pooling.

Short-term follow-up: Short-term results were available for pooling from 5 high quality studies (6 comparisons, 1003 patients) [121,160,162,168,169], showing no significant effect of exercise in reducing work disability (OR = 0.80, 95% CI 0.51 – 1.25). The addition of one low quality study [171] did not substantially change the overall effect estimate (OR = 0.68, 95% CI 0.42 – 1.10).

Intermediate-term follow-up: 4 high quality studies (5 comparisons, 971 patients) [121,160,162,168] provided results for pooling at the intermediate-term follow-up showing no significant effect of exercise in reducing work disability (OR = 0.78, 95% CI 0.45 – 1.34).

Long-term follow-up: 8 high quality studies (10 comparisons, 1992 patients) [121,160,164-168,170] presented long-term follow-up results showing a statistically significant overall effect in favour of exercise on work disability (OR = 0.66, 95% CI 0.48 – 0.92). There was no evidence for publication bias at long-term follow-up shown by symmetric funnel plots.

4.3.3 Influence of exercise characteristics

The 8 high quality studies with long-term follow-up [121,160,164-168,170] providing data on 1149 patients receiving an exercise intervention and 843 patients receiving usual care were included for this analysis. All comparisons were between different outpatient rehabilitation programs, and all used individually designed exercises. In one comparison stretching exercises were instructed, another used stabilisation exercise, and in the remaining 8 comparisons mixed exercises were used. A second overall analysis, which did not include the patient sample from the trial of Steenstra et al. that had already received a workplace intervention (WI group), showed reduced statistical heterogeneity (I^2 = 60.4%, p = 0.007) and increased the effect estimate (OR = 0.59, 95% CI 0.45 – 0.78). The effect of delivery type, exercise dose, work context and behavioural treatment approach was analysed with and without the WI group showing different results, although within the statistical error margin. Pooled effects for the four exercise characteristics hypothesised to influence work disability (delivery type, exercise dose, work context, behavioural treatment approach) became higher for supervised exercise, and a behavioural treatment approach. However, none of the

variables were statistically significant in meta-regression analysis, although there was a trend observed favouring home exercises (p=0.11).

4.3.4   Comparison of different exercise interventions

Of the 13 studies comparing 15 different exercise interventions, 6 were of low quality [161,172,174,177,178,180] and 7 were of high quality. 6 high quality studies presented long-term follow-up data and were used for pooling [18,166,170,173,175,176]. We defined the exercise intervention with more contact hours as the standard intervention. All standard interventions used individually designed supervised exercises, 5 of them with mixed exercises, and 5 were conducted in an outpatient setting. There was significant statistical heterogeneity in these trials (I-squared = 65.5%, p = 0.013). The overall effect of exercise interventions with more contact hours was not significant (OR = 1.07 95% CI 0.67-1.72). Within the trials comparing exercise interventions with more contact hours to exercise interventions with fewer contact hours, the 3 trials applying exercise within a behavioural treatment approach showed some benefit (OR = 0.75 95% CI 0.47-1.20) compared to the trials without this characteristic (OR = 1.74 95% CI 0.71-4.30). One trial applying work-related exercise in an inpatient setting [19] showed a significant effect on work disability (OR = 0.53 95% CI 0.30-0.93) compared to exercise not specifically designed to restore work-related physical capacity (OR = 1.25 95% CI 0.80-1.97). None of these characteristics showed statistical significance in meta-regression analysis. The funnel plots that were conducted did not reveal evidence of funnel plot asymmetry.

5.   Discussion

The overall aim of this thesis was to examine relevant work-related assessments and the effect of exercise to reduce work disability of patients with non-acute NSLBP. In the following, a discussion with respect to methodology and main results will be made.

5.1   Methodological considerations

5.1.1   Prospective cohort study (paper I)

This study investigating the validity of the SFS in a European rehabilitation was embedded within an RCT investigating the effectiveness of a function-centred treatment compared with a pain-centred treatment with 3 and 12 month follow-up for working status [18,19]. In the design of this RCT particular emphasis was given to obtain a representative patient sample referred for inpatient rehabilitation by not excluding patients with mother tongues other than those nationally spoken. All questionnaires in this study therefore had to be either available in all required languages or be independent of language. The FABQ was therefore translated into German, Italian, Serbo-Croatian, Albanian, Turkish, French, Spanish and Portuguese. This may have introduced measurement bias as not all of these translated versions of the FABQ were cross-culturally adapted. In view of the expected higher correlation of the picture based SFS score with the FABQ work score than with FABQ activity score we felt the patients had understood the questions and we, consequently, interpreted the findings as valid.

Another essential issue of this study is the measurement of the number of days at work. As no national database exists to obtain such information we had to resort to questionnaires sent to both the primary physicians and the employers, who were blinded to the treatment group. This approach may have introduced more random measurement error but also allowed the assessment of the initial days of each period of sick leave. Databases used in other studies did not cover the first 7 [166] or 16 days [161,181] of each work absence.

5.1.2   Cross-sectional study (papers II and III)

We thoroughly considered whether the questions posed in papers II and III could be investigated in one paper. However, substantial methodological differences among the FCE tests inhibited such a combined analysis. A full Isernhagen FCE consists of 28 tests [74]. These can be categorised into manual handling capacity tests, work postures, hand capacity and ambulation. The aim of paper II was to investigate the influences of 'nonorganic-somatic-components', together with physical and other psychosocial factors, on the results of a FCE. We used one representative test of each FCE test category ('floor to waist lifting', 'forward bend standing', 'grip strength' and 'six minute walking distance'). Paper III aimed to assess the contribution of 'nonorganic-somatic-components' and 'submaximal effort' specifically for lifting performance and to determine the concurrent validity of the 'nonorganic-somatic-components' with 'submaximal effort' during FCE. Maximal effort in lifting tests is determined by observational criteria. For work postures and ambulation, other observational criteria are used to diagnose sincerity of effort [114]. To determine sincerity of effort during hand capacity

tests an interpretation of the grip curve in five different handle positions is performed [182]. We therefore had to resort to two separate analyses.

A weakness of the cross-sectional study is that it was conducted in a single country with a highly established social insurance system reducing its external validity. One might also criticise that 52% of the patient population's previous jobs had been heavy or very heavy and this was not accounted for in the analysis as a possible physical risk factor for long term disability. Instead, we used salary in the previous job as an explanatory variable. Both are mutually associated but salary in the previous job is also associated with other known physical risk factors such as repetitive assembly work or monotonous static work, and is an important compensation issue in Switzerland as it serves for final disability determination. However, we do admit that the inclusion of further physical work factors might explain a higher amount of variation in FCE performance.

It might be further argued that the assessment of the dependent variable 'grip strength' by the first assessor is a weakness of paper II. The intention was to blind the FCE therapist towards the frequently used grip curve assessment as a proxy for 'nonorganic-somatic-components' within grip curve assessment. To prevent prejudice in the first assessor, he assessed the 'nonorganic-somatic-signs' before assessing grip strength. We therefore feel that the assessment of one independent variable by the first assessor is justified and does not diminish the validity of the study's findings.

A weakness of paper III is that the FCE assessors were potentially influenced in their determination of effort by the observed low performance, thus explaining the high contribution of 'submaximal effort' to the FCE result. The possibility that the observational criteria were not consequently applied cannot be ruled out. However, we assessed as an external objective measure of effort the change of heart rate during lifting from 'floor to waist', 'waist to crown', and 'horizontal'. In all three lifting tests, we found significantly lower changes in heart rate among the patients classified as giving submaximal effort than among the patients giving maximal effort. Heart rate was an observational criterion. Therefore, these findings can be interpreted as an indication that the FCE assessors did adhere to the observational criteria. The addition of further external measures such as a control group as used in other studies investigating the validity of effort level determination during FCE [106-108] was beyond the scope of this cross-sectional study.

Major strengths of the cross-sectional study are the multi-centre setting involving a large number of assessors, the inclusion of patients of several nationalities, which raises the external validity of the study's findings, and the blinding of the two assessors against each other's results.

5.1.3 Systematic review (paper IV)

A weakness of the meta-analysis is the high proportion of total unexplained variance that could be attributed to study heterogeneity. We thoroughly considered this weakness but concluded that patients, social support and outcomes showed satisfactory homogeneity. All but one of the studies were performed in Europe, in countries with comparable social system.

All patients were diagnosed with non-acute NSLBP, were of working age and available for the job market. Despite the wide variety of used work disability outcomes, this meta-analysis is based on the pooled results of just two different outcome measures. We performed a stratified analysis in three studies providing both outcome measures and found no relevant differences in odds ratios, both in favour of exercise. Furthermore, using mean values and standard deviations for further statistical analysis in data with a skewed distribution is usually regarded as inappropriate. Data regarding sick days have a skewed distribution, but this was similar in both groups in treatment comparisons which reduces the risk of systematic bias [183]. To address the problem of statistical heterogeneity, we performed a random effects meta-analysis. There remains the possible error of substantial variation in standard deviations across studies leading to an over- or underestimation of the odds ratios.

Our study has several strengths. The search strategy was based on the recommendations of the Cochrane Back Review Group. We planned the analysis a priori based on the findings of previous meta-analyses and assessed study quality based on key components of methodological quality (concealed allocation, blinded assessor, intention to treat analysis) as recommended by Juni et al. [127]. Studies affected by biases have previously been shown to exaggerate treatment effects [127]. We, therefore, excluded low quality studies from meta-regression analysis to avoid a possible overestimation of the effect of different exercise characteristics.

5.2   Assessment of patients with low back pain

5.2.1   Assessing patients' perceptions with a picture-based questionnaire

This is, to our knowledge, the first study assessing the validity of a picture-based questionnaire assessing perceived ability for work tasks in a European rehabilitation setting. The majority of the patients were accustomed to heavy work, came from 10 different nations, were poorly educated, and had insufficient knowledge of the Swiss national languages [18]. We consider this pan-European patient sample as a major strength of this study. We found a high internal consistency, acceptable unidimensionality and concurrent validity, as well as a good diagnostic accuracy for future work status. Furthermore, correlations of the SFS score with the FABQ work score were higher than with FABQ activity score and the SFS was able to capture the expected changes in the two different treatment approaches. These findings add further validity to this questionnaire specifically assessing PFA to perform work tasks. Based on these findings we can recommend the SFS for clinical use in a European rehabilitation setting treating patients with different mother tongues and variate literacy level.

5.2.2   Association of perceived functional ability for work tasks with RTW and FCE

The high follow-up rate of work status at 3 and 12 months allowed a thorough evaluation of the diagnostic accuracy of the SFS for future work status. We found discriminating power of an SFS score of less than 100 indicating a minimal perceived working capacity. This cut-off

score of <100 showed a high positive predictive value of 81.6% for non RTW and specificity of 78.8% at three month follow-up, resp. 79.7% and 80% at 12 month follow-up. In view of the high specificity and the high prevalence of patients with non-acute NSLBP not returning to work, and both factors being determinants of the positive predictive value, we can recommend using an SFS score <100 for screening purposes in clinical practice identifying patients with high probability of not returning to work. We also found in paper II significant associations of PFA with lifting performance, forward bend standing and grip strength. These findings, together with the high association of the SFS with the FABQ for work activities, can be interpreted as an indication for the construct validity of the SFS as a questionnaire assessing PFA for work tasks.

The knowledge of low PFA for work tasks allows choice of appropriate rehabilitation interventions. Previous research has shown that rehabilitation programs focusing on pain relief diminish PFA for work tasks while a function-centred treatment approach improved it [18,19]. The SFS was able to capture these changes in PFA in the two treatment groups shown by a positive SRM for a higher perceived functional ability in the function-centred treatment group and a negative SRM for the pain-centred treatment group. Such an assessment approach is in line with the recommended screening for psychosocial yellow flags [9,51-53] which need to be addressed so that the risk of developing long-term disability and work loss can be reduced [54].

5.2.3   Influences of physical and non-physical factors on Functional Capacity Evaluation

This study confirms previous findings showing that FCE test results are influenced by physical as well as by non-physical factors [93]. PFA for work tasks was the most important predictor for lifting performance and also contributed significantly to grip strength and forward bend standing. A new finding of this study is the consistent independent prediction of the 'nonorganic-somatic-signs' for FCE performance in all tests. Their influence was most significant on forward bend standing and walking distance, and less significant on grip strength and lifting performance. The physical factors of age and/or gender were strongly associated with grip strength and lifting, less with walking distance and not at all with forward bend standing. As hypothesised, the inclusion of 'nonorganic-somatic-components' has led to substantially higher amounts of explainable variation in lifting performance (54%) than were found in a previous model [93] consisting of perceived disability due to pain, age, and gender explaining 20%, 6%, respectively 5% of lifting performance.

A striking finding of this study is the influence of 'nonorganic-somatic-components' on tests without any specific load on the spine, as observed in grip strength. Without this additional information on the presence of 'nonorganic-somatic-components' physical disability relating to grip strength would have been interpreted. The same applies to the lower mean walking distance observed compared to a considerably older population of healthy adults [88], suggesting disability in walking arising from LBP. However, 'nonorganic-somatic-components' had the most significant and clinical meaningful influence, with a one point

increase in 'nonorganic-somatic-signs' accounting for a 27m decrease in walking distance. We therefore hypothesise that FCE influenced by 'nonorganic-somatic-components' should not be interpreted solely as a reflection of the remaining physical function of patients with back pain but indeed as behavioural tests influenced by non-physical factors.

This statement must not be misunderstood as a call to use the 'nonorganic-somatic-signs' as a screening tool for malingering. In accordance with previous authors, we interpret such behaviour as a form of communication between patient and examiner influenced by expectations [105] and possibly arising from pain, fear of injury or neuromuscular inhibition [105,120]. Some of these patients may require a more careful examination and management of the psychosocial and behavioural aspects of their illness [105].

The question arises as to what cut-off point in the number of 'nonorganic-somatic-signs' should be used to declare an FCE as not being representative of physical abilities. The recommendation of Waddell et al. to interpret three or more positive categories as indicators of 'nonorganic-somatic-components' within the physical assessment requires at least three positive 'nonorganic-somatic-signs' out of the total eight signs (see Table 4). This leads to clinically meaningful changes in three out of the four test results: i.e. 61.6sec decrease in forward bend standing, 4.5kg decrease in grip strength, and 81m decrease in walking distance. We therefore feel encouraged to use the proposed cut-off point as defined by Waddell et al. [94].

5.2.4 Interpretation of lifting performance during FCE

We have been able to demonstrate that 'nonorganic-somatic-components' and 'observational criteria' for physical effort contribute independently to lifting performance without significant interactions in patients with chronic NSLBP undergoing FCE to determine fitness-for-work. The contribution of 'submaximal effort' to lifting performance was higher than that of 'nonorganic somatic components'. This was shown by markedly higher changes in coefficients if 'submaximal effort' was added to the model compared to the lower changes if 'nonorganic-somatic-components' was added, as well as by higher standardised beta values of -0.49 to –0.61 in the final models compared with – 0.21 to –0.26. Concurrent validity of the 'nonorganic-somatic-components' with 'submaximal effort' was low as shown by the low sensitivity. Thus, the two measures appear to reflect different aspects of physical performance during FCE of lifting and should therefore not be interchangeably used.

Despite the results suggesting that the two measures appear to reflect different aspects of physical performance during FCE, the overlap between them should not be dismissed. 53%-63% of the patients classified as presenting 'nonorganic-somatic-components' were also diagnosed as giving submaximal effort. This overlap might be an explanation for the association found in previous study of 'nonorganic-somatic-components' with decreased functional performance [99,150]. We can only hypothesise on possible explanations for this overlap. 'Nonorganic-somatic-components' should be viewed as a behavioural response to examination which could be driven by fear of further injury [105]. It is theoretically plausible that fear of injury might is also be an underlying cause for submaximal performance during FCE.

Since both measurements were taken on the same day during fitness-for-work evaluation, another possible explanation is that personal expectations arising from this evaluation may have led to overt illness behaviour. Overreaction, such as disproportionate verbalisation, facial expressions, muscle tension and tremors is one criterion for 'nonorganic-somatic-components' [94]. Overreaction can also be observed in FCE during handling of leight weights. Such behaviour hinders the occurrence of body reactions indicative of maximal physical efforts and will consequently lead to a 'submaximal effort' classification.

A major point that can be taken from this study is that 'nonorganic-somatic-signs' testing should not be used to assess 'physical effort'. This is illustrated by the low concurrent validity and the markedly lower contribution of 'nonorganic-somatic-components' to FCE performance. Without the direct observation of the effort level during the lifting tests, the estimated influence of submaximal effort based on 'nonorganic-somatic-signs' testing would have been substantially smaller and may have lead to incorrect classification of disability and, consequently, inadequate decisions regarding vocational rehabilitation and physical fitness for work. Based on the study's results and on the different constructs that were to be measured using these two methods, we do not recommend the use of 'nonorganic-somatic-signs' testing for effort evaluation. The use of the previously described observational criteria for this purpose seems appropriate.

5.3   The Effect of exercise on work disability

This meta-analysis provides continuous support for the use of exercise interventions to achieve long-term benefits on work disability in patients with non-acute NSLBP. The odds ratio of 0.66 suggests that the odds for 'improvement' in work disability are in the long-term 34% lower if only usual care (rather than exercise) is given. No significant effect was observed in short- and intermediate-term follow-ups. Meta-regression showed no significant differences between different exercise types. Interestingly, home exercises seem to be at least as effective as supervised programs. As the meta-regression is only explorative, no conclusions can be drawn regarding exercise types.

5.3.1   Effectiveness at different times of follow-up

Exercise interventions did not show a significant effect on work disability at short- and intermediate-term follow-up. However, these findings are not conclusive. The mean odds ratios for short- and intermediate-term results were both below unity but with wide confidence intervals. Therefore, a significant effect might remain undetected based on ineffectiveness, heterogeneity or limited power of the pooled studies. Possible explanations for a lack of effect at short- and intermediate-term follow-up are the time required to improve physical capacity, to modify pain behaviour, or to search for work. Furthermore, the process of care has a substantial effect on work disability as shown in a recent study comparing a graded activity program with usual care [168]. The interaction between a prior workplace intervention and graded activity, together with a delay in the start of the graded activity intervention, explained most of the delay in RTW [168]. This study introduced relevant clinical heterogeneity

in this meta-analysis. All of the other trials investigated the primary treatments for this occurrence of back pain while half of the patients in the trial from Steenstra et al. [168] had already received a workplace intervention (WI group), which has been shown to be effective on RTW [184]. Herbert & Bo [185] propose that study quality can also be assessed on how interventions are administered. There were clearly problems in the implementation of the graded activity program in the trial by Steenstra et al. [168] leading to a potentially false conclusion if the whole patient sample had been included in this meta-analysis. In view of these considerations we feel it legitimate to interpret the findings without the results of the WI group.

The author's recommendation of paying special attention to the structure and process of care in implementing graded activity [168] does have clinical relevance when conducting medical interventions aiming for early RTW. An open and fast access to such interventions prevents unnecessary waiting time before an RTW can be attempted. This might also be a possible explanation for why individually designed home exercises seem to be more effective than supervised exercise interventions in reducing work disability. Home exercise may facilitate RTW as the patients are able to continue their daily routine without spending extra time on medical intervention.

5.3.2   Exercise dose and effectiveness

Interestingly, this meta-analysis did not show a greater effect of higher dose exercise interventions (≥ 17 contact hours) compared to lower dose exercise interventions on work disability (< 17 contact hours). This finding is contrary to exercise physiology postulating a dose and effect relation [186], as well as to previous findings that only intensive (> 100 hours of therapy) multidisciplinary biopsychosocial rehabilitation with functional restoration improves function in patients with chronic LBP, whereby inconclusive results were found on vocational outcomes [116]. The only study included in the systematic review by Guzman et al. [116] supporting the use of functional restoration to reduce work disability [179] is not included in the performed meta-regression because only high quality studies were used. Moreover, in the meantime new studies with low contact hours administering home exercise have been published showing a positive effect on work disability. This might be an explanation for the different findings. However, a cautionary comment must be made on the missing effect of exercise dose found in this review. As in other systematic reviews, incomplete reporting in the primary studies presents important limitations and prevented the calculation of the exercise dose in home exercise programs. It must be assumed that the actual exercise dose in home exercise interventions was higher than the one calculated.

5.3.3   Behavioural treatment components

We have not been able to confirm the positive effects of exercises performed within a behavioural treatment approach on work disability postulated in previous reviews [119,123,128]. In the comparison of exercises with usual care we found stronger treatment effects for such exercises. However, this was not statistically significant in the meta-regression. The missing confirmation might be due to the differing study inclusion criteria and the analysis performed.

All previous reviews based their conclusion on a qualitative assessment, at least partly based on the evidence found by Lindström et al. [121]. Hayden et al. [123] also included the results of Staal et al. [167], while Schonstein et al. 2003 [128] included the results of three more studies which were excluded from this analysis due to the risk of bias or missing inclusion criteria. In addition to the studies of Staal et al. and Lindström et al. [121,167] we included the findings of six more studies [160,162,164,166,168,169] that contained nine treatment comparisons with a total of 1316 patients providing sufficient power for the meta-regression. However, it must be emphasised that the presented meta-regression analysis is only explorative and does not allow any conclusions.

The comparison of different exercise interventions also did not reveal a significant effect of a behavioural treatment approach. There might be a superior effect if exercises are performed within a behavioural treatment approach and are specifically designed to restore work-related capacity as shown by the study of Kool et al. [19]. This is also in line with Schonstein et al. [117], who hypothesise a positive effect of such a combination.

6.   Conclusions and clinical implications

▪ Perceived functional ability for work-related tasks can validly be assessed with the SFS in a European rehabilitation setting, and is predictive for future work status.

  - The SFS can be used in daily practice in patients with different mother tongues and literacy level.

  - The SFS may be used to identify patients at risk of not returning to work and consequently guide rehabilitation interventions.

▪ For a comprehensive FCE of patients with non-acute NSLBP referred for fitness for work evaluation, an assessment of the patient's perceived functional ability for work tasks as well as the presence of 'nonorganic-somatic-components' should be considered.

  - Information about the patient's perceived functional ability for work tasks and the presence of 'nonorganic-somatic-components' would allow an interpretation of the validity of FCE results as a measure of physical fitness for work within the four FCE categories.

  - 'Nonorganic-somatic-signs' testing should not be used to assess 'physical effort'. The use of the described observational criteria for this purpose seems appropriate.

▪ Exercise interventions have a significant effect on work disability in patients with non-acute NSLBP in the long term. No conclusions can be drawn regarding exercise types.

  - The structure and process of care when implementing exercise interventions aiming for early return-to-work should be taken into account. An open and fast access to exercise interventions might prevent unnecessary waiting time, thereby facilitating an early return to work.

7.   Further research

Spinal Function Sort: Principal component analysis has shown reasonable evidence for unidimensionality of the SFS. However, 4 items asking for manual handling capacity of 50kg had higher loading for another dimension. These 4 questions showed also low corrected total item correlation in the assessment of internal consistency. We, therefore, recommend further research using RASCH analysis to investigate unidimensionality and consequently reduce questionnaire items.

Functional Capacity Evaluation: Despite promising results for the validity of the observational criteria applied during FCE [107] and the results found in this study, further research on 'physical effort evaluation' is needed. Serious medico-legal consequences might result for a patient if he is labelled as exerting 'submaximal effort' during a physical fitness-for-work evaluation.

Exercise interventions: Previous papers [117] as well as isolated findings within this systematic review [19] suggest a superior effect of exercises if these are performed within a behavioural treatment approach and are specifically designed to restore work-related capacity. We recommend further evaluation of the combined effects of individually designed home exercises applied within a behavioural treatment approach aiming to specifically restore work-related physical capacity and that special attention given to a fast and open access to such treatment.

8.   References

1.   Kendall N, Linton S, Main C. Guide to assessing psychosocial yellow flags in acute low back pain: risk factors for long-term disability and work loss. ed. Wellington NZ: Accident rehabilitation and compensation insurance corporation of New Zealand and the National Health Committee, 1997.

2.   Wittink H. Functional Capacity Testing in Patients With Chronic Pain. The Clinical Journal of Pain 2005;21:197-9.

3.   Soer R, van der Schans CP, Groothoff JW, et al. Towards consensus in operational definitions in functional capacity evaluation: a Delphi Survey. J Occup Rehabil 2008;18:389-400.

4.   Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. Bull World Health Organ 2003;81:646-56.

5.   Olafson S. Social and Personal costs of Arthritis and Rheumatic Diseases - An explanatory survey undertaken for the Nordic Rheuma Council (NRR): Social Reseach Institute University of Iceland, 2008:2-77.

6.   Wieser S, Horisberger B, Schmidhauser S, et al. Cost of low back pain in Switzerland in 2005. Eur J Health Econ 2011;12:455-67.

7.   OECD. Sickness, Disability and Work: Breaking the Barriers. A Synthesis of findings across OECD countries.ed: OECD Publishing, 2010.

8.   Fordyce W. Back Pain in the Workplace: Management of Disability in Nonspecific Conditions A Report of the Task Force on Pain in the Workplace of the International Association for the Study of Pained. Seattle, WA: IASP Press, 1995.

9.   Airaksinen O, Brox JI, Cedraschi C, et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 2006;15 Suppl 2:S192-300.

10.   Balague F, Mannion AF, Pellise F, et al. Non-specific low back pain. Lancet 2011.

11.   Abenhaim L, Rossignol M, Valat JP, et al. The role of activity in the therapeutic management of back pain. Report of the International Paris Task Force on Back Pain. Spine 2000;25:1S-33S.

12.   Koes BW, van Tulder M, Lin CW, et al. An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. Eur Spine J 2010;19:2075-94.

13.   Milhous RL, Haugh LD, Frymoyer JW, et al. Determinants of vocational disability in patients with low back pain. Arch Phys Med Rehabil 1989;70:589-93.

14.   McIntosh, Frank J, Hogg-Johnson S, et al. Prognostic Factors for Time Receiving Workers' Compensation Benefits in a Cohort of Patients With Low Back Pain. Spine 1999;25:147–57.

15.   Cocchiarella L, Turk MA, Andersson G. Improving the evaluation of permanent impairment. Journal of the American Medical Association 2000;283:532-3.

16.   Hunt DG, Zuberbier OA, Kozlowski AJ, et al. Are components of a comprehensive medical assessment predictive of work disability after an episode of occupational low back trouble? Spine 2002;27:2715-9.

17.   Bolderson H, Mabbett D, Hvinden B. Definitions of Disability in Europe: A Comparative Analysised. Brussels, Belgium: European Commission Directorate General for Employment and Social Affairs, 2002.

18.   Kool JP, Oesch PR, Bachmann S, et al. Increasing days at work using function-centered rehabilitation in nonacute nonspecific low back pain: a randomized controlled trial. Arch Phys Med Rehabil 2005;86:857-64.

19.   Kool J, Bachmann S, Oesch P, et al. Function-centered rehabilitation increases work days in patients with nonacute nonspecific low back pain: 1-year results from a randomized controlled trial. Archives of physical medicine and rehabilitation 2007;88:1089- 94.

20. Oesch P, Kool J, Wunderlin B, et al. Rehabilitation von Patienten mit chronischen Rückenbeschwerden: Assessment Ergebnisse und prädiktive Faktoren. Phys Rehab Kur Med 1997;7:224-30.

21. Kool JP, Oesch PR, de Bie RA. Predictive tests for non-return to work in patients with chronic low back pain. Eur Spine J 2002;11:258-66.

22. Bachmann S, Oesch P, Kool J, et al. Treatment of patients with chronic low back pain in a functional restoration program: work related function parameters, pain parameters and the working status after 12 months (German). Phys Rehab Kur Med 2003;13:263-70.

23. Bachmann S, Wieser S, Oesch P, et al. Three-year cost analysis of function-centred versus pain-centred inpatient rehabilitation in patients with chronic non-specific low back pain. J Rehabil Med 2009;41:919-23.

24. Oesch PR, Kool JP, Bachmann S, et al. The influence of a Functional Capacity Evaluation on fitness for work certificates in patients with non-specific chronic low back pain. Work 2006;26:259-71.

25. Scheermesser M, Kool J, Oesch P, et al. Rehabilitation of low back pain in patients with a Southeast European cultural back-ground – Patient's perspectives on rehabilitation using a constructivist approach. Swiss Med Wkly 2009;139:5.

26. Guo HR, Chang YC, Yeh WY, et al. Prevalence of musculoskeletal disorder among workers in Taiwan: a nationwide study. J Occup Health 2004;46:26-36.

27. Louw QA, Morris LD, Grimmer-Somers K. The prevalence of low back pain in Africa: a systematic review. BMC Musculoskelet Disord 2007;8:105.

28. El-Sayed AM, Hadley C, Tessema F, et al. Back and Neck Pain and Psychopathology in Rural Sub-Saharan Africa: Evidence From the Gilgel Gibe Growth and Development Study, Ethiopia. Spine (Phila Pa 1976) 2010.

29. Mousavi SJ, Akbari ME, Mehdian H, et al. Low back pain in Iran: a growing need to adapt and implement evidence-based practice in developing countries. Spine (Phila Pa 1976) 2011;36:E638-46.

30. Walker BF. The prevalence of low back pain: a systematic review of the literature from 1966 to 1998. J Spinal Disord 2000;13:205-17.

31. Jeffries LJ, Milanese SF, Grimmer-Somers KA. Epidemiology of adolescent spinal pain: a systematic overview of the research literature. Spine (Phila Pa 1976) 2007;32:2630-7.

32. Loney PL, Stratford PW. The prevalence of low back pain in adults: a methodological review of the literature. Phys Ther 1999;79:384-96.

33. Manek NJ, MacGregor AJ. Epidemiology of back disorders: prevalence, risk factors, and prognosis. Curr Opin Rheumatol 2005;17:134-40.

34. Raspe H, Matthis C, Croft P, et al. Variation in back pain between countries: the example of Britain and Germany. Spine 2004;29:1017-21; discussion 21.

35. Deyo RA, Mirza SK, Martin BI. Back pain prevalence and visit rates: estimates from U.S. national surveys, 2002. Spine (Phila Pa 1976) 2006;31:2724-7.

36. Freburger JK, Holmes GM, Agans RP, et al. The rising prevalence of chronic low back pain. Arch Intern Med 2009;169:251-8.

37. BFS. Gesundheit und Gesundheitsverhalten in der Schweiz 2007. [serial online]. Available from: http://www.bfs.admin.ch/bfs/portal/de/index/themen/14/22/publ.html?publicationID=4092. Accessed 12. October 2011, 2010.

38. Brage S, Laerum E. Low back disorders in Norway - An epidemiological description. Tidsskr Nor Laegeforen 1999;Apr 30:1619-23.

39.   Frymoyer JW, Cats-Baril WL. An overview of the incidences and costs of low back pain. Orthop Clin North Am 1991;22:263-71.

40.   BSV. IV-Statistik 2003ed. Bern: Bundesamt für Sozialversicherungen, 2003.

41.   Breitenmoser B, Foffa F, Guggisberg G, et al. «Eingliederung vor Rente» – realisierbares Ziel oder bloss wohltönender Slogan? Soziale Sicherheit 1999;6:288-92.

42.   Leuenberger M, Huber-Hotz A. Botschaft über die 4. Revision des Bundesgesetzes über die Invalidenversicherung. 2001.

43.   Hagen KB, Thune O. Work incapacity from low back pain in the general population. Spine 1998;23:2091-5.

44.   Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 2008;8:8-20.

45.   van Tulder MW, Koes BW, Bouter LM. A cost-of-illness study of back pain in The Netherlands. Pain 1995;62:233-40.

46.   Spitzer WO. Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Report of the Quebec Task Force on Spinal Disorders. Spine 1987;12:S1-S59.

47.   Chou R, Huffman LH. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. Ann Intern Med 2007;147:492-504.

48.   Koes BW, van Tulder MW, Ostelo R, et al. Clinical guidelines for the management of low back pain in primary care: an international comparison. Spine 2001;26:2504-13.

49.   Staal JB, Hlobil H, van Tulder MW, et al. Occupational health guidelines for the management of low back pain: an international comparison. Occup Environ Med 2003;60:618-26.

50.   Waddell G. The Back Pain Revolutioned. Edinburgh: Churchill Livingstone, 1998.

51.   Waddell G, Burton AK. Occupational health guidelines for the management of low back pain at work: evidence review. Occup Med (Lond) 2001;51:124-35.

52.   ACC. New Zealand Acute Low Back Pain Guide: New Zealand Accident Compensation Cooperation, 2004.

53.   Australian. Evidence-Based Management of Acute Musculoskeletal Pain Guidelines for Cliniciansed. Bowen Hills, Queensland: Australian Academic Press, 2004.

54.   Kendall NA. Psychosocial approaches to the prevention of chronic pain: the low back paradigm. Baillieres Best Pract Res Clin Rheumatol 1999;13:545-54.

55.   Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. Clin J Pain 1998;14:209-15.

56.   Nicholas MK, Linton SJ, Watson PJ, et al. Early identification and management of psychological risk factors ("yellow flags") in patients with low back pain: a reappraisal. Phys Ther 2011;91:737-53.

57.   Rahman A, Reed E, Underwood M, et al. Factors affecting self-efficacy and pain intensity in patients with chronic musculoskeletal pain seen in a specialist rheumatology pain clinic. Rheumatology (Oxford) 2008;47:1803-8.

58.   Bandura A. Self-efficacy. In Ramachaudran V ed. Encyclopedia of human behavior. New York: Academic Press, 1994:71-88.

59.   Hoogendoorn WE, van Poppel MN, Bongers PM, et al. Physical load during work and leisure time as risk factors for back pain. Scand J Work Environ Health 1999;25:387-403.

60.   Matheson L, Matheson M, Grant N. Development of a measure of perceived functional ability. Journal of Occupational Rehabilitation 1993;3:15-30.

61.   Adams J, Bradley S, J C. Literacy levels required to complete the patients reported functional outcomes measures in rheumatology. Ann Rheum Dis 2009;68:771.

62.   Oliveri M. Arbeitsbezogene funktionelle Leistungsfähigkeit (Evaluation der funktionellen Leistungsfähigkeit EFL). In Hildebrandt J, Müller G, Pfingsten M eds. Lendenwirbelsäule. Ursachen, Diagnostik und Therapie von Rückenschmerzen. München: Urban & Fischer, 2004:220-35.

63.   Sufka A, Hauger B, Trenary M, et al. Centralization of low back pain and perceived functional outcome. J Orthop Sports Phys Ther 1998;27:205-12.

64.   Robinson RC, Kishino N, Matheson L, et al. Improvement in postoperative and nonoperative spinal patients on a self-report measure of disability: the Spinal Function Sort (SFS). J Occup Rehabil 2003;13:107-13.

65.   Gibson L, Strong J. The reliability and validity of a measure of perceived functional capacity for work in chronic back pain. Journal of Occupational Rehabilitation 1996;6:159-7.

66.   OECD. Transforming Disability into Ability ed: OECD Publishing, 2003.

67.   OECD. Sickness, Disability and Work: Breaking the Barriers (Vol.1) - Norway, Poland and Switzerlanded: OECD Publishing, 2006.

68.   Meyer-Blaser U. Der Rechtsbegriff der Arbeitsunfähigkeit und seine Bedeutung in der Sozialversicherung, namentlich für den Einkommensvergleich in der Invaliditätsbemessung. In Prof. B. Ehrenzeller PRS, Prof. I. Schwander ed. Schmerz und Arbeitsunfähigkeit. St. Gallen: Schriftenreihe des Institutes für Rechtswissenschaft und Rechtspraxis, 2003:35-57.

69.   DWP. Raising expectations and increasing support: reforming welfare for the futureed: Departement for Work and Pensions, 2008.

70.   Wind H, Gouttebarge V, Kuijer PP, et al. Complementary value of functional capacity evaluation for physicians in assessing the physical work ability of workers with musculoskeletal disorders. Int Arch Occup Environ Health 2009;82:435-43.

71.   Verbeek J, van Dijk F. Assessing the ability to work. Bmj 2008;336:519-20.

72.   Innes E, Straker L. Validity of work-related assessments. Work 1999;13:125-52.

73.   Innes E, Straker L. Reliability of work-related assessments. Work 1999;13:107-24.

74.   Isernhagen SJ. Functional capacity evaluation. Work Injury: Management and Prevention. Gaithersburg: Aspen Publishers Inc, 1988:139–91.

75.   Gouttebarge V, Wind H, Kuijer PP, et al. Reliability and validity of Functional Capacity Evaluation methods: a systematic review with reference to Blankenship system, Ergos work simulator, Ergo-Kit and Isernhagen work system. Int Arch Occup Environ Health 2004;77:527-37.

76.   Innes E. Reliability and Validity of Functional Capacity Evaluations: An Update. International Journal of Disability Managment Research 2006;1:135-48.

77.   Isernhagen SJ. Functional Capacity Evaluation: Rationale, Procedure, Utility of the Kinesiophysical Approach. Journal of Occupational Rehabilitation 1992;2:157-68.

78.   ILO. Technical and Ethical Guidelines for Workers' Health Surveillance [International Labour Organisation]. 2-9 September 1997, 1997. Available from: http://www.ilo.org/public/english/protection/safework/cops/english/download/e000017.pdf. Accessed 22.11.2011.

79.   Matheson LN, Isernhagen SJ, Hart DL. Relationships among lifting ability, grip force, and return to work. Phys Ther 2002;82:249-56.

80.   Gross DP, Battie MC, Cassidy JD. The prognostic value of functional capacity evaluation in patients with chronic low back pain: part 1: timely return to work. Spine (Phila Pa 1976) 2004;29:914-9.

81. Gross DP, Battie MC. The prognostic value of functional capacity evaluation in patients with chronic low back pain: part 2: sustained recovery. Spine (Phila Pa 1976) 2004;29:920-4.

82. Oliveri M, Jansen T, Oesch P, et al. Comment on Gross et al: The prognostic value of functional capacity evaluation in patients with chronic low back pain: part 1: timely return to work. And part 2: sustained recovery. Spine 2005;30:1232-3; author reply 3-4.

83. Streibelt M, Blume C, Thren K, et al. Value of functional capacity evaluation information in a clinical setting for predicting return to work. Arch Phys Med Rehabil 2009;90:429-34.

84. Reneman MF, Kool J, Oesch P, et al. Material handling performance of patients with chronic low back pain during functional capacity evaluation: a comparison between three countries. Disabil Rehabil 2006;28:1143-9.

85. Soer R, van der Schans CP, Geertzen JH, et al. Normative values for a functional capacity evaluation. Arch Phys Med Rehabil 2009;90:1785-94.

86. Mathiowetz V, Kashman N, Volland G, et al. Grip and pinch strength: normative data for adults. Arch Phys Med Rehabil 1985;66:69-74.

87. Troosters T, Gosselink R, Decramer M. Six minute walking distance in healthy elderly subjects. Eur Respir J 1999;14:270-4.

88. Bohannon R. Six-Minute Walk Test A Meta-Analysis of Data From Apparently Healthy Elders. Topics in Geriatric Rehabilitation 2007;23:155-60.

89. Gross DP, Battie MC. Construct validity of a kinesiophysical functional capacity evaluation administered within a worker's compensation environment. J Occup Rehabil 2003;13:287-95.

90. Hart D. Relation between Three Measures of Function in Patients with Chronic Work Related Pain Syndromes. J Rehabil Outcomes Measures 1998;2:1-14.

91. Reneman MF, Jorritsma W, Schellekens JM, et al. Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. J Occup Rehabil 2002;12:119-29.

92. Lackner JM, Carosella AM. The relative influence of perceived pain control, anxiety, and functional self efficacy on spinal function among patients with chronic low back pain. Spine 1999;24:2254-60; discussion 60-1.

93. Gross D, Battié M. Factors influencing results of functional capacity evaluations in workers' compensation claimants with low back pain. Phys Ther 2005;85:315-22.

94. Waddell G, McCulloch JA, Kummel E, et al. Nonorganic physical signs in low-back pain. Spine 1980;5:117-25.

95. Kummel BM. Nonorganic signs of significance in low back pain. Spine 1996;21:1077-81.

96. Novy DM, Collins HS, Nelson DV, et al. Waddell signs: distributional properties and correlates. Arch Phys Med Rehabil 1998;79:820-2.

97. Dickens C, Jayson M, Creed F. Psychological correlates of pain behavior in patients with chronic low back pain. Psychosomatics 2002;43:42-8.

98. Baker P, Goodman G, Ekelman B, et al. The effectiveness of a comprehensive work hardening program as measured by lifting capacity, pain scales, and depression scores. Work. 2005;24:21-31.

99. Fishbain DA, Cole B, Cutler RB, et al. A structured evidence-based review on the meaning of nonorganic physical signs: Waddell signs. Pain Med 2003;4:141-81.

100. Shaw WS, Pransky G, Fitzgerald TE. Early prognosis for low back disability: intervention strategies for health care providers. Disabil Rehabil 2001;23:815-28.

101. Karas R, McIntosh G, Hall H, et al. The relationship between nonorganic signs and centralization of symptoms in the prediction of return to work for patients with low back pain. Phys Ther 1997;77:354-60; discussion 61-9.

102.   Lechner DE, Bradbury SF, Bradley LA. Detecting sincerity of effort: a summary of methods and approaches. Phys Ther 1998;78:867-88.

103.   Fishbain DA, Cutler RB, Rosomoff HL, et al. Is there a relationship between nonorganic physical findings (Waddell signs) and secondary gain/malingering? Clin J Pain 2004;20:399-408.

104.   Sullivan M. Exaggerated pain behavior: by what standard? Clin J Pain. 2004;20:433-9.

105.   Main CJ, Waddell G. Behavioral responses to examination. A reappraisal of the interpretation of "nonorganic signs". Spine 1998;23:2367-71.

106.   Jay MA, Lamb JM, Watson RL, et al. Sensitivity and specificity of the indicators of sincere effort of the EPIC lift capacity test on a previously injured population. Spine 2000;25:1405-12.

107.   Reneman MF, Fokkens AS, Dijkstra PU, et al. Testing lifting capacity: validity of determining effort level by means of observation. Spine (Phila Pa 1976) 2005;30:E40-6.

108.   Lemstra M, Olszynski WP, Enright W. The sensitivity and specificity of functional capacity evaluations in determining maximal effort: a randomized trial. Spine 2004;29:953-9.

109.   Liebenson C, Yeomans S. Quantification of Physical Performance Ability. In Liebenson C ed. Rehabilitation of the Spine: a practitioner's manual 2ed. Philadelphia: Lippincott Williams & Wilkins, 2007:226-59.

110.   Reneman M, Wittink H. Functional Performance Testing. In Nordin M, Pope M, Andersson G eds. Musculoskeletal disorders in the workplace: principles and practice. Philadelphia: Mosby, 2007:397-408.

111.   Robinson ME, Dannecker EA. Critical issues in the use of muscle testing for the determination of sincerity of effort. Clin J Pain 2004;20:392-8.

112.   Fishbain DA, Cutler R, Rosomoff HL, et al. Chronic pain disability exaggeration/malingering and submaximal effort research. Clin J Pain 1999;15:244-74.

113.   Kaplan GM, Wurtele SK, Gillis D. Maximal effort during functional capacity evaluations: an examination of psychological factors. Arch Phys Med Rehabil 1996;77:161-4.

114.   SAR. Functional Capacity Evaluation - Course Manual 1ed. Bellikon, Switzerland: Verein IG Ergonomie SAR, 2007.

115.   Karjalainen K, Malmivaara A, van Tulder M, et al. Multidisciplinary biopsychosocial rehabilitation for subacute low back pain among working age adults. Cochrane Database Syst Rev 2003:CD002193.

116.   Guzman J, Esmail R, Karjalainen K, et al. Multidisciplinary rehabilitation for chronic low back pain: A systematic review. British Medical Journal 2001;322:1511-6.

117.   Schonstein E, Kenny DT, Keating J, et al. Work conditioning, work hardening and functional restoration for workers with back and neck pain. Cochrane Database Syst Rev 2003:CD001822.

118.   Henschke N, Ostelo RW, van Tulder MW, et al. Behavioural treatment for chronic low-back pain. Cochrane Database Syst Rev 2010:CD002014.

119.   Ostelo R, van Tulder M, Vlaeyen J, et al. Behavioural treatment for chronic low-back pain. Cochrane Database of Systematic Reviews 2005;No.: CD002014. DOI: 10.1002/14651858.CD002014.pub2.

120.   Mayer T, Gatchel R. Functional Restoration for Spinal Disorders: The Sports Medicine Approach. Philadelphia: Lea and Febiger, 1988.

121.   Lindström I, Öhlund C, Eek C, et al. Mobility, strength, and fitness after a graded activity program for patients with subacute low back pain. A randomized prospective clinical study with a behavioral therapy approach. Spine 1992;17:641-52.

122. Hayden JA, van Tulder MW, Tomlinson G. Systematic review: strategies for using exercise therapy to improve outcomes in chronic low back pain. Ann Intern Med 2005;142:776-85.

123. Hayden JA, van Tulder MW, Malmivaara AV, et al. Meta-analysis: exercise therapy for nonspecific low back pain. Ann Intern Med 2005;142:765-75.

124. Kool J, de Bie R, Oesch P, et al. Exercise reduces sick leave in patients with non-acute non-specific low back pain: a meta-analysis. J Rehabil Med 2004;36:49-62.

125. Hlobil H, Staal JB, Spoelstra M, et al. Effectiveness of a return-to-work intervention for subacute low-back pain. Scand J Work Environ Health 2005;31:249-57.

126. van Tulder M, Furlan A, Bombardier C, et al. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine 2003;28:1290-9.

127. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. Bmj 2001;323:42-6.

128. Schonstein E, Kenny D, Keating J, et al. Physical conditioning programs for workers with back and neck pain: a cochrane systematic review. Spine 2003;28:E391-5.

129. Biering-Sorensen F. Physical measurements as risk indicators for low-back trouble over a one-year period. Spine (Phila Pa 1976) 1984;9:106-19.

130. Keller A, Hellesnes J, Brox JI. Reliability of the isokinetic trunk extensor test, Biering-Sorensen test, and Astrand bicycle test: assessment of intraclass correlation coefficient and critical difference in patients with chronic low back pain and healthy individuals. Spine 2001;26:771-7.

131. Perret C, Poiraudeau S, Fermanian J, et al. Validity, reliability, and responsiveness of the fingertip-to-floor test. Arch Phys Med Rehabil 2001;82:1566-70.

132. Reese NB, Bandy WD. Use of an inclinometer to measure flexibility of the iliotibial band using the Ober test and the modified Ober test: differences in magnitude and reliability of measurements. J Orthop Sports Phys Ther 2003;33:326-30.

133. Matheson LN, Matheson M. Spinal function sort, rating of perceived capacity. Performance Assessment and Capacity Testing PACTed. Trabuco Canyon, California: Performance Assessment and Capacity Testing PACT, 1989.

134. Waddell G, Newton M, Henderson I, et al. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993;52:157-68.

135. Storheim K, Brox JI, Holm I, et al. Predictors of return to work in patients sick listed for sub-acute low back pain: a 12-month follow-up study. Journal of rehabilitation medicine : official journal of the UEMS European Board of Physical and Rehabilitation Medicine 2005;37:365-71.

136. Ferraz MB, Quaresma MR, Aquino LR, et al. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. J Rheumatol 1990;17:1022-4.

137. van der Giezen AM, Bouter LM, Nijhuis FJ. Prediction of return-to-work of low back pain patients sicklisted for 3-4 months. Pain 2000;87:285-94.

138. Isernhagen SJ, Hart DL, Matheson LM. Reliability of independent observer judgments of level of lift effort in a kinesiophysical Functional Capacity Evaluation. Work 1999;12:145-50.

139. Reneman MF, Jaegers SM, Westmaas M, et al. The reliability of determining effort level of lifting and carrying in a functional capacity evaluation. Work 2002;18:23-7.

140. Reneman MF, Dijkstra PU, Westmaas M, et al. Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. Journal of Occupational Rehabilitation 2002;12:269-75.

141. Brouwer S, Reneman MF, Dijkstra PU, et al. Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in patients with chronic low back pain. Journal of Occupational Rehabilitation 2003;13:207-18.

142. Reneman MF, Brouwer S, Meinema A, et al. Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in healthy adults. J Occup Rehabil 2004;14:295-305.

143.   Gross D, Battié M. Reliability of safe maximum lifting determinations of a functional capacity evaluation. Phys Ther 2002;82:364-71.

144.   Mathiowetz V, Weber K, Volland G, et al. Reliability and validity of grip and pinch strength evaluations. J Hand Surg [Am] 1984;9:222-6.

145.   ATS. American Thoracic Society statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med 2002;166:111-7.

146.   King MB, Judge JO, Whipple R, et al. Reliability and responsiveness of two physical performance measures examined in the context of a functional training intervention. Phys Ther 2000;80:8-16.

147.   Kennedy DM, Stratford PW, Wessel J, et al. Assessing stability and change of four performance measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty. BMC Musculoskelet Disord 2005;6:3.

148.   Zweig M, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry 1993;39:561-77.

149.   Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.

150.   Oesch P, Meyer K, Janssen B, et al. What is the Role of 'Nonorganic-Somatic-Components' in Functional Capacity Evaluations in Patients with Chronic Non-Specific Low Back Pain Undergoing Fitness for Work Evaluation? Spine (Phila Pa 1976) 2011.

151.   Bryman A. Social Research methods. 2 ed. Oxford: Oxford University Press, 2004.

152.   Polit D, Beck C. Developing and Testing Self-Report Scales. Nursing Research, Generating and Assessing Evidence for Nursing Practice. 8 ed. Philadelphia: Lippingcott Williams & Wilkins, 2008:474-505.

153.   Isernhagen S. Contemporary issues in functional capacity evaluation. In Isernhagen S 1ed. The Comprehensive Guide To Work Injury Management. Gaithersburg, MD: Aspen Publishers, 1995:410-29.

154.   Hamilton L. Regression with Graphics: A Second Course in Applied Statisticsed. Pacific Grove, California: Brooks/Cole Publishing Company, 1991.

155.   Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. Stat Med 2000;19:3127-31.

156.   Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. Psychol Bull 1995;117:167-78.

157.   DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177-88.

158.   Reitman RD, Emerson RH, Higgins LL, et al. A multimodality regimen for deep venous thrombosis prophylaxis in total knee arthroplasty. J Arthroplasty 2003;18:161-8.

159.   Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. J Clin Epidemiol 2001;54:1046-55.

160.   Hagen EM, Eriksen HR, Ursin H. Does early intervention with a light mobilization program reduce long-term sick leave for low back pain? Spine (Phila Pa 1976) 2000;25:1973-6.

161.   Harkäppää K, Jarvikoski A, Mellin G, et al. A controlled study of the outcome of inpatient and outpatient treatment of low back pain. Part 1. Pain Disability, compliance and reported treatment benefits three months after treatment. Scand J Rehabil Med 1989;21:81-90.

162.   Heymans MW, de Vet HC, Bongers PM, et al. The effectiveness of high-intensity versus low-intensity back schools in an occupational setting: a pragmatic randomized controlled trial. Spine 2006;31:1075-82.

163.   Hurri H. The Swedish back school in chronic low back pain. Part I. Benefits. Scand J Rehabil Med 1989;21:33-40.

164. Karjalainen K, Malmivaara A, Pohjolainen T, et al. Mini-intervention for subacute low back pain: a randomized controlled trial. Spine 2003;28:533-40; discussion 40-1.

165. Niemisto L, Lahtinen-Suopanki T, Rissanen P, et al. A randomized trial of combined manipulation, stabilizing exercises, and physician consultation compared to physician consultation alone for chronic low back pain. Spine 2003;28:2185-91.

166. Skouen JS, Grasdal AL, Haldorsen EM, et al. Relative cost-effectiveness of extensive and light multidisciplinary treatment programs versus treatment as usual for patients with chronic low back pain on long-term sick leave: randomized controlled study. Spine 2002;27:901-9.

167. Staal JB, Hlobil H, Twisk JW, et al. Graded activity for low back pain in occupational health care: a randomized, controlled trial. Ann Intern Med 2004;140:77-84.

168. Steenstra IA, Anema JR, Bongers PM, et al. The effectiveness of graded activity for low back pain in occupational healthcare. Occup Environ Med 2006;63:718-25.

169. Storheim K, Brox JI, Holm I, et al. Intensive group training versus cognitive intervention in sub-acute low back pain: short-term results of a single-blind randomized controlled trial. J Rehabil Med 2003;35:132-40.

170. Torstensen TA, Ljunggren AE, Meen HD, et al. Efficiency and costs of medical exercise therapy, conventional physiotherapy, and self-exercise in patients with chronic low back pain. A pragmatic, randomized, single-blinded, controlled trial with 1-year follow-up. Spine 1998;23:2616-24.

171. White AW. Low back pain in men receiving workmen's compensation. Can Med Assoc J 1966;95:50-6.

172. Alaranta H, Rytokoski U, Rissanen A, et al. Intensive physical and psychosocial training program for patients with chronic low back pain. A controlled clinical trial. Spine 1994;19:1339-49.

173. Aure OF, Nilsen JH, Vasseljen O. Manual therapy and exercise therapy in patients with chronic low back pain: a randomized, controlled trial with 1-year follow-up. Spine 2003;28:525-31; discussion 31-2.

174. Jousset N, Fanello S, Bontoux L, et al. Effects of functional restoration versus 3 hours per week physical therapy: a randomized controlled study. Spine 2004;29:487-93; discussion 94.

175. Kääpä EH, Frantsi K, Sarna S, et al. Multidisciplinary group rehabilitation versus individual physiotherapy for chronic nonspecific low back pain: a randomized trial. Spine 2006;31:371-6.

176. Petersen T, Kryger P, Ekdahl C, et al. The effect of McKenzie therapy as compared with that of intensive strengthening training for the treatment of patients with subacute or chronic low back pain: A randomized controlled trial. Spine 2002;27:1702-9.

177. Roche G, Ponthieux A, Parot-Shinkel E, et al. Comparison of a functional restoration program with active individual physical therapy for patients with chronic low back pain: a randomized controlled trial. Arch Phys Med Rehabil 2007;88:1229-35.

178. Bendix AF, Bendix T, Ostenfeld S, et al. Active treatment programs for patients with chronic low back pain: a prospective, randomized, observer-blinded study. Eur Spine J 1995;4:148-52.

179. Bendix AF, Bendix T, Vaegter K, et al. Multidisciplinary intensive treatment for chronic low back pain: a randomized, prospective study. Cleve Clin J Med 1996;63:62-9.

180. Bendix T, Bendix A, Labriola M, et al. Functional restoration versus outpatient physical training in chronic low back pain: a randomized comparative study. Spine 2000;25:2494-500.

181. Härkäpää K, Mellin G, Jarvikoski A, et al. A controlled study on the outcome of inpatient and outpatient treatment of low back pain. Part III. Long-term follow-up of pain, disability, and compliance. Scand J Rehabil Med 1990;22:181-8.

182. Shechtman O, Sindhu BS, Davenport PW. Using the force-time curve to detect maximal grip strength effort. J Hand Ther 2007;20:37-47; quiz 8.

183.   Boneau C. The effects of violations of assumptions underlying the t-test. Psychol Bull 1960:49-64.

184.   Anema JR, Steenstra IA, Bongers PM, et al. Multidisciplinary rehabilitation for subacute low back pain: graded activity or workplace intervention or both: a randomized controlled trial [with consumer summary]. Spine 2007;32:291-8.

185.   Herbert RD, Bo K. Analysis of quality of interventions in systematic reviews. Bmj 2005;331:507-9.

186.   Wilmore J, Costill D, Kenney W. Physiology of Sport and Exercise Science. 4 ed. Hong Kong: Human Kinetics, 2008.

Paper I

Perceived functional ability assessed with the spinal function sort: is it valid for European rehabilitation settings in patients with nonspecific non-acute low back pain?

Paper II

What is the Role of 'Nonorganic-Somatic-Components' in Functional Capacity Evaluations in Patients with Chronic Non-Specific Low Back Pain Undergoing Fitness for Work Evaluation?

Oesch P, Meyer K, Jansen B, Mowinckel P, Bachmann S, Hagen KB

What is the role of 'nonorganic-somatic-components' in Functional Capacity Evaluations in patients with chronic non-specific low back pain undergoing fitness for work evaluation?

Peter Oesch, MSc[1,2]; Kathrin Meyer, MPH[5]; Beatrice Jansen, MSc[4]; Petter Mowinckel, MSc[6], Stefan Bachmann, MD[1,2,3]; Kare Birger Hagen, PhD[6,7]

1 Research Department, Rehabilitation Centre Valens, Switzerland; 2 Department of Rheumatology, Rehabilitation Centre Valens, Switzerland; 3 Department of Geriatrics, Inselspital, Bern University Hospital, Switzerland; 4 Departement of Work Rehabilitation, Rehaklinik Bellikon; 5 Division of Rheumatology and Institute of Physical Medicine, Zurich University Hospital; 6 National Resource Centre for Rehabilitation in Rheumatology, Diakonhjemmet Hospital, Norway; 7 Institute of Health and Society, Department of Health Sciences, University of Oslo Norway

Abstract

Study Design: Analytical cross-sectional study

Objective: To assess the association of 'nonorganic-somatic-components' together with physical and other psychosocial factors on Functional Capacity Evaluation (FCE) in patients with chronic non-specific low back pain (NSLBP) undergoing fitness-for-work evaluation.

Summary of Background Data: FCE is increasingly used for physical fitness-for-work evaluation in patients with chronic NSLBP, but results seem to be influenced by physical as well as psychosocial factors. The influence of 'nonorganic-somatic-components' together with physical and other psychosocial factors on FCE performance has not yet been investigated.

Methods: 126 patients with chronic NSLBP referred for physical fitness-for-work evaluation were included. The four FCE tests were lifting from floor to waist, forward bend standing, grip strength, and six minute walking. 'Nonorganic-somatic-components' were assessed with the eight 'nonorganic-somatic-signs' as defined by Waddell, and were adjusted for age, gender, days off work, salary in the previous occupation, pain intensity, fear avoidance belief, and perceived functional ability in multivariate regression analyses.

Results: Between 42% and 58% of the variation in the FCE tests was explained in the final multivariate regression models. 'Nonorganic-somatic-components' were consistent independent predictors for all tests. Their influence was most important on forward bend standing and walking distance, and less on grip strength and lifting performance. The physical factors of age and/or gender were strongly associated with grip strength and lifting, less with walking distance and not at all with forward bend standing. The influence of at least one other psychosocial factor was observed in all FCE tests, having the highest proportion in the six minute walking test.

Conclusions: 'Nonorganic-somatic-components' seem to be consistent independent predictors in FCE testing and should be considered for interpretation of test results.

Introduction

Low back pain (LBP) continues to be a major health problem in Western countries, causing an increase in rehabilitation allowances, sickness absence and disability pensions [1,2]. A decision on a patient's fitness-for-work is mostly based upon a medical evaluation and the resulting medical diagnosis. However, there are now reforms in progress in many countries to move away from an 'essentialist' diagnostic approach [3]. Fitness-for-work certificates based on diagnosis are criticised on the grounds that few objective physical or biomechanical measures are associated with return-to-work [4-6] and a person's potential work ability is not explored [3]. Due to these shortcomings of the diagnosis-based physical fitness-for-work evaluations, the so-called Functional Capacity Evaluation (FCE) has been developed. This standardized battery of clinical tests intends to measure a patient's safe physical ability to carry out work-related activities [7].

There is increasing evidence that not only physical but also psychosocial factors influence FCE results. Performance during FCE is associated with pain intensity [8], perceived disability [8-10], and functional self-efficacy [11]. A previous study investigating factors influencing results of FCE in Workers' Compensation Claimants with LBP found that FCE reflects physical capacity to some degree but also found influences of perceived disability and pain intensity. It is therefore proposed that FCE should be considered as behavioral tests influenced by multiple factors, including physical ability and psychosocial factors. However, the authors state that there must have been some important determinants of physical performance that were not measured as they were unable to explain large amounts of the variation in FCE performance [12]. A possible confounding factor related to FCE performance might have been the presence of 'nonorganic-somatic-components' within the physical examination. Waddell et al. 1980 described eight 'nonorganic-somatic-signs' (see Table 1) that are distinguishable from the standard clinical signs of physical pathology in patients with low back pain. By helping to separate the physical from the nonorganic, they clarify the assessment of purely physical pathologic conditions [13].

The identification of 'nonorganic-somatic-components' within the physical examination has a long standing history in medical examination but is debated. A major criticism is that such findings are frequently interpreted as evidence of malingering [14-16]. In a later published reappraisal of the interpretation of their 'nonorganic-somatic-signs', the authors emphasise that these should not be used as evidence of simulation for the purpose of financial gain, but should be viewed as a form of communication between patient and examiner [17]. Such a behavioral response to examination is influenced by expectations and must be understood in the context of the patient's history. It has been shown that 'nonorganic-somatic-components' correlate with illness behavior and distress [13,18], as well as with increased disability [19-21] and a poorer rate of return to work [22,23].

To our knowledge, the influences of 'nonorganic somatic components', together with physical and other psychosocial factors, on the results of an FCE have not yet been investigated and this,

therefore, is the purpose of this study. We hypothesise that lower FCE performance in patients with chronic non-specific low back pain (NSLBP) is associated with physical factors such as sex and age, as well as with higher levels of perceived disability and symptom reporting. However, we also theorize that the inclusion of 'nonorganic-somatic-components' as an explanatory variable will substantially increase the explained variation in FCE performance.

Table 1: The eight 'nonorganic-somatic-signs' 13

Test categories	'nonorganic-somatic-signs'
Tenderness	1. 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. 2. 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	3. 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. 4. 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. In the presence of root irritation, leg pain may be produced and should be discounted.
Distraction Test	5. Straight Leg Raising Straight leg raising is the most useful distraction test. The subject whose back pain has a nonorganic component shows marked improvement in straight leg raising on distraction as compared with formal testing.
Regional disturbances	6. Sensory Sensory disturbances include diminished sensation to light touch, pinprick, and sometimes other modalities fitting a "stocking" rather than a dermatomal pattern. 7. 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	8. Overreaction during examination may take the form of disproportionate verbalisation, facial expressions, muscle tension and tremor, collapsing, or sweating. 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.

A category is positive if at least one 'nonorganic-somatic-signs' in that category is positive. Three positive categories are required indicating that a patient with LBP does not have a straightforward physical problem.

Materials and Methods

A cross-sectional study was performed in three rehabilitation clinics in Switzerland. Evaluations were performed by two independent assessors. The first assessor administered the questionnaires and assessed 'nonorganic-somatic-signs' and 'grip strength'. The second assessor tested lifting, forward bend standing and walking. All second assessors were trained FCE therapists. In addition, a FCE instructor discussed eight FCE video sequences with the

FCE therapists regarding test performance and trained the first assessors to assess the eight 'nonorganic-somatic-signs' (see Table 1).

Patients with chronic NSLBP [24,25] as their primary complaint, aged 20 to 60 years, referred for fitness-for-work evaluation, and with sufficient understanding of German, French or Italian to follow the instructions during FCE were included. Patients with specific LBP due to nerve root compression, vertebral fracture, tumor, infection, inflammatory diseases, spondylolisthesis, spinal stenosis and definite instability [24,25] or with relevant comorbidity such as cardio-respiratory, psychiatric or other musculoskeletal problems affecting work ability were excluded.

Dependent variables

An FCE as described by Isernhagen [7] employs 28 tests administered over two days. While an early review found only limited scientific evidence for the reliability and validity of FCE [26,27] there is now an increasing body of knowledge on the clinimetric properties of the Isernhagen FCE providing the most comprehensive coverage of all aspects of reliability and validity among FCE methods [28]. The 28 tests can be categorized in manual handling capacity tests, work postures, hand capacity and ambulation [7]. For the purpose of this study, one test from each category was employed. We included 'manual handling from floor to waist', thereby allowing a comparison with previous research investigating the influence of non-physical factors on lifting performance. In addition, we wanted to investigate whether representative tests from the other FCE categories were also influenced by 'nonorganic-somatic-components' and chose the ones with reported reliability.

Lifting from floor to waist

Patients were instructed to perform five lifts, in which they would move as much weight as safely possible from a shelf at waist height to the floor, and then place the weight back on the shelf. The test began with a light weight which was increased. Five lifts were repeated with each weight until the maximum safe weight was reached, or the patient stopped lifting [7]. The maximum weight that was lifted five times was recorded. Lifting capacity assessment from floor to waist has been shown to be reliable [29] and predictive for return to work [30,31]

Forward bend standing

Patients were instructed to maintain a 30 degree forward bending trunk position for 5 minutes without straightening the hips or back while performing a hand activity. Test performance in seconds was measured by the observer with a stopwatch. The test showed acceptable test-retest reliability in patients with chronic low back pain [32].

Grip strength measured in kg

Grip strength measurements in kilogram were taken with a handheld G200 Dynamometer (Biometrics Ltd, Tampa, FL, version 8 software), according to a standardized and reliable protocol [33,34]. Patients were instructed to squeeze the Dynamometer as hard as possible for three tests in alternating order with the right and left hand in 5 different handle positions. The mean grip-strength measurements of each handle position were calculated and the maximum grip strength was recorded.

Six minute walking test

The six minute walking test was performed according to the recommendations of the American Thorax Society [35]. Participants were instructed that the objective was to walk as far as possible for six minutes, without running or jogging, by walking back and forth along a distance of 30 meters marked by traffic cones. The test performance was measured by the observer in meters walked. Reliability of the six minute walking test has been proven high in various patient groups [36,37].

Independent variables

Physical factors

A subject's age and sex were shown to be performance predictors in lifting from floor to waist, grip strength and forward bend standing [38] as well as for six minute walking distance [39].

Psychosocial factors

Psychosocial yellow flags are known risk factors for long-term disability and work loss associated with low back pain [40]. We screened for the following psychosocial yellow flags:

Nonorganic-somatic-components'

The 'nonorganic-somatic-components' were assessed with the eight 'nonorganic-somatic-signs' (see Table 1). These are grouped into five categories. A category is positive if at least one 'nonorganic-somatic-sign' in that category is positive. Three categories are required to be positive to indicate that a patient with LBP does not have a straightforward physical problem [13]. To optimize the homogeneity and reliability of the score for the purpose of this study, we summed up the individual signs according to Apeldoorn et al. (score 0 – 8) instead of using the dichotomized score [41]. This approach would also allow within linear regression analysis to estimate the increase in the dependent variables corresponding to a one or more point increase in 'nonorganic-somatic-signs.

Compensation issues

Economic aspects have an impact on return to work [42]. Low previous salary was assumed as having a negative influence on FCE performance. We collected information about patients' salaries in their previous job and days off work from the clinical database.

Pain intensity

Pain intensity was measured with a numeric rating scale (0 - 10) [43] at the moment of assessment.

Fear avoidance beliefs (FABs)

High FABs are negative predictors for return to work [44]. The FAB questionnaire provides a score for fear of physical activity ranging from 0 – 24 and one for work activities ranging from 0 – 42 [45]. We chose the FABQ score for work activities as an independent variable because of the work-related functional testing.

Perceived functional ability

The Spinal Function Sort (SFS) was used to capture perceived functional ability for work tasks. This questionnaire was developed in the United States [46] and consists of 50 graphically depicted tasks with simple descriptions. Its validity in a European rehabilitation setting has been shown [47].

Power calculation

Based on the results found by Gross et al. [12] we assumed R^2 to be at least 0.15 giving an effect size of 0.18. With alpha defined as 0.05 and beta as 0.9 and 8 independent variables we calculated a minimal sample size of 114 subjects needed for multiple regression analysis.

Data analyses

Our main goal was to assess the association of 'nonorganic-somatic-components' with lifting from floor to waist, forward bend standing, grip strength and six minute walking distance in a regression analysis. We chose to correct for age, gender, days off work, salary in the previous job, pain intensity, fear avoidance belief for work, and perceived functional ability in our analysis. Model building started with bivariate models for each dependent variable. We planned to retain all independent variables with p<0.2 in the first multivariate model and then remove the variables with the highest p-values. If a slope coefficient for a variable changed by more than 30% when removed, it was retained in the analysis as a confounder. The final model included the confounders plus the variables with p≤05. An examination of the models revealed violations of the normality as well as homoscedasticity assumptions. Consequently, we resorted to Robust regression analysis [48]. The analysis is an iterative procedure that seeks to identify the outliers and down-weights their influence on the regression coefficient estimates. This analysis was performed using Number Cruncher Statistical System (NCSS, version 07.1.19, Kaysville, Utah). All other analyses were performed with Statistical Package for Social Sciences (SPSS Version 18). Data analysis was performed by an independent statistician (PMo) who was not involved in conducting the study and data collection.

Ethical approval

Was obtained from the three regional ethics committees (EKSG 08/029/2B; SPUK N°. 784, EKAG 08/058) where the rehabilitation clinics are located.

Results

678 FCEs were performed in the three rehabilitation clinics from March 2009 to August 2010. Treating physicians identified 203 subjects with LBP of whom 26 suffered from specific LBP, and 16 from relevant comorbidity affecting work ability. 5 were excluded due to language problems, 11 subjects did not give informed consent, 2 were older than 60 and 13 were missed for inclusion, leaving 130 subjects with chronic NSLBP of which full data was available in 126 cases. 11 physiotherapists with a mean professional experience of 12.1 years (SD 6.9) acted as first assessors. 17 physiotherapists performed FCE. These had extensive experience in FCE testing, having performed on average 36.5 FCEs (SD 10.0) in the two years prior to this study.

Mean time off work was over two years. 52% of the included patients had performed heavy or very heavy work in their previous job. Mean lifting capacity was comparable to earlier published samples of patients with chronic low back pain from Switzerland [49] and from Canada [12]. Time for forward bend standing was comparable to a patient sample with chronic LBP achieving 141 sec (SD 101.5) [32]. Grip strength was within the norm [34] and mean walking distance in six minutes was below a reference value of 499 meters (95%CI 480–519m) established from healthy persons over 60 years of age [39]. There was a high level of fear avoidance beliefs and low perceived functional ability (see Table 2).

Table 2: Subjects characteristics (n=126)

Independent variables	n, mean (SD)	
Male/female	94/32	
Age (years)	44.1	(10.4)
Previous salary per month (CHF)	4481.9	(2209.4)
Days off work	670.1	(1031.0)
Fear avoidance belief work activities (FABQ 0-42)	32.7	(9.0)
Percevied functional ability (SFS 0-200)	96.3	(50.9)
Pain intensity (NRS 0 – 10)	5.1	(2.2)
'Nonorganic-somatic-signs' (1-8)	2.3	(2.1)
Dependent variables		
Floor to waist lift (kg)	16.8	(10.7)
Forward bend standing (sec)	188.7	(101.5)
Six minute walking test (meter)	462.3	(144.5)
Grip strength dominant hand (kg)	33.3	(15.3)

CHF indicates Swiss francs; FABQ, Fear avoidance belief questionnaire; SFS, Spinal Function Sort; NRS, Numeric rating scale

In the bivariate regression analysis, p was below 0.2 in all variables. All variables were entered in the first multivariate models explaining between 37% and 61% of the variance in FCE performance (see Table 3). Further reduction of variables with high p-values led to four significant models explaining between 42% and 58% of the variance (see Table 4). The 'nonorganic-somatic-signs' were significant independent predictors for all FCE results.

Self-perceived functional ability was the most important predictor for lifting performance whereby a 10 point increase in SFS score was associated with a 1.1kg weight increase, followed by the 'nonorganic-somatic-signs' accounting for a 1kg decrease, and female lifting 4.7kg less than male. The 'nonorganic-somatic-signs' were the most influential predictors for forward bend standing. A one point increase was associated with a 20.5sec decrease in performance. Days off work and perceived functional ability were also retained as significant predictors in the final model. Gender was most influential for grip strength with women having 16kg less mean grip strength than men.

Table 3: The first multivariate model containing all independent variables (n=126)

Independent Variable	Lifting from floor to waist			Forward bend standing			Grip strength			Six minute walking distance		
Adjusted R2	**0.37**			**0.38**			**0.61**			**0.43**		
	Unst. Coef	Std. Error	Sig.	Unst. Coef	Std. Error	Sig.	Unst. Coef	Std. Error	Sig.	Unst. Coef	Std. Error	Sig.
'Nonorganic-somatic-components'	-0.88	0.39	0.03	-17.31	4.45	<.01	-1.13	0.49	0.02	-27.17	5.45	<.01
Age	0.01	0.06	0.90	0.65	0.73	0.38	-0.22	0.08	0.01	-1.72	0.88	0.05
Gender	4.20	1.57	0.01	-12.43	17.97	0.49	15.84	2.00	<.01	23.43	21.44	0.28
Days off work	0.00	0.00	0.55	-0.03	0.01	<.01	0.00	0.00	0.17	-0.01	0.01	0.18
Salary previous job	0.00	0.00	0.15	0.00	0.00	0.97	0.00	0.00	0.16	0.01	0.00	0.04
Pain	-0.09	0.41	0.83	-4.52	4.66	0.33	-1.25	0.54	0.02	-9.81	5.69	0.09
FAB's work activities	-0.04	0.08	0.65	-0.71	0.91	0.44	0.04	0.10	0.68	-2.21	1.08	0.04
Perceived functional ability	0.11	0.02	<0.01	0.22	0.21	0.29	0.08	0.02	<.01	0.14	0.25	0.58

Unst. Coef indicates unstandardized coefficients

A 10 point increase in SFS score was associated with a 1.1kg grip strength increase, and a one point increase in 'nonorganic-somatic-signs' with 1.5kg decrease in grip strength. Younger patients had more grip strength than older patients. The 'nonorganic-somatic-signs' were also the most influential predictor for walking distance whereby a one point increase was associated with a 27m decrease in distance walked. Three further psychosocial factors of similar importance were associated with walking distance. A 10 year increase in age was associated with 19m decrease in walking distance.

Table 4: Final robust regression models (n=126)

FCE tests	Adj. R2	Final model	Unst. Coef	95% CI for the unst. Coefficient	Sig.	Std. Coeff.
Lifting from floor to waist (kg)	**0.54**	Perceived functional ability	0.11	(0.08 , 0.14)	0.001	0.57
		Gender (male)	4.73	(1.90 , 7.55)	<.001	0.20
		'Nonorganic-somatic-components'	-0.95	(-1.66 , -0.24)	0.009	-0.20
Forward bend standing (sec)	**0.42**	'Nonorganic-somatic-components'	-20.49	(-28.13 , -12.85)	<.001	-0.44
		Days off work	-0.03	(-0.04 , -0.02)	<.001	-0.30
		Perceived functional ability	0.31	(-0.02 , 0.63)	*0.065	0.16
Grip strength dominant hand (kg)	**0.58**	Gender (male)	15.97	(12.03 , 19.90)	<.001	0.47
		Perceived functional ability	0.11	(0.07 , 0.15)	<.001	0.38
		'Nonorganic-somatic-components'	-1.53	(-2.51 , -0.54)	0.003	-0.23
		Age	-0.25	(-0.41 , -0.08)	0.005	-0.17
Six minute walking distance (m)	**0.52**	'Nonorganic-somatic-components'	-27.13	(-37.16 , -17.11)	<.001	-0.43
		Salary previous job	0.01	(0.00 , 0.02)	0.002	0.20
		Pain intensity	-11.65	(-21.26 , -2.04)	0.018	-0.19
		Fear avoidance belief work activities	-2.50	(-4.47 , -0.52)	0.014	-0.17
		Age	-1.90	(-3.56 , -0.24)	0.025	-0.15

* Not significant, but a confounder

Unstandardized Regression Coefficient is the slope of the regression line. It is the increase in y (the dependent variable) corresponding to a unit increase in x the (independent) variable, Std. Coeff, Standardized Coefficients are an attempt to make the regression coefficients more comparable. These were used to list the variables in order of importance, but not for interpretation.

Discussion

This study confirms previous findings showing that FCE test results are influenced by physical as well as by non-physical factors [12]. A new finding of this study is the consistent independent prediction of the 'nonorganic-somatic-signs' for FCE performance in all tests. Their influence was most significant on forward bend standing and walking distance, and less significant on grip strength and lifting performance. The physical factors age and/or gender were strongly associated with grip strength and lifting, less with walking distance and not at all with forward bend standing. As hypothesized, the inclusion of 'nonorganic-somatic-components' has led to substantially higher amounts of explained variation in lifting performance (54%) than were found in a previous model [12] consisting of perceived disability due to pain, age, and gender explaining 20%, 6%, respectively 5% of lifting performance.

A striking finding of this study is the influence of 'nonorganic-somatic-components' on tests without any specific load on the spine, as observed in grip strength. Without this additional information on the presence of 'nonorganic-somatic-components' physical disability relating to grip strength would have been interpreted. The same applies to the lower mean walking distance observed compared to a considerably older population of healthy adults, suggesting disability in walking arising from low back pain. However, 'nonorganic-somatic-components' had the most significant and clinical meaningful influence, with a one point increase in 'nonorganic-somatic-signs' accounting for a 27m decrease in walking distance. We therefore hypothesize that FCE influenced from 'nonorganic-somatic-components' should not solely be interpreted as a reflection of the remaining physical function of patients with back pain but indeed as behavioral tests influenced by non-physical factors. This statement must not be misunderstood as a call to use the 'nonorganic-somatic-signs' as a screening tool for malingering. In accordance with previous authors, we interpret such behavior as a form of communication between patient and examiner influenced by expectations [17] and possibly arising from pain, fear of injury or neuromuscular inhibition [17,50]. Some of these patients may require a more careful examination and management of the psychosocial and behavioral aspects of their illness [17].

The question arises as to what cut-off point in the number of 'nonorganic-somatic-signs' should be used to declare an FCE as not being representative of physical abilities. The recommendation of Waddell et al. to interpret three or more positive categories as indicators of 'nonorganic-somatic-components' within the physical assessment requires at least three positive 'nonorganic somatic signs' out of the total eight signs (see Table 1). This leads to clinically meaningful changes in three out of the four test results: i.e. 61.6sec decrease in forward bend standing, 4.5kg decrease in grip strength, and 81m decrease in walking distance. We therefore feel encouraged to use the proposed cut-off point as defined by Waddell et al. [13].

One might criticize that 52% of the patient population's previous jobs had been heavy or very heavy and this was not accounted for in the analysis as a possible physical risk factor for long term disability. Instead, we used salary in the previous job as an explanatory variable. Both are associated with each other but salary in the previous job is also associated with other known

physical risk factors such as repetitive assembly work or monotonous static work, and is an important compensation issue in Switzerland as it serves for final disability determination. However, we do admit that the inclusion of further physical work factors might explain a higher amount of variation in FCE performance.

A weakness of this study is that it was conducted in a single country with a highly established social insurance system reducing its external validity. It might be further argued that the assessment of the dependent variable 'grip strength' by the first assessor is a weakness of this study. The intention was to blind the FCE therapist towards the frequently used grip curve assessment as a proxy for 'nonorganic-somatic-components' within grip curve assessment. To prevent prejudice in the first assessor, he assessed the 'nonorganic-somatic-signs' before assessing grip strength. We therefore feel that the assessment of one independent variable by the first assessor is justified and does not diminish the validity of the study's findings.

Strengths of this study are the multicenter setting involving a large number of assessors, the inclusion of patients of several nationalities, which raises the external validity of the study's findings, and the blinding of the two assessors against each other's results.

Further research is needed into the confusing mix of methods and terminologies to assess 'nonorganic-somatic-components' within an FCE. A validity check identifying 'nonorganic-somatic-components' is implemented within the FCE design by assessing the level of effort a patient had given through observational criteria [7]. These might identify the same patients as the 'nonorganic-somatic-signs'.

Literature

1. Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. Bull World Health Organ 2003;81:646-56.

2. Manchikanti L, Singh V, Datta S, et al. Comprehensive review of epidemiology, scope, and impact of spinal pain. Pain Physician. 2009;12:E35-70.

3. Bolderson H, Mabbett D, Hvinden B. Definitions of Disability in Europe: A Comparative Analysised. Brussels, Belgium: European Commission Directorate General for Employment and Social Affairs, 2002.

4. Milhous RL, Haugh LD, Frymoyer JW, et al. Determinants of vocational disability in patients with low back pain. Arch Phys Med Rehabil 1989;70:589-93.

5. McIntosh, Frank J, Hogg-Johnson S, et al. Prognostic Factors for Time Receiving Workers' Compensation Benefits in a Cohort of Patients With Low Back Pain. Spine 1999;25:147–57.

6. Hunt DG, Zuberbier OA, Kozlowski AJ, et al. Are components of a comprehensive medical assessment predictive of work disability after an episode of occupational low back trouble? Spine 2002;27:2715-9.

7. Isernhagen SJ. Functional capacity evaluation. Work Injury: Management and Prevention. Gaithersburg: Aspen Publishers Inc, 1988:139–91.

8. Gross DP, Battie MC. Construct validity of a kinesiophysical functional capacity evaluation administered within a worker's compensation environment. J Occup Rehabil 2003;13:287-95.

9. Hart D. Relation between Three Measures of Function in Patients with Chronic Work Related Pain Syndromes. J Rehabil Outcomes Measures 1998;2:1-14.

10. Reneman MF, Jorritsma W, Schellekens JM, et al. Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. J Occup Rehabil 2002;12:119-29.

11. Lackner JM, Carosella AM. The relative influence of perceived pain control, anxiety, and functional self efficacy on spinal function among patients with chronic low back pain. Spine 1999;24:2254-60; discussion 60-1.

12. Gross D, Battié M. Factors influencing results of functional capacity evaluations in workers' compensation claimants with low back pain. Phys Ther 2005;85:315-22.

13. Waddell G, McCulloch JA, Kummel E, et al. Nonorganic physical signs in low-back pain. Spine 1980;5:117-25.

14. Lechner DE, Bradbury SF, Bradley LA. Detecting sincerity of effort: a summary of methods and approaches. Phys Ther 1998;78:867-88.

15. Fishbain DA, Cutler RB, Rosomoff HL, et al. Is there a relationship between nonorganic physical findings (Waddell signs) and secondary gain/malingering? Clin J Pain 2004;20:399-408.

16. Sullivan M. Exaggerated pain behavior: by what standard? Clin J Pain. 2004;20:433-9.

17. Main CJ, Waddell G. Behavioral responses to examination. A reappraisal of the interpretation of "nonorganic signs". Spine 1998;23:2367-71.

18. Kummel BM. Nonorganic signs of significance in low back pain. Spine 1996;21:1077-81.

19. Novy DM, Collins HS, Nelson DV, et al. Waddell signs: distributional properties and correlates. Arch Phys Med Rehabil 1998;79:820-2.

20. Dickens C, Jayson M, Creed F. Psychological correlates of pain behavior in patients with chronic low back pain. Psychosomatics 2002;43:42-8.

21. Baker P, Goodman G, Ekelman B, et al. The effectiveness of a comprehensive work hardening program as measured by lifting capacity, pain scales, and depression scores. Work. 2005;24:21-31.

22. Karas R, McIntosh G, Hall H, et al. The relationship between nonorganic signs and centralization of symptoms in the prediction of return to work for patients with low back pain. Phys Ther 1997;77:354-60; discussion 61-9.

23. Kool JP, Oesch PR, de Bie RA. Predictive tests for non-return to work in patients with chronic low back pain. Eur Spine J 2002;11:258-66.

24. Fordyce W. Back Pain in the Workplace: Management of Disability in Nonspecific Conditions A Report of the Task Force on Pain in the Workplace of the International Association for the Study of Pained. Seattle, WA: IASP Press, 1995.

25. Spitzer WO. Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Report of the Quebec Task Force on Spinal Disorders. Spine 1987;12:S1-S59.

26. Innes E, Straker L. Validity of work-related assessments. Work 1999;13:125-52.

27. Innes E, Straker L. Reliability of work-related assessments. Work 1999;13:107-24.

28. Innes E. Reliability and Validity of Functional Capacity Evaluations: An Update. International Journal of Disability Managment Research 2006;1:135-48.

29. Gross D, Battié M. Reliability of safe maximum lifting determinations of a functional capacity evaluation. Phys Ther 2002;82:364-71.

30. Matheson LN, Isernhagen SJ, Hart DL. Relationships among lifting ability, grip force, and return to work. Phys Ther 2002;82:249-56.

31. Gross DP, Battie MC, Cassidy JD. The prognostic value of functional capacity evaluation in patients with chronic low back pain: part 1: timely return to work. Spine (Phila Pa 1976) 2004;29:914-9.

32. Brouwer S, Reneman MF, Dijkstra PU, et al. Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in patients with chronic low back pain. Journal of Occupational Rehabilitation 2003;13:207-18.

33. Mathiowetz V, Weber K, Volland G, et al. Reliability and validity of grip and pinch strength evaluations. J Hand Surg [Am] 1984;9:222-6.

34. Mathiowetz V, Kashman N, Volland G, et al. Grip and pinch strength: normative data for adults. Arch Phys Med Rehabil 1985;66:69-74.

35. ATS. American Thoracic Society statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med 2002;166:111-7.

36. King MB, Judge JO, Whipple R, et al. Reliability and responsiveness of two physical performance measures examined in the context of a functional training intervention. Phys Ther 2000;80:8-16.

37. Kennedy DM, Stratford PW, Wessel J, et al. Assessing stability and change of four performance measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty. BMC Musculoskelet Disord 2005;6:3.

38. Soer R, van der Schans CP, Geertzen JH, et al. Normative values for a functional capacity evaluation. Arch Phys Med Rehabil 2009;90:1785-94.

39. Bohannon R. Six-Minute Walk Test A Meta-Analysis of Data From Apparently Healthy Elders. Topics in Geriatric Rehabilitation 2007;23:155-60.

40. Kendall N, Linton S, Main C. Guide to assessing psychosocial yellow flags in acute low back pain: risk factors for long-term disability and work loss. ed. Wellington NZ: Accident rehabilitation and compensation insurance corporation of New Zealand and the National Health Committee, 1997.

41. Apeldoorn AT, Bosselaar H, Blom-Luberti T, et al. The reliability of nonorganic sign-testing and the Waddell score in patients with chronic low back pain. Spine 2008;33:821-6.

42. van der Giezen AM, Bouter LM, Nijhuis FJ. Prediction of return-to-work of low back pain patients sicklisted for 3-4 months. Pain 2000;87:285-94.

43. Ferraz MB, Quaresma MR, Aquino LR, et al. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. J Rheumatol 1990;17:1022-4.

44. Storheim K, Brox JI, Holm I, et al. Predictors of return to work in patients sick listed for sub-acute low back pain: a 12-month follow-up study. Journal of rehabilitation medicine : official journal of the UEMS European Board of Physical and Rehabilitation Medicine 2005;37:365-71.

45. Waddell G, Newton M, Henderson I, et al. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993;52:157-68.

46. Matheson LN, Matheson M. Spinal function sort, rating of perceived capacity. Performance Assessment and Capacity Testing PACTed. Trabuco Canyon, California: Performance Assessment and Capacity Testing PACT, 1989.

47. Oesch PR, Hilfiker R, Kool JP, et al. Perceived functional ability assessed with the spinal function sort: is it valid for European rehabilitation settings in patients with non-specific non-acute low back pain? Eur Spine J 2010;2010:21.

48. Hamilton L. Regression with Graphics: A Second Course in Applied Statisticsed. Pacific Grove, California: Brooks/Cole Publishing Company, 1991.

49. Reneman MF, Kool J, Oesch P, et al. Material handling performance of patients with chronic low back pain during functional capacity evaluation: a comparison between three countries. Disabil Rehabil 2006;28:1143-9.

50. Mayer T, Gatchel R. Functional Restoration for Spinal Disorders: The Sports Medicine Approach. Philadelphia: Lea and Febiger, 1988.

Paper III

Comparison of two methods for interpreting lifting performance
during Functional Capacity Evaluation.

Oesch P, Meyer K, Bachmann S, Hagen KB, Vollestad NK.

Comparison of two methods for interpreting lifting performance during Functional Capacity Evaluation

Peter Oesch, MSc[1,2], Kathrin Meyer, MPH[3], Stefan Bachmann, MD[1,2,4,] Kare Birger Hagen, PhD[5,6]; Nina K. Vollestad, PhD[6]

[1]Research Department, Rehabilitation Centre Valens, Switzerland; [2]Department of Rheumatology, Rehabilitation Centre Valens, Switzerland; [3]Division of Rheumatology and Institute of Physical Medicine, Zurich University Hospital; [4]Department of Geriatrics, Inselspital, Bern University Hospital, Switzerland; [5]National Resource Centre for Rehabilitation in Rheumatology, Diakonhjemmet Hospital; [6]Institute of Health and Society, Department of Health Sciences, University of Oslo Norway

Abstract

Background: Functional Capacity Evaluation (FCE) requires an effort determination by observation of effort indices for performance interpretation. 'Nonorganic-somatic-components' identified by 'nonorganic-somatic-signs' have shown to be associated with decreased functional performance. The question arises whether effort determination by observational criteria and 'nonorganic-somatic-signs testing' can be interchangeably used to interpret lifting performance.

Objectives: To assess whether 'submaximal-effort' and 'nonorganic-somatic-components' contribute independently to lifting performance and to determine their concurrent validity.

Design: Analytical cross-sectional study

Methods: 130 patients with chronic nonspecific low back pain referred for fitness-for-work evaluation were included. Physical effort determination based on observational criteria was performed during FCE of lifting from 'floor to waist', 'waist to crown', and 'horizontal'. A second tester conducted 'nonorganic-somatic-signs' testing to identify 'nonorganic-somatic-components'. Multivariate linear regression analysis was used to determine the contribution of 'nonorganic-somatic-components' and 'submaximal-effort' to lifting performance. Age and gender were covariates. Concurrent validity of 'nonorganic-somatic-components' with 'submaximal-effort' was assessed by calculating sensitivity and specificity.

Results: 'Nonorganic-somatic-components' and 'submaximal-effort' were independent contributors to lifting performance. The contribution of 'submaximal-effort' was higher than that of 'nonorganic-somatic-components', shown by a higher change of coefficients ranging between 42–58% when 'submaximal-effort' was added to the model compared to 14–17% when 'nonorganic-somatic-components' was added as well as by higher standardized beta values of –0.49 to –0.61 in the final model compared to –0.21 to –0.26. Between 53%-63% of the patients with 'nonorganic-somatic-components' were classified as showing 'submaximal-effort'.

Limitations: Assessor variability could have influenced the study results.

Conclusions: In patients with chronic nonspecific low back pain, 'nonorganic-somatic-signs' testing and determination of physical effort by observational criteria should not be interchangeably used for interpreting lifting performance during FCE.

Key words: Low back pain – Functional Capacity Evaluation – 'nonorganic-somatic-components' – 'effort' – 'lifting performance'

Introduction

Effort determination has been a widely used attempt to validate the findings of physical performance tests [1-6]. Results from physical performance tests that are biased by 'submaximal-effort' may lead to false classifications of disability and, consequently, incorrect care as well as unwarranted disability compensation [2,7].

Various research has been performed within the fields of effort determination during muscle testing. Two literature reviews found a total of 61 studies investigating a wide variety of methods used to determine 'submaximal-effort' [7,8]. Robinson et al. [7] concluded that, despite some promising aspects of methods examining motion variability, radial/ulnar force output ratios, difference scores of eccentric-concentric ratios, and electromyography, there is not sufficient empirical evidence to support the clinical application of muscle testing for this purpose.

Physical effort determination is also attempted during Functional Capacity Evaluation (FCE) [2-4,9,10] in order to interpret the performance results. FCE is a standardized battery of clinical tests that purport to measure a patient's safe physical ability for work-related activity. Maximum effort of the client is required to obtain valid results in these physical performance tests [9]. The Isernhagen FCE uses observational criteria for physical effort level determination during manual handling tests to judge the weight load as 'light – to moderate', 'heavy' or 'maximal' [11]. 'Submaximal-effort' is assumed if a patient stops the manual handling test before the criteria indicative of a maximum weight are observed [9]. While an earlier review found only limited scientific evidence for the reliability and validity of FCEs [12,13], there is now an increasing body of knowledge on the clinimetric properties of the Isernhagen FCE providing the most comprehensive coverage of all aspects of reliability and validity of all FCE methods [14].

Another frequently used method to validate the findings of a physical examination is the assessment for 'nonorganic-somatic-components'. Waddell et al. 1980 described eight 'nonorganic-somatic-signs' that are distinguishable from the standard clinical signs of physical pathology in patients with low back pain. Multiple signs suggest that a patient does not have a straightforward physical problem, but that illness behavior and psychological factors also need to be considered [15]. 'Submaximal-effort' [3,7,8] and 'nonorganic-somatic-components' [16,17] have both been shown to be associated with decreased functional performance. However, the eight 'nonorganic-somatic-signs' were not intended to determine physical effort. In a reappraisal of the 'nonorganic-somatic-signs' the authors emphasize that these should be understood as responses to examination affected by fear in the context of recovery from injury and the development of chronic incapacity, and should be interpreted with reference to other psychological and behavioral information [18]. Despite this reasoning, 'nonorganic-somatic-signs' testing has been used as a mean for effort determination [1,4,10].

To our knowledge, it has not been investigated whether 'nonorganic-somatic-components' identified by 'nonorganic-somatic-signs' testing and 'submaximal-effort' determined by observational criteria during FCE contribute independently to lifting performance. It is also not known whether patients identified by 'nonorganic-somatic-signs' testing as presenting

'nonorganic-somatic-components' will be classified by observational criteria as showing 'submaximal-effort'. The objectives of this study are, therefore, to assess the contribution of 'nonorganic-somatic-components' and 'submaximal-effort' to lifting performance and to determine the concurrent validity of the 'nonorganic somatic components' with 'submaximal-effort' during FCE in patients with chronic nonspecific low back pain.

Materials and Methods

Study design, setting and participants

A cross sectional study was performed in three rehabilitation clinics in Switzerland. All clinics are national competence centres for FCEs, each performing more then 100 FCEs per year. For the purpose of this study, evaluations were performed by two independent assessors. These were blinded to each others results. The first assessor administered the questionnaires and assessed the independent variable 'nonorganic-somatic-signs'. The second assessor undertook three FCE tests, which consisted of lifting from 'floor to waist', 'waist to crown', and 'horizontal', and determined the effort level by applying the observational criteria . Eleven physiotherapists with a mean professional experience of 12.1 years (SD 6.9) acted as first assessors while 17 physiotherapists performed FCE. All FCE assessors completed a two-day course on FCE and had extensive experience in FCE testing, having performed on average 36.5 FCEs (SD 10.0) in the 2 years prior to this study. In addition, for the purpose of this study a trainer of FCE discussed eight FCE video sequences with the FCE assessors relating to observational criteria for effort level during manual handling tasks. The first assessors were also trained in the assessment of the eight 'nonorganic-somatic signs' as described by Waddell et al.

Participants between 20 and 60 years of age, referred for physical fitness for work evaluation, presenting with chronic nonspecific low back pain as their primary complaint, and understanding enough German, French or Italian to follow the instructions during FCE were included. Participants with specific LBP due to nerve root compression, vertebral fracture, tumor, infection, inflammatory diseases, spondylolisthesis, spinal stenosis and definite instability [19,20] or with relevant comorbidity such as cardio-respiratory, psychiatric or other musculoskeletal problems affecting work ability were excluded.

Dependent variables

Functional Capacity Evaluation

The lifting tests were conducted using a kinesiophysical approach as proposed by Isernhagen [9]. Within this approach, lifting performance is determined using observational criteria indicative of physical effort at 'light – to moderate', 'heavy' or 'maximal' levels of demand. Patients were instructed that the objective was to perform five lifts with as much weight as safely possible. For safety reasons all tests were commenced with a light weight. Weight was then increased by the FCE therapist and five lifts were repeated until the maximum safe weight was reached. The maximum safe weight load was determined by the FCE assessor when the

following criteria were observed: muscle bulging of prime movers and accessory muscles, very wide base, marked counterbalance, substantial increase in heart rate and respiration, safe weight handling but inability to maintain control if any more weight was to be added, and slowest pace [9]. Heart rate was measured with a Polar watch. If the patient stopped the manual handling test before the criteria indicative of a maximum weight were observed, the highest weight that a patient was willing to handle five times was recorded. The reliability of these observational criteria to judge lifting performance has been established in several studies based on video observation [21,22] or on direct observation by one [23-25] or more observers [26]. In the study of Gross et al. [26] five experienced FCE assessors tested patients simultaneously but independently. Their agreement in the application of the Isernhagen observational criteria [9] to determine the maximal safe lifting performance during kinesiophysical FCE was excellent shown by ICC values ranging from 0.95 to 0.98 [26]. One study using the observational criteria during video recordings of lifting from 'floor to waist' in patients with chronic low back pain and healthy subjects serving as a control group concluded that the effort level can be determined validly by means of visual observation [3].

Lifting from 'floor to waist'
Participants had to lift the weight receptacle from a shelf at waist height, make a 90 degree turn, lower the weight to the floor, stand up again lifting the weight, and then place it back on the shelf.

Lifting from 'waist to crown'
Participants had to lift the weight receptacle from a shelf at waist height to a second shelf at shoulder height, and place it back on the shelf at waist height.

Lifting 'horizontal'
Participants had to lift the weight receptacle from a shelf at waist height, walk 1.5m with the weight receptacle, place it on a second shelf at waist height, and walk back with the weight receptacle to the first shelf.

Independent variables

'Nonorganic-somatic-components'
'Nonorganic-somatic-components' were assessed with the eight 'nonorganic-somatic-signs' according to Waddell et al [15]. These are grouped into five categories. A category is positive if at least one 'nonorganic-somatic-sign' in that category is positive. Waddell et al. stress that isolated nonorganic signs should not be over interpreted. At least three positive categories are required to indicate that a patient with LBP presents 'nonorganic somatic components [15]. Patients presenting less than three positive categories are not viewed as presenting relevant 'nonorganic-somatic-components'. Waddell et al [15] report good intratester and intertester reliability of the 'nonorganic-somatic-signs'. A recent study by trained observers found good intratester reliability and moderate intertester reliability [27]. There are inconsistent conclusions regarding the validity of the 'nonorganic-somatic-components' in assessing illness behavior

and psychological factors. The most recent review [28] investigating the construct validity of the Waddell score found satisfactory cross-sectional construct validity. This contrasts, to some extent, with the review of Fishbain et al. who conclude that the 'nonorganic-somatic-signs' are not associated with psychological distress and abnormal illness behaviour but are a reflector of elevated pain and diminished functional physical capacities [16].

'Submaximal-effort'

Patients were classified by the FCE assessors as giving maximal effort if, based on their overall impression, maximum safe weight load according to the observational criteria was reached. 'Submaximal-effort' was assumed if a patient stopped the manual-handling test before these criteria were observed.

Physical factors

A subject's age and sex were shown to be performance predictors in manual-handling tests [29]. We hypothesized that decreased performance due to age or gender could be a confounder in effort evaluation during FCE.

Data analyses

The contribution of 'nonorganic-somatic-components' and 'submaximal-effort' to lifting performance was assessed with multivariate linear regression analysis. The dependent variables were the maximal weights handled during lifting from 'floor to waist', 'waist to crown', and 'horizontal'. The explanatory variables were 'submaximal-effort' and 'nonorganic-somatic-components'. We included age and gender in the model. We then added the explanatory variables stepwise and observed the change in coefficients if one variable was added over the other. Interaction of 'nonorganic-somatic-components' with 'submaximal-effort' was tested with product terms. Examination of the full models revealed that assumptions of the normality as well as homoscedasticity were respected. Taking previous research into account, [17] we assumed R2 to be at least 0.30 giving an effect size of 0.43. With alpha defined as 0.05 and beta as 0.9 and 4 independent variables, we calculated a minimum sample size of 41 subjects needed for multiple regression analysis.

We assessed the concurrent validity of the 'nonorganic-somatic-components' with 'submaximal-effort' by calculating sensitivity and specificity. Sensitivity within this study refers to the proportion of patients who presented 'nonorganic-somatic-components' and were classified as giving 'submaximal-effort'. Specificity was calculated as the proportion of the patients not presenting 'nonorganic-somatic-components' and classified as giving maximal effort.

Statistical analysis was performed with SPSS for Windows version 18.

Ethical approval

Ethical approval was obtained from the three regional ethics committees (EKSG 08/029/2B; SPUK N°. 784, EKAG 08/058) where the rehabilitation clinics are located.

Results

Descriptives

678 FCEs were performed in three rehabilitation clinics from March 2009 until August 2010. Treating physicians identified 203 subjects with low back pain, of which 60 did not fulfil inclusion criteria, 13 were missed for inclusion leaving 130 subjects with chronic nonspecific low back pain (see figure 1).

```
┌──────────────────────────┐
│ Referred for FCE, n=678  │
└──────────────────────────┘
            │         ┌────────────────────────────────────┐
            │────────▶│ LBP not primary complaint, n=475   │
            ▼         └────────────────────────────────────┘
┌──────────────────────────┐
│ Patients with LBP, n=203 │
└──────────────────────────┘
            │         ┌────────────────────────────────────────────────┐
            │         │ Inclusion criteria not fulfilled               │
            │         │  ▪ Specific back pain, n=26                    │
            │────────▶│  ▪ Comorbidity affecting work ability, n=16    │
            │         │  ▪ Language problems, n=5                      │
            │         │  ▪ No informed consent, n=11                   │
            │         │  ▪ > 60 years of age, n=2                      │
            │         └────────────────────────────────────────────────┘
            │         ┌────────────────────────────────────────────────┐
            │         │ Exclusion for other reasons                    │
            │────────▶│  ▪ Missed for inclusion, n=13                  │
            │         └────────────────────────────────────────────────┘
            ▼
┌──────────────────────────┐
│ Included in study, n=130 │
└──────────────────────────┘
```
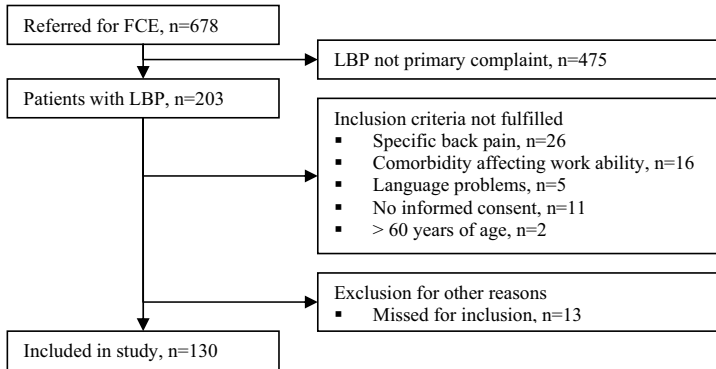
Figure 1: Study flow

93 men and 37 women with a mean age of 44.4 years (SD 10.3) were included. Mean time off work was 670 days (SD 1031) being representative of a patient sample referred for fitness for work evaluation in Switzerland. Mean lifting capacity from 'floor to waist' was 16.8kg (SD 10.6), from 'waist to crown' 13.2kg (SD 7.0), and 'horizontal' 20.0kg (SD 12.0). This was comparable to earlier published samples of patients with chronic low back pain from Switzerland [30] and from Canada [30,31]. Among the patients classified as giving submaximal effort, the change of heart rate during lifting from 'floor to waist', 'waist to crown', and 'horizontal' was 21.2 (SD12.8), 13.5 (SD8.3), resp. 8.5 (SD7.3). This was significantly lower ($p<0.001$) than the change of heart rate among the patients classified as giving maximal effort 38.2 (SD17.6), 25.9 (SD12.1), 24.9 (SD11.7).

Apart from age during lifting from 'floor to waist', all variables showed significant independent contribution to lifting performance in the final multivariate models (Table 1) explaining between 48% and 64% of the variance. No significant interactions between 'nonorganic-somatic-components' and 'submaximal-effort' was found (p= 0.11, 0.16, resp. 0.3).

Patients classified as giving 'submaximal-effort' lifted on average 10.4kg less from 'floor to waist', 8.2kg less from 'waist to crown' and 14.9kg less 'horizontal' than patients diagnosed as giving maximal effort. The influence of 'nonorganic-somatic-components' on lifting performance was substantially smaller, accounting for 5.9kg less weight lifted from 'floor to waist', 3.2kg less from 'waist to crown' and 5.3kg less 'horizontal' (see Table 1).

	Adj R2	Final model	Unstd. Coeff.	95.% CI	Sig.	Std Coeff
Lifting from 'floor to waist'	.48	Submaximal effort	-10.4	(-13.3, -7.5)	<.001	-.49
		Gender (male)	8.2	(5.1, 11.3)	<.001	.34
		'Nonorganic-somatic-components'	-5.9	(-9.0, -2.7)	<.001	-.26
		Age	-.1	(-.2, 0.0)	.103	-.11
Lifting from 'waist to crown'	.60	Submaximal effort	-8.2	(-10.0, -6.3)	<.001	-.57
		Gender (male)	6.4	(4.6, 8.2)	<.001	.40
		'Nonorganic-somatic-components'	-3.2	(-5.1, -1.2)	.002	-.21
		Age	-.1	(-.2, -.0)	.005	-.16
Lifting 'horizontal'	.64	Submaximal effort	-14.9	(-17.9, -12.0)	<.001	-.61
		Gender (male)	10.7	(7.8, 13.6)	<.001	.39
		'Nonorganic-somatic-components'	-5.3	(-8.4, -2.2)	.001	-.21
		Age	-.2	(-.3, -.1)	.007	-.15

Table 1: Final multivariate linear regression models for lifting performance
Unstd. Coeff. = Unstandardized Regression Coefficient, 95%CI = 95% Confidence Interval for the unstandardized coefficient. Std. Coeff. = Standardized Coefficients

'Submaximal-effort' had the highest influence on lifting performance shown by a high change in 'nonorganic-somatic-components' coefficients ranging from 42–58% if added to the model. The change in 'submaximal-effort' coefficients ranging from 14–17% was considerably smaller if 'nonorganic-somatic-components' was added to the model. The higher independent contribution of 'submaximal-effort' is also shown in the final models by the highest values of the standardized beta coefficients ranging between -0.49 to –0.61.

The first assessors classified 33% of the patients as presenting with 'nonorganic-somatic-components'. 53%-63% of these patients were classified by the FCE therapists as making 'submaximal-effort' during the three manual handling tests. Specificity for classification of maximal effort ranged from 84%-85% (see Table 2).

		Lifting from 'floor to waist'		Lifting from 'waist to crown'		Lifting 'horizontal'	
'Nonorganic-somatic-components' (n)		Maximal effort (n)	Sub-maximal effort (n)	Maximal effort (n)	Sub-maximal effort (n)	Maximal effort (n)	Sub-maximal effort (n)
Negative	87	59	28	69	18	68	19
Positive	43	11	32	12	31	12	31
Total (n)	130	70	60	81	49	80	50
Sensitivity		53%		63%		62%	
Specificity		84%		85%		85%	

Table 2: Sensitivity and specificity to assess the concurrent validity of 'nonorganic-somatic-components' with 'submaximal-effort'

Discussion

This study found independent contribution of 'nonorganic-somatic-components' and of 'submaximal-effort' for lifting performance without significant interactions in patients with chronic nonspecific low back pain undergoing FCE to determine fitness-for-work. The contribution of 'submaximal-effort' to lifting performance was higher than that of 'nonorganic-somatic-components'. This was shown by markedly higher changes in coefficients if 'submaximal-effort' was added to the model' compared to the lower changes if 'nonorganic-somatic-components' was added, as well as by higher standardized beta values of -0.49 to – 0.61 in the final models compared to – 0.21 to –0.26. Concurrent validity of the 'nonorganic-somatic-components' with 'submaximal effort' was low as shown by the low sensitivity. Thus, the two measures appear to reflect different aspects of physical performance during FCE of lifting.

A weakness of this study is that the FCE assessors were potentially influenced in their determination of effort by the observed low performance, thus explaining the high contribution of 'submaximal-effort' to the FCE result. The possibility that the observational criteria were not consequently applied cannot be ruled out. However, we assessed as an external objective measure of effort the change of heart rate during lifting from 'floor to waist', 'waist to crown', and 'horizontal'. In all three lifting tests, we found significantly lower changes in heart rate among the patients classified as giving submaximal effort than among the patients giving maximal effort. Heart rate was an observational criterion. Therefore, these findings can be interpreted as an indication that the FCE assessors adhered to the observational criteria. The addition of further external measures such as a control group as used in other studies investigating the validity of effort level determination during FCE [2-4] was beyond the scope of this cross-sectional study. A further weakness of this study is that it was conducted in a single country with a highly established social insurance system thereby reducing the generalizability of the study findings to other countries. Major strengths of this study are the blinding of the two assessors towards each other's findings, as well as the multicenter setting involving a large number of assessors.

Our study does not allow the identification of the different aspects of physical performance measured with 'nonorganic-somatic-signs' testing and observational criteria for physical effort level. The study's findings are in accordance with the different theoretical backgrounds of these two methods. 'Nonorganic-somatic-signs' are thought to assess illness behavior and psychological factors [15,18,28]. Effort is assumed to reflect a person's ability to perform at his or her maximal level [1]. Measurement of effort can, thus, be achieved through assessment of performance indices [2-4,9].

Despite the results suggesting that the two measures appear to reflect different aspects of physical performance during FCE, the overlap between them should not be neglected. 53%-63% of the patients classified as presenting 'nonorganic-somatic-components' were also diagnosed as giving submaximal effort. This overlap might be an explanation for the association found in previous studies of 'nonorganic-somatic-components' with decreased

functional performance [16,17]. We can only hypothesize on possible explanations for this overlap. 'Nonorganic somatic components' should be viewed as a behavioral response to examination which could be driven by fear of further injury [18]. It is theoretically plausible that fear of injury is also an underlying cause for submaximal performance during FCE. Since both measures were taken on the same day during fitness-for-work evaluation, another possible explanation is that personal expectations arising from this evaluation may have led to overt illness behavior. Overreaction, such as disproportionate verbalisation, facial expressions, muscle tension and tremors, is one criterion for 'nonorganic-somatic-components' [15]. Overreaction can also be observed in FCE during handling of low weights. Such behavior hinders the occurrence of body reactions indicative of maximal physical efforts and will, consequently, lead to 'submaximal-effort' classification.

A major point that can be taken from this study is that 'nonorganic-somatic-signs' testing should not be used to assess 'physical effort'. This is illustrated by the low concurrent validity and the markedly lower contribution of 'nonorganic-somatic-components' to FCE performance. Without the direct observation of the effort level during the lifting tests, the estimated influence of submaximal effort based on 'nonorganic-somatic-signs' testing would have been substantially smaller and may have lead to incorrect classification of disability and, consequently, inadequate decisions regarding vocational rehabilitation and physical fitness for work. Based on the study's results and on the different constructs that were to be measured with these two methods, we do not recommend the use of 'nonorganic-somatic-signs' testing for effort evaluation. The use of the previously described observational criteria for this purpose seems appropriate.

However, we must emphasize that despite promising results for the validity of the observational criteria applied during FCE [3] and the results found in this study, further research on 'physical effort evaluation' is needed. Serious medico-legal consequences might result for a patient if he or she is labelled as exerting 'submaximal-effort' during a physical fitness-for-work evaluation. Leemstra et al.[4] were only able to demonstrate the validity of five commonly used maximal effort tests to individually differentiate between maximal and 'submaximal-effort' during FCE. This is in line with Robinson et al. [7] who concluded in their literature review on the use of muscle testing for effort determination that other explanatory variables such as fear of injury, pain, medications, work satisfaction, and other motivational factors should also be considered when effort is determined.

References

1. Lechner DE, Bradbury SF, Bradley LA. Detecting sincerity of effort: a summary of methods and approaches. Phys Ther 1998;78:867-88.

2. Jay MA, Lamb JM, Watson RL, et al. Sensitivity and specificity of the indicators of sincere effort of the EPIC lift capacity test on a previously injured population. Spine 2000;25:1405-12.

3. Reneman MF, Fokkens AS, Dijkstra PU, et al. Testing lifting capacity: validity of determining effort level by means of observation. Spine (Phila Pa 1976) 2005;30:E40-6.

4. Lemstra M, Olszynski WP, Enright W. The sensitivity and specificity of functional capacity evaluations in determining maximal effort: a randomized trial. Spine 2004;29:953-9.

5. Liebenson C, Yeomans S. Quantification of Physical Performance Ability. In Liebenson C ed. Rehabilitation of the Spine: a practitioner's manual 2ed. Philadelphia: Lippincott Williams & Wilkins, 2007:226-59.

6. Reneman M, Wittink H. Functional Performance Testing. In Nordin M, Pope M, Andersson G eds. Musculoskeletal disorders in the workplace: principles and practice. Philadelphia: Mosby, 2007:397-408.

7. Robinson ME, Dannecker EA. Critical issues in the use of muscle testing for the determination of sincerity of effort. Clin J Pain 2004;20:392-8.

8. Fishbain DA, Cutler R, Rosomoff HL, et al. Chronic pain disability exaggeration/malingering and submaximal effort research. Clin J Pain 1999;15:244-74.

9. Isernhagen SJ. Functional Capacity Evaluation: Rationale, Procedure, Utility of the Kinesiophysical Approach. Journal of Occupational Rehabilitation 1992;2:157-68.

10. Kaplan GM, Wurtele SK, Gillis D. Maximal effort during functional capacity evaluations: an examination of psychological factors. Arch Phys Med Rehabil 1996;77:161-4.

11. SAR. Functional Capacity Evaluation - Course Manualed. Bellikon, Switzerland: Verein IG Ergonomie SAR, 2007.

12. Innes E, Straker L. Validity of work-related assessments. Work 1999;13:125-52.

13. Innes E, Straker L. Reliability of work-related assessments. Work 1999;13:107-24.

14. Innes E. Reliability and Validity of Functional Capacity Evaluations: An Update. International Journal of Disability Managment Research 2006;1:135-48.

15. Waddell G, McCulloch JA, Kummel E, et al. Nonorganic physical signs in low-back pain. Spine 1980;5:117-25.

16. Fishbain DA, Cole B, Cutler RB, et al. A structured evidence-based review on the meaning of nonorganic physical signs: Waddell signs. Pain Med 2003;4:141-81.

17. Oesch P, Meyer K, Janssen B, et al. What is the Role of 'Nonorganic-Somatic-Components' in Functional Capacity Evaluations in Patients with Chronic Non-Specific Low Back Pain Undergoing Fitness for Work Evaluation? Spine (Phila Pa 1976) 2011.

18. Main CJ, Waddell G. Behavioral responses to examination. A reappraisal of the interpretation of "nonorganic signs". Spine 1998;23:2367-71.

19. Fordyce W. Back Pain in the Workplace: Management of Disability in Nonspecific Conditions A Report of the Task Force on Pain in the Workplace of the International Association for the Study of Pained. Seattle, WA: IASP Press, 1995.

20. Spitzer WO. Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Report of the Quebec Task Force on Spinal Disorders. Spine 1987;12:S1-S59.

21. Isernhagen SJ, Hart DL, Matheson LM. Reliability of independent observer judgments of level of lift effort in a kinesiophysical Functional Capacity Evaluation. Work 1999;12:145-50.

22. Reneman MF, Jaegers SM, Westmaas M, et al. The reliability of determining effort level of lifting and carrying in a functional capacity evaluation. Work 2002;18:23-7.

23. Reneman MF, Dijkstra PU, Westmaas M, et al. Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. Journal of Occupational Rehabilitation 2002;12:269-75.

24. Brouwer S, Reneman MF, Dijkstra PU, et al. Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in patients with chronic low back pain. Journal of Occupational Rehabilitation 2003;13:207-18.

25. Reneman MF, Brouwer S, Meinema A, et al. Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in healthy adults. J Occup Rehabil 2004;14:295-305.

26. Gross D, Battié M. Reliability of safe maximum lifting determinations of a functional capacity evaluation. Phys Ther 2002;82:364-71.

27. Apeldoorn AT, Bosselaar H, Blom-Luberti T, et al. The reliability of nonorganic sign-testing and the Waddell score in patients with chronic low back pain. Spine 2008;33:821-6.

28. Apeldoorn AT, Ostelo RW, Fritz JM, et al. The Cross-sectional Construct Validity of the Waddell Score. Clin J Pain 2011.

29. Soer R, van der Schans CP, Geertzen JH, et al. Normative values for a functional capacity evaluation. Arch Phys Med Rehabil 2009;90:1785-94.

30. Reneman MF, Kool J, Oesch P, et al. Material handling performance of patients with chronic low back pain during functional capacity evaluation: a comparison between three countries. Disabil Rehabil 2006;28:1143-9.

31. Gross D, Battié M. Factors influencing results of functional capacity evaluations in workers' compensation claimants with low back pain. Phys Ther 2005;85:315-22.

Paper VI

Effectiveness of exercise on work disability in patients with non-acute nonspecific low back pain: Systematic review and meta-analysis of randomised controlled trials.

Oesch P, Kool J, Hagen KB, Bachmann S.

REVIEW ARTICLE

EFFECTIVENESS OF EXERCISE ON WORK DISABILITY IN PATIENTS WITH NON-ACUTE NON-SPECIFIC LOW BACK PAIN: SYSTEMATIC REVIEW AND META-ANALYSIS OF RANDOMIZED CONTROLLED TRIALS

Peter Oesch, MSc, PT[1,2], Jan Kool, PhD, PT[3], Kare Birger Hagen, PhD, PT[4] and
Stefan Bachmann, MD[1,2,5]

From the [1]Research Department, [2]Department of Rheumatology, Rehabilitation Centre Valens, Valens, [3]Zurich University of Applied Science, School of Physiotherapy, Zurich, Switzerland, [4]University of Oslo, Institute of Health Sciences and Nursing, Oslo, Norway and [5] Department of Geriatrics, Inselspital, Bern University, and University of Bern, Switzerland

Objectives: To determine whether exercise is more effective than usual care to reduce work disability in patients with non-acute non-specific low back pain, and if so, to explore which type of exercise is most effective.

Methods: Systematic review and meta-analysis of randomized controlled trials investigating the effectiveness of exercise in non-acute non-specific low back pain, and reporting on work disability. Data sources: MEDLINE, EMBASE, PEDro, Cochrane Library databases, NIOSHTIC-2, and PsycINFO until August 2008. Work disability data were converted to odds ratios. Random effects meta-analyses were conducted.

Results: A total of 23 trials met the inclusion criteria, 20 of which were suitable for inclusion in meta-analysis allowing 17 comparisons of exercise interventions with usual care and 11 comparisons of 2 different exercise interventions. A statistically significant effect in favour of exercise on work disability was found in the long term (odds ratio (OR) = 0.66, 95% confidence interval (CI) 0.48–0.92) but not in the short (OR = 0.80, 95% CI 0.51–1.25) and intermediate term (OR = 0.78, 95% CI 0.45–1.34). Meta-regression indicated no significant effect of specific exercise characteristics.

Conclusion: Exercise interventions have a significant effect on work disability in patients with non-acute non-specific low back pain in the long term. No conclusions can be made regarding exercise types.

Key words: low back pain; exercise; meta-analysis; vocational rehabilitation; sick leave.

J Rehabil Med 2010; 42: 193–205

Correspondence address: Peter Oesch, Research Department, Rehabilitation Centre Valens, CH-7317 Valens, Switzerland. E-mail: p.oesch@klinik-valens.ch

Submitted April 7, 2009; accepted December 8, 2009

INTRODUCTION

Low back pain is the most prevalent of musculoskeletal conditions. It affects almost everyone during their lifetime and has become a major socioeconomic problem in western countries (1). Exercise is consistently recommended in modern treatment guidelines for non-acute non-specific low back pain (NSLBP)

defining return to work as the primary treatment goal (2, 3). Exercises applied in the treatment of patients with NSLBP encompass a wide variety of interventions and are applied with different rationales. The sports medicine approach applies exercise based on the principles of exercise physiology, and is used in functional restoration programs with the goal of restoring physical function and thereby enabling patients to return to work (4). Behavioural treatment programmes use exercise with the aim of modifying pain behaviour. Patients learn that it is safe to move, while restoring function by receiving continuous feedback and positive reinforcement (5).

Until the year 2000 no evidence was found for the effectiveness of specific exercises in the management of chronic low back pain. Abenhaim et al. (2) state: "it appears that the key to success is physical activity itself, i.e. activity of any form, rather than any specific activity". A later review revealed that individually designed stretching or strengthening exercises delivered with supervision may improve pain and function in chronic NSLBP. The authors recommended further testing with this multivariable model and further assessment with specific patient-level characteristics and exercise types (6). A systematic review of trials with positive outcomes on work disability revealed that all had significant cognitive behavioural components combined with intensive physical training. The authors, however, advised caution when interpreting this *post-hoc* analysis and recommended further investigation into the contribution of these exercise characteristics (7). Whereas additional reviews found limited evidence for the effectiveness of behavioural graded activity in improving absenteeism outcomes (8, 9), strong evidence has been found that exercise reduces work disability in patients with NSLBP (7, 10, 11). These reviews were based on studies published prior to 2004 that did not evaluate the effectiveness of different exercise characteristics.

Although new studies have been published in the meantime, the effect of specific exercise characteristics on work disability is still unclear; a more up-to-date review is required. The objective of this review is to use recent research results to determine whether exercise is more effective than usual care to reduce work disability in patients with non-acute NSLBP, and if so, to explore which type of exercise is most effective.

METHODS

Searching

The search strategy was based on the recommendations of the Cochrane Back Review Group (12). We searched MEDLINE, EMBASE, PEDro, the Cochrane Library, PsycINFO (2002–August 2008) and NIOSHTIC-2 (until August 2008). This search was combined with a previous search performed in December 2002 covering MEDLINE (1966–October 2002), EMBASE (1988–October 2002), PEDro (until December 2002), the Cochrane Library (2002, Issue 4) and PsycLIT (1984–December 2002) (10). References were checked for further trials.

Selection, validity assessment, and data abstraction

Studies were included if randomized controlled trials were performed, the primary diagnosis in all patients was non-acute NSLBP with a duration of at least 4 weeks, the experimental treatments used exercise alone or as a part of a multidisciplinary treatment, work disability was the primary outcome, and if at least 90% of the patients under treatment were available for the job market, in that they were either employed or unemployed but seeking work. Excluded were studies that did not report work disability, investigated the effect of treatments that did not contain any form of exercise such as respondent psychological interventions, included patients with thoracic pain, cervical pain or specific low back pain, such as nerve root compression, vertebral fracture, tumour, infection, inflammatory diseases, spondylolisthesis, spinal stenosis and definite instability, and studies that included pregnant women with low back pain.

Two authors (PO and JK) independently applied the admission criteria for the studies and assessed risk of bias. Disagreements were solved through discussion involving a third researcher (StB). Authors were contacted if the information regarding the eligibility of a trial, quality criteria, or work disability were unclear.

Study quality was assessed according to Juni et al. (13), who stated that the internal validity of a study was threatened by detection bias, attrition bias, selection bias, and performance bias. Thus, the following 3 criteria were rated as "met", "unclear" or "not met": Concealed allocation, blinding of the outcome assessor, and intention to treat analysis. Performance bias was not assessed as it is not strictly possible to blind the treatment provider and recipient in clinical trials investigating the effect of exercise to treatment allocation. The internal validity of the included studies were then evaluated on methodological overall assessment. Studies were classified as high-quality studies if 2 or 3 of the criteria were met, while studies were classified as of low quality if one or none of the criteria were met.

For each study, 2 of the authors (PO and StB) independently extracted data from all included studies and defined exercise characteristics. Four criteria designed by Hayden et al. (6) were used, namely programme design, delivery type, dose, and type. Additionally, 2 criteria proposed by Schonstein et al. (14) were used, namely work context and exercise administration within a cognitive behavioural approach. A further criterion was the setting in which exercise was applied (see Table I).

Quantitative data synthesis

Work-related outcomes were converted into odds ratios (OR) using the method described by Chinn (15) and Hasselblad & Hedges (16). This method is based on the fact that, when assuming logistic distributions and equal variances in the 2 treatment groups, the log OR corresponds to a constant multiplied by the standardized difference between means. The ''metan'' command for Stata statistical software (Stata Corporation, College Station, TX V10) was used to conduct DerSimonian and Laird random effects meta-analyses (17). To use all available means, we estimated missing standard deviations (SD) from other included studies. We assessed treatment effects at 3 different times of follow-up (short-term = closest to 4 weeks, intermediate-term = closest to 6 months, long-term = closest to 12 months). Between-trial heterogeneity was quantified using the I^2 statistic, which can be understood as the proportion of the total variation in estimated ORs that is due to between-trial heterogeneity rather than chance (18). The extent

Table I. *Exercise intervention characteristics*

Programme design (according to Hayden et al. (6))
- "Individually designed", in which the treating therapist completed a clinical history and physical examination and delivered an exercise programme specifically designed for the individual participant.
- "Standard design", in which a fixed exercise programme was delivered to all participants.

Delivery type (according to Hayden et al. (6))
- Home exercises: participants performed their exercises at home with no direct supervision by the therapist.
- Supervised exercises: participants performed their exercises either under 1-on-1 supervision or attended exercise therapy sessions with 2 or more participants.
- Exercise therapy programmes that included both types of delivery will be classified according to their main delivery type.

Exercise dose (hours of intervention time) (according to Hayden et al. (6))
- We will calculate the exercise dose from the exercise duration and the number of treatment sessions received. Home exercise dose will only be included in exercise dose calculation if the home training was controlled (i.e. by using an exercise diary or by follow-up visits). If the study adherence information is not reported, we will assume an adherence rate of 50%. We will then dichotomize the exercise interventions into high- and low-dose exercise.

Types of exercises (according to Hayden et al. (6))
- Strengthening exercises
- Stretching exercises
- Mobilizing or flexibility exercises
- Aerobic exercises
- Stabilization exercises
- Programmes that included different exercise types will be classified as mixed exercises.

Setting
- Inpatient
- Outpatient

Work context (proposed by Schonstein et al. (14))
- Exercises will be classified as work-related if these were specifically designed to restore work-related physical capacity.

Behavioural treatment approach (proposed by Schonstein et al. (14))
- Exercises will be considered as administered within a behavioural treatment approach if this was specifically stated or if at least 3 of the following behavioural treatment modalities were applied: positive reinforcement of healthy behaviours (i.e. reassurance that it is safe to move, encouragement for early return to work); goal-contingent instead of pain-contingent exercise administration (i.e. exercise intensity was progressively increased to pre-set goals despite pain provocation); patients were given self-responsibility for treatment; patient education about a multidimensional view of pain (i.e. explanation of pain mechanisms); pain-coping strategies were applied (i.e. relaxation techniques were a consistent feature of the exercise programme).

to which one or more study characteristics explained between-trial heterogeneity was explored using meta-regression. The following explanatory variables were considered according to an *a priori* statistical analysis plan: exercise design (individual vs standard care), dose (high- vs low-dose exercise), delivery type (home-based exercises vs supervised exercises), type (specific vs mixed), administration within a cognitive behavioural approach (yes/no), work context (yes/no), and setting (in- vs outpatient) in bivariate models. In addition, we assessed the effect of methodological quality (low vs high). For work disability we included the variables above in meta-regression models and conducted random effects meta-analyses within each subgroup. Differences between small and large trials were assessed using funnel plots (19).

Table II. *Study characteristics*

Study	Participants	Interventions (exercise characteristic classification)	Work disability outcomes and terms of follow-up	Results
Alaranta et al., 1994 (Finland) (20)	293 patients with NSLPB > 6 months, age 30–47 years	A: Akseli Programme: 3 weeks, 37 h/week of guided or self-controlled physical exercise, 5 h/week cognitive-behavioural disability management groups, no passive treatments: (ID, SE, HD, MxE, IP, WCy, Bby) B: Current national type of inpatient rehabilitation: 3 weeks, 15–20 h/week physical activity, large amount of passive treatments: (ID, SE, HD, MxE, IP, WCn, BTn)	Number of subjects with more than 30 sick days at one year FU	No difference regarding sick leave (A: 26%, B: 23%) after one year
Aure et al., 2003 (Norway) (21)	49 patients sick-listed 8–24 weeks due to NSLPB, age 20–60 years	A: Manual therapy and specific exercise: 16 session lasting 45 min including spinal manipulation, specific mobilization, and stretching techniques as well as 5 general and 6 specific localized exercises for spinal segments and pelvic girdle, individual home exercises: (ID, SE, LD, SaE, OP, WCn, BTn) B: Supervised exercise therapy: 16 sessions consisting of 10 min warm-up on an exercise bicycle and 35 min. exercise session for the trunk, and lower limb muscles, suited to the clinical findings, individual home exercises: (ID, SE, LD, MxE, OP, WCn, BTn)	Proportion of partly or fully sick-listed patients at 4 weeks, 6 and 12 months FU	Significant advantage at 4 weeks FU for A vs B (30% vs 57%), at 6 months (11% vs 62%), and at 12 months (19% vs 59%).
Bendix et al., 1995, 2000, 1998 (Denmark) (22–24)	132 patients with disabling NSLPB > 6 months, threatened job situation or out of work	A: Functional restoration: full-time intensive 3-week multidisciplinary programme for 39 h per week, including active physical and ergonomic training and psychological pain management, followed by one 6-h day weekly for the subsequent 3 weeks: (ID, SE, HD, MxE, OP, WCy, BTy) B: Active physical training: Two sessions a week for 6 weeks, consisting of 45 min aerobics class, 45 min progressive resistance training, and 1 h of theoretical back lessons every second session: (SD, SE, LD, MxE, OP, WCn, BTn) C: Psychological pain management: 2 sessions lasting 75 min, twice a week for 6 weeks, combining active physical training of 15 min warm-up exercises, and 30 min progressive resistance training with psychological pain management: (SD, SE, LD, MxE, OP, WCn, BTy)	Sick days (median, IQR) and work capability as judged by a physician at 4 and 13 months FU	Sick days: significant advantage after 4 months for A and B vs C. A: 25 (0–103), B: 13 (0–122), C: 122 (60–122), and after 13 months for A vs B and C and for B vs C. A: 52 (0–127), B: 100 (0–390), C: 295 (0–390). Significantly better work capability in A vs B and C after 4 months (A: 75%, B: 48%, C: 40%) and 2 years (A: 80%, B: 55%, C: 44%)
Bendix et al., 1998, 1996 (Denmark) (24, 25)	106 patients with > 6 months of disabling back pain, age 18–59 years	A: Functional restoration: full-time intensive 3-week multidisciplinary programme for 39 h per week, including active physical and ergonomic training and psychological pain management, followed by 1 6-hour day weekly for the subsequent 3 weeks: (ID, SE, HD, MxE, OP, WCy, BTy) B: Usual care	Sick days (median, IQR) and work capability as judged by a physician at 4 months and 2 years FU	Sick days: significant difference at 4 months FU. A: 10 (0–122), B: 122 (24.5–122). Work capability: significantly larger improvement at 4 months. A: from 27% to 64%, B: from 16% to 29%. No difference after 2 years. A: 52%, B: 51%
Bendix et al., 2000 (Denmark) (26)	138 patients, precarious work situation due to NSLPB	A: Functional restoration: full-time intensive 3-week multidisciplinary programme for 39 h per week, including active physical and ergonomic training and psychological pain management, followed by one 6-hour day weekly for the subsequent 2 weeks, and a third day after 2ⁿᵈ month: (ID, SE, HD, MxE, OP, WCy, BTy) B: Outpatient intensive physical training: aerobics and strengthening exercises for 1.5 h 3 times/week for 8 weeks: (SD, SE, HD, MxE, OP, WCn, BTn)	Sick days (median, IQR) and work capability as judged by a physician at one year FU. No statistical comparison between groups. Results disregard 32 of 138 randomized patients who did not finish treatment	Sick days: No difference at 1 year FU. A: median in 34/48 patients 5.5 (0–113), B: median in 40/50 patients 2.5 (0–301). Work capability: pre-post A: from 28 to 36/48, B: from 21 to 35/51
Hagen et al., 2000, 2003 (Norway) (27, 28)	457 patients with NSLPB and sick leave 8–12 weeks, age 18–60 years	A: Light mobilization programme: examination at spine clinic for 1 hour, information, advice to stay active and go on daily walks; individual instructions lasting 1–1.5 h on stretching and training at home by the physical therapist: (ID, HE, LD, ScE, OP, WCn, BTy) B: Usual care by GP	Days of sickness compensation (mean, 95% CI), numbers of workers returning to full-duty work at 3, 6, 12, 24, and 36 months FU	Significant fewer sick days at one year FU in A vs B (A: 95.5, 104.5, B: 133.7, 112); RTW: significant improvement after 6 months in A vs B (A: 61.2 %, B: 45.0 %) and 1 year (A: 68.4 %, B: 56.4 %). No significant difference at 3-year FU (A: 36.2%, B: 38.2%)

Study	Population	Intervention	Outcome measure	Results
Härkäpää et al, 1989, 1990 (Finland) (29, 30)	476 blue-collar workers, age 35–54 years, sick leave due to NSLPB during the last 2 years	A: Inpatient rehabilitation: groups of 6–8 patients, 3 weeks, 4 sessions Swedish back school, 15 back exercise sessions, 9 relaxation sessions, heat or electrotherapy prior to exercise, 2 structured group discussions, home programme, rehearsal after 1.5 years (2 weeks): (SE, MxE, IP, WCn, BTn) B: Outpatient treatment at the workplace or local health centre: 15 sessions in 2 months, groups of 6–8 patients, 4 sessions Swedish back school, 15 back exercise sessions, 9 relaxation sessions, heat or electrotherapy prior to exercise, 2 structured group discussions, home programme, rehearsal after 1.5 years (8 sessions): (SE, MxE, OP, WCn, BTn) C: Physician examination: no systematic treatment, written and oral instructions on back exercises and ergonomics.	Sick days at 15 months FU disregarding the first 7 days of each episode of sickness absence leading to a considerable underestimation of the days lost from work	Sick days after 1.5 years: A vs C no difference. (A: 5.5, 25; C: 7.5, 25). B vs C no difference (B: 5.8, 25, C: 7.5, 25.0). A vs B no difference (A: 5.5, 25.0, B: 5.8, 25.0)
Heymans et al, 2006 (Netherlands) (31)	299 workers sick-listed 3 to 6 weeks because of NSLPB, age 18–65 years; able to complete Dutch questionnaires	A: High-intensity back school: 2 sessions of 1 hour per week during 8 weeks, supervised by a physiotherapist. Instruction of individual strength training exercises, work simulation and home exercises during the treatment period. A cognitive-behavioural therapy approach was used: (ID, SE, LD, SgE, OP, WCy, BTy) B: Low-intensity back school: 4 group sessions based on the Swedish model, once a week, 4 consecutive weeks. Sessions were divided into 30 min education and a 90-minute practical part, comprised of a standardized exercise programme consisting of strength training and home exercises: (SD, SE, LD, SgE, OP, WCn, BTn) C: Usual care: provided by occupational physician according to the Dutch guidelines for the occupational health management of patients with LBP.	Cumulative percentages of RTW during 6 month FU. Sick days (mean, SD personal communication) at 6 month FU	Faster RTW was observed in B compared with C and A with hazard ratios of 1.4 (p=0.06) and 1.3 (p=0.09) respectively. No statistically significant difference regarding sick days between the 3 groups at 6 month FU (A: 96.1, 63.3; B: 81.8, 55.6; C: 92.5, 65.5)
Hurri, 1989 (Finland) (32)	80 female employees with NSLPB > 12 months	A: Back school given by a physical therapist; 6 sessions of 1 hour, consisting of education and exercise, in groups of 11 patients. Two review sessions of 1 hour after 6 months: (SD, SE, LD, MxE, OP, WCn, BTn) B: Usual care by GP and provision of instruction material from back school.	Sick days due to LBP at 12 months FU (mean, SD personal communication)	No difference during the 1st year (A: 8.1, 26.9; B: 11.1, 26.6) and during the 2nd year (A: 9.0, 23.6; B: 9.5, 25.0)
Jousset et al., 2004 (France) (33)	86 workers, threatened job situation by LBP, age 18–50 years, no limited work contract	A: Functional restoration programme: 6 h per day, 5 days/week during 5 weeks, muscular warm-up and stretching techniques, aerobic exercises, progressively increasing strengthening exercises, work simulation, Balneology for 30 min per day. Psychological counselling if required and dietary advice: (ID, SE, HD, MxE, WCy, BTn) B: Active individual therapy: Three 1-hour sessions per week during 5 weeks, supervised by private practice physiotherapists, instruction of a 50 min active home exercise programme twice a week, and recommendation of sports activities: (ID, SE, HD, MxE, OP, WCn, BTn)	Sick days (mean, SD) at 6 months FU and RTW (%) within one week	Sick days: advantage of A vs B (A. 28.7, 44.6; B: 48.3, 66.0). RTW no difference between the 2 groups (A 69.2%, B 66.7%)
Kääpä et al., 2006 (Finland) (34)	120 women with daily or nearly daily LBP, age 22–57 years	A: Outpatient multidisciplinary rehabilitation: programme of 70 h duration conducted in groups of 6 to 8 patients during 8 weeks. The intervention comprised cognitive-behavioural treatment, back school education including occupational intervention, and a physical exercise programme for 2–3 h per day: (ID, SE, HD, MxE, OP, WCn, BTy) B: Individual physiotherapy: Ten 1-hour treatment sessions during 6–8 weeks. 30–40 min passive pain treatment and 15–20 min light active exercise per session, advice to progressively increase regular daily activities, recommendation of general physical training, instruction of a light home-exercise programme: (ID, SE, LD, MxE, OP, WCn, BTn)	Percentage of patients in different categories of sick leave at all terms of FU (0 days, 1–30 days, > 30 days) at 6, 12 and 24 months FU	No statistically significant differences of A vs B at all terms of FU (1 year: 0 days = A: 61.5%, B: 56.4%; 1–30 days = A: 32.7%, B: 34.6%; > 30 days = A: 5.8%, B: 9.0%; 2 years: 0 days = A: 68%; B: 63.8%; 1–30 days = A: 20%; B: 25.6%; >30 days = A: 12%; B: 10.6%.

Study	Population	Intervention	Outcome measure	Results
Karjalainen et al., 2003, 2004 (Finland) (35, 36)	164 patients, age 25–60, working difficulties between 4 and 12 weeks	A: Mini-intervention: 60 min physician examination, information on the good prognosis of back pain, and disadvantages of rest, encouragement to stay active followed by a 90-minute physiotherapy session; appraising of daily back-straining activities and training of special movements required at work, and instruction of a daily exercise programme: (ID, HE, LD, MxE, OP, WCy, BTy) B: Work site visit: mini intervention plus work site visit C: Usual care by GP	Sick days (mean, SD estimated as 2/3 of mean) and percentage of patients on sick leave at 12 and 24 months FU	Significantly fewer sick days at 1 year FU in A vs C (A: 19, 12; C 41, 27) and at 2 year FU. Significantly fewer patients on sick leave at 1 year FU (A: 45%; C: 61%)
Kool et al., 2005, 2007 (Switzerland) (37, 38)	174 patients with non-acute LBP, out of work > 6 weeks, age 20–55 years	A: Function centred treatment: Inpatient rehabilitation of 3 weeks duration, 4 h per day, focusing on improved function not pain reduction. Work simulation, strength and endurance training, sports therapy, and self-exercises: (ID, SE, HD, MxE, IP, WCy, BTy) B: Pain-centred treatment: inpatient rehabilitation of 3 weeks, 2.5 h per day, focusing on pain reduction. Individually selected passive and active exercises, mini back school, passive pain modulating treatments, and relaxation techniques: (ID, SE, HD, MxE, IP, WCn, BTn)	Sick days (mean, SD personal communication) at 12 months FU	Advantage after 12 months for A vs B (A: 247, 137; B: 291, 115)
Lindström et al., 1992, 1995, 1992 (Sweden) (5, 39, 40)	103 blue-collar workers, sick-listed 6 weeks, no NSLPB sick-listing in the prior 12 weeks	A: Graded activity programme: treatment until RTW achieved, no predefined treatment duration. Functional capacity testing, workplace visit, back school, individual sub-maximal gradually increasing exercise programme to teach the patient that it is safe to move while restoring function: (ID, SE, HD, MxE, OP, WCn, BTy) B: Usual care by GP	Number of patients RTW at 6 and 12 weeks FU, and sick days (mean, SD) during 2nd year FU	Significantly larger proportion of patients RTW in A vs B after 6 weeks (A: 59%, B: 40%) and 12 weeks (A: 80%; B: 58%). Sick days: significant advantage during the 2nd follow-up year of A vs B (A: 60, 92; B: 98, 103.5)
Niemisto et al 2003, 2005 (Finland) (41, 42)	204 patients with chronic LBP, age 24–64 years, Oswestry Disability Index score > 16%	A: Combined treatment: physician's consultation incl. explanation of clinical findings, instructions regarding posture and exercise, and an educational booklet combined with manipulative treatment, and instruction of pain-free stabilizing exercises 4 times in the course of 4 week: (ID, SE, LD, SaE, OP, WCn, BTn) B: Physician's consultation: 1 hour at entry and 5 months FU	Sick days (mean, SD) at 1 and 2 years FU	No significant differences of A vs B at 1 year FU (A: 13.9, 26.6; B: 18.5, 38.8) and 2 years FU (A: 12.3, 20.5; B: 14.8, 38.0)
Petersen et al., 2002, 2007 (Denmark) (43, 44)	260 people with LBP > 8 weeks, age 18–60 years	A: McKenzie treatment: examination by physiotherapist 1 hour, treatment sessions 30 min (maximum 15 sessions) consisting of self-mobilization with repeated movements, mobilization by physiotherapist. Instruction to continue self-treatment sessions for 2 months: (ID, SE, LD, MoE, OP, WCn, BTn) B: Strengthening training: sessions lasting 60–90 min, twice a week for 8 weeks (maximum 15 sessions). Intensive training of trunk flexors and extensors in groups of 6 patients after warm-up exercises. Instruction to continue self-treatment sessions for 2 months: (SD, SE, LD, SgE, OP, WCn, BTn)	Percentage of patients on sick leave after 2, 8 and 14 months FU	No difference after 2 months (A: 10%; B: 14% and 8 months (A: 7%; B: 8%) and 14 months
Roche et al., 2007 (France) (45)	132 patients with chronic LBP on sick leave or at risk of work disability, age 18–50 years	A: Functional restoration programme: 6 h per day, 5 days/week during 5 weeks, muscular warm-up and stretching techniques, aerobic exercises, progressively increasing strengthening exercises, and work simulation. Balneology for 30 min per day. Psychological counselling if required and dietary advice: (ID, SE, HD, MxE, nc, Wcy, BTn) B: Active individual therapy for 1 hour, 3 times a week during 5 weeks and individual exercises to be performed at home for 50 min twice a week: (ID, SE, HD, MxE, OP, WCn, BTn)	RTW and self-assessed ability to work at end of treatment	No difference regarding RTW (A: 85.7%; B: 86.8%). Significant advantage in self-assessed ability to work of A vs B (A: 78.1%; B: 95.5%).

Study	Population	Intervention	Outcome measure	Results
Skouen et al., 2002 (Norway) (46)	195 employees with NSLPB sick-listed > 8 weeks	A: Extensive multidisciplinary programme: 6 h/day, 5 days/week, 4 weeks. Cognitive behavioural modification in group sessions, education, exercise, occasional workplace interventions: (ID, SE, HD, MxE, OP, WCn, BTy). B: Light multidisciplinary programme: information about exercise, fear avoidance. Instruction of a personal exercise programme based on physical tests. FU visits at 3 and 6 months. Further physiotherapy, appointment with psychologist and workplace visits if necessary: (ID, HE, LD, MxE, OP, WCn, BTy) C: Usual care by GP	Sick days (mean, SD) and percentage of patients with full RTW at 12 and 28 months FU	Sign. differences in men for RTW in B vs C but not for women. Overall results: sick days during the first year: A: 167.3; 228.1; B: 149, 228.1; C: 203.8, 223.6); RTW: A: 60%, B: 69%, C: 49%.
Staal et al., 2004, 2005, 2007 (Netherlands) (47–49)	134 sick-listed workers with NSLPB > 4 weeks in succession	A: Graded activity programme: physical examination followed by 1 hour exercise sessions twice per week (graded activity exercises regardless of the amount of pain, maximum therapy duration of 3 months: (ID, SE, LD, MxE, OP, WCy, BTy) B: Usual care provided by OP and GP according to the low back guidelines of the Dutch College of General Practitioners	Sick days (mean, SD personal communication) until full return to regular work and subjects sick-listed at 6, 12 and 36 months FU	No significant difference of sick days (mean, SD) at 12mth FU (A: 86.4/90.6; B: 128.5/122.7) and of subjects sick-listed (A: 5; B: 8).
Steenstra et al., 2006 (Netherlands) (50, 51)	112 patients with NSLPB included in a previous trial (44) but remaining sick-listed for > 8 weeks, age 18–65	A: Graded activity programme: individual, submaximal, gradually increasing exercise programme based on the demands from the patients' work with an operant-conditioning behavioural approach consisting of two 1-hour sessions a week, 26 sessions maximally: (ID, SE, HD, MxE, OP, WCy, BTy) B: Usual care provided by OP according to the Dutch OP guidelines for LBP. Patients GP were informed to minimize co-interventions	Persons at work at 26 weeks FU and sick days (mean, SD separated according to previous intervention WI/UC) until full return to regular work at 52 weeks FU	A prolonged RTW at 26 weeks FU (A: 59%; B: 76%), more sick days at 52 weeks (WI: A 160.8, 78.7; B: 109.9, 62.6; UC: A: 172.75, 85.8; B: 151.4, 105.1).
Storheim et al., 2003 (Norway) (52)	93 patients with sub-acute NSLPB, sick-listed for 8–12 weeks	A: Cognitive intervention: 2 consultations, each lasting between 30 and 60 min providing functional examination with individual feedback and advice, explanation of pain mechanisms including reassurance that it is safe to move, and instruction in trunk stabilizing behaviour and squat lifting technique B: Intensive group training: based on the Norwegian Aerobic Fitness Model but modified for patients with LBP lasting 1 hour for 15 weeks, 2–3 times per week. No pain focus: (SD, SE, HD, MxE, OP, WCn, BTy) C: Usual care by the GP	Sick days (mean, SD estimated as 2/3 of mean) at 18 weeks FU	No difference between B and C (B: 63.5, 44; C: 56.3, 37).
Torstensen et al., 1998 (Norway) (53)	137 patients with NSLPB, sick-listed 4–8 weeks, employment, age 20–65 years, birth in Norway	A: Medical exercise therapy: 36 treatments of 1 hour duration for 12 weeks, groups of 5 patients, 6–9 exercises using specially designed exercise equipment with approximately 1000 repetitions: (ID, SE, HD, MxE, OP, WCn, BTn) B: Conventional physiotherapy: 36 treatments of 1 hour duration for 12 weeks, heat or cold, massage, traction, electrotherapy, few exercises on the treatment table: (ID, SE, LD, MxE, OP, WCn, BTn) C: Physician examination and information on the importance of activity and walking 3 times a week	Total number of sick days in each group at 15 months FU	No significant difference regarding sick days between the 3 groups after 15 months (A: 11757, B: 9967, C: 13567).
White, 1966 (Canada) (54)	194 men sick-listed 6–52 weeks for NSLPB, age 19–60 years	A: Inpatient rehabilitation: up to 6 weeks, 7 h per day, progressive treatment in 4 stages (bed rest–light–medium–heavy physical exercise) to the level demanded by the requirements of the man's job: (ID, SE, HD, MxE, IP, WCy, BTn) B: Usual care by GP	Numbers of workers with satisfactory RTW at 3 months FU	Significantly better at 3 months (A: 42%, B: 16%).

ID: individually designed; SD: standard design; HE: home exercises; SE: supervised exercise; SgE: strengthening exercises; ScE: stretching exercises; SaE: stabilization exercises; MoE: mobilization exercises; HD: high-dose exercise; LD: low-dose exercise; MxE: mixed exercises; IP: inpatient; OP: outpatient; WCy/n: work context yes/no; BTy/n: behavioural treatment yes/no; FU: follow-up; LBP: low back pain; RTW: return to work.

RESULTS

Trial flow

Of the 838 articles retrieved from the literature search, we evaluated 87 articles in detail, of which 64 did not meet the inclusion criteria. Consequently, we included 23 studies in this review. Sixteen were trials with 2 study arms and 7 were trials with 3 study arms. Table II summarizes the characteristics of the included studies. Twenty studies were included in the meta-analysis, allowing 17 comparisons of exercise interventions with usual care (5, 27, 29, 31, 32, 35, 41, 46, 47, 50, 52–54) and 11 comparisons of 2 different exercise interventions (20, 21, 29, 31, 33, 34, 37, 43, 45, 46, 53). Three studies (22, 25, 26) were excluded from meta-analysis as "days of sick leave" were presented as median and interquartile range, thereby preventing pooling (Fig. 1).

Trials comparing 2 different exercise interventions with usual care were treated as 2 trials with the sample size of the usual care group equally divided between the 2 exercise intervention groups: inpatient rehabilitation and outpatient treatment (29); low and high intensity back school (31); light and extensive multidisciplinary programme (46); conventional physiotherapy and medical exercise therapy (53). One study (50) presented results of 2 patient groups defined by the previous intervention (UC: usual care; WI: workplace intervention) receiving the same exercise intervention.

Validity assessment

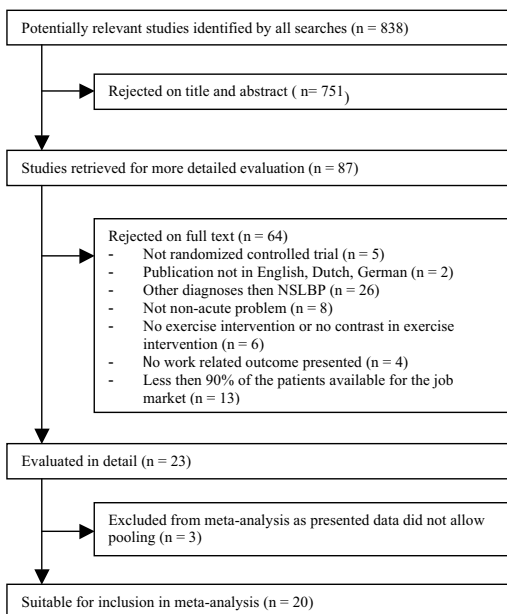According to the previously mentioned criteria, 14 (61%) of the studies were found to be of high quality and 9 (39%) of low quality (Table III). The 3 studies excluded from meta-analysis were all of low quality.

Work disability data

Data on work disability varied between the different studies and included self-assessed work ability, days of sick leave, days at work, physician's judgement of work capability, and days of sickness compensation or numbers of workers returning to full-duty work. These were obtained from insurance databases whereby national legal requirements may have influenced the recordings. The data used for pooling were the number of people who returned and did not returned to work at the time of the follow-up, or the total number of sick days within the follow-up period (Table II).

Exercise characteristics

Thirty-five different exercise interventions were used. Exercise design, dose and setting were reported unclear in 6% of the investigated exercise interventions. Twenty-six (74%) of the exercise interventions were individually designed; 32 (91%) were primarily performed as supervised exercise; 28 (80%) interventions used mixed exercise types, 2 stabilization, 3 strengthening, 1 mobilization, and 1 stretching exercise; 27 (77%) were conducted in an outpatient setting; 10 (29%) were work-related; and 14 (40%) of the exercise interventions were administered within a cognitive behavioural approach (Table II).

Unfortunately, none of the studies using home exercise reported adherence rate or sufficient information to estimate home exercise dose. Therefore, the calculation of the exercise dose is based on the number of the supervised treatment ses-

Fig. 1. Trial flow diagram to summarize the stages of the systematic review.

Table III. *Risk of bias and study quality*

Study, reference	Selection bias	Detection bias	Attrition bias	Study quality
Alaranta et al., 1994 (20)	Yes	No	Yes	Low
Aure et al., 2003 (21)	No	Yes	No	High
Bendix et al.,1995 (22–24)	Yes	Yes	Yes	Low
Bendix et al.,1996 (24, 25)	Yes	Yes	Yes	Low
Bendix et al., 2000 (26)	Yes	Yes	Yes	Low
Hagen et al., 2000 (27, 28)	No	No	No	High
Härkäpää et al., 1989 (29, 30)	Yes	No	Yes	Low
Heymans et al., 2006 (31)	No	No	No	High
Hurri, 1989 (32)	Yes	Yes	No	Low
Jousset et al., 2004 (32)	Yes	Yes	Yes	Low
Kääpa et al., 2006 (34)	No	Yes	No	High
Karjalainen et al.,2003 (35, 36)	No	Yes	No	High
Kool et al., 2005 (37, 38)	No	No	No	High
Lindström et al., 1992 (5, 39, 40)	No	No	No	High
Niemisto et al., 2003 (41, 42)	No	Yes	No	High
Petersen et al. 2002 (43, 44)	No	Yes	No	High
Roche et al., 2007 (45)	Yes	Yes	No	Low
Skouen et al., 2002 (46)	No	No	Yes	High
Staal et al., 2004 (47–49)	No	No	No	High
Steenstra et al., 2006 (50, 51)	No	No	No	High
Storheim et al., 2003 (52)	No	No	No	High
Torstensen et al., 1998 (53)	No	No	No	High
White, 1966 (54)	Yes	Yes	Yes	Low

sions and their duration only. Such calculated exercise dose varied widely between the different exercise interventions, ranging between 1.5 and 210 h. The median exercise dose was 17 h. We classified exercise interventions with ≥17 h of contact time as high-dose exercise ($n = 18$), and those with less than 17 h of contact time into low-dose exercises ($n = 17$). A cut-off point of 14 and 20 h resulted in less than a 10% change in exercise dose classification.

Qualitative comparison of exercise interventions and usual care

Seven studies reported work disability data on a short-term follow-up. Two high-quality studies (5, 27) and 2 low-quality studies (25, 54) reported a positive effect, 1 high-quality study no effect (52) and 2 high-quality studies a negative effect (31, 50). Five studies reported work disability data on an intermediate-term follow-up. Three high-quality studies (5, 27, 47) reported a positive effect. This was observed in the study by Staal et al. (47) from approximately 50 days after randomization onwards. Two high-quality studies reported a negative effect (31, 50). Long-term results were presented by 11 studies. Positive effects were found in 3 high-quality studies (5, 27, 35). No significant effects were observed in 3 low-quality studies

(25, 29, 32) and 4 high-quality studies (41, 46, 47, 53). One high-quality study reported a negative effect (50).

Quantitative data synthesis

Comparison of exercise interventions vs usual care. Thirteen studies allowing 17 comparisons between an exercise intervention and usual care with a total of 3181 patients were available for pooling.

Comparison 01: Short-term follow-up. Short-term results were available for pooling from 5 high-quality studies (6 comparisons, 1030 patients) (5, 27, 31, 50, 52), showing no significant effect of exercise reducing work disability (OR = 0.80, 95% confidence interval (CI) 0.51–1.25). The addition of one low-quality study (54) did not substantially change the overall effect estimate (OR = 0.68, 95% CI 0.42–1.10)

Comparison 02: Intermediate-term follow-up. Four high-quality studies (5 comparisons, 971 patients) (5, 27, 31, 50) provided results for pooling at the intermediate-term follow-up showing no significant effect of exercise in reducing work disability (OR = 0.78, 95% CI 0.45–1.34).
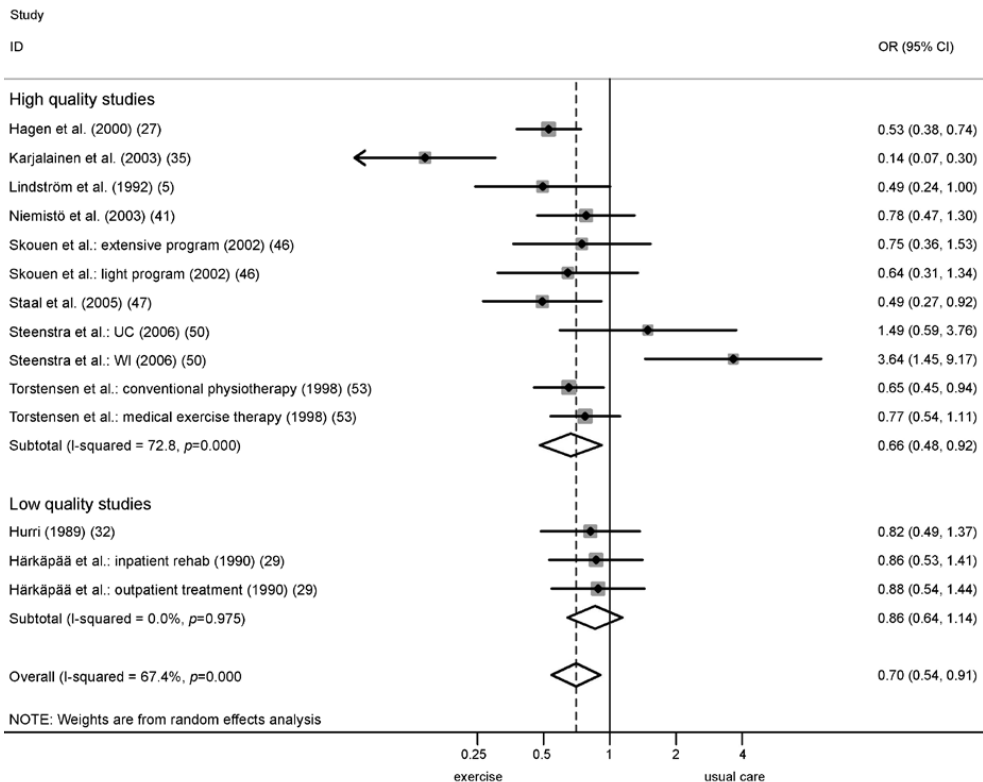

Fig. 2. Meta-analysis of 10 trials with long-term follow-up comparing exercise interventions with usual care.
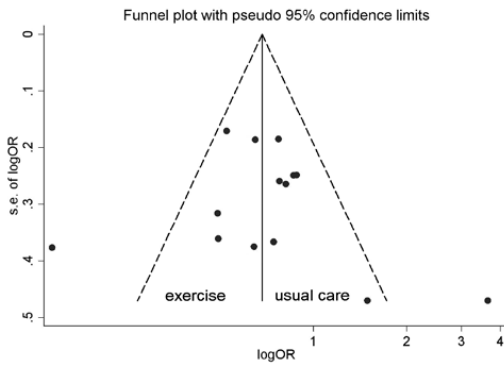
Fig. 3. Funnel plot of 10 trials with long-term follow-up comparing exercise interventions with usual care.

Comparison 03: Long-term follow-up. Eight high-quality studies (10 comparisons, 1992 patients (5, 27, 35, 41, 46, 47, 50, 53) presented long-term follow-up results showing a statistically significant overall effect in favour of exercise on work disability (OR=0.66, 95% CI 0.48–0.92). The addition of 2 low-quality studies (29, 32) did not substantially change the overall effect estimate (OR=0.70, 95% CI 0.54–0.91) (Fig. 2). The funnel plots did reveal evidence of asymmetry at short- and intermediate-term follow-ups but not at long-term follow-up (Fig. 3).

Comparison 04: Influence of exercise characteristics in high-quality trials with long-term follow-up. The 8 high-quality studies (5, 27, 35, 41, 46, 47, 50, 53) providing data on 1149 patients receiving an exercise intervention and 843 patients receiving usual care were included for this analysis. All comparisons were between different outpatient rehabilitation programmes, and all used individually designed exercises, in one comparison stretching exercises were instructed, another used stabilization exercise, and in the remaining 8 comparisons

mixed exercises were used. A second overall analysis, which did not include the patient sample from the trial of Steenstra et al. (50) that had already received a workplace intervention (WI group), showed reduced statistical heterogeneity ($I^2 = 60.4\%$, $p = 0.007$) and increased the effect estimate (OR=0.59, 95% CI 0.45–0.78). The effect of delivery type, exercise dose, work context and behavioural treatment approach was analysed with and without the WI group showing different results, although within the statistical error margin. Pooled effects for the 4 exercise characteristics hypothesized to influence work disability (delivery type, exercise dose, work context, behavioural treatment approach) became higher and more significant for supervised exercise, and a behavioural treatment approach. However, none of the variables were statistically significant in meta-regression analysis, although there was a trend observed favouring home exercises ($p = 0.11$) (Table IV).

Comparison of different exercise interventions. Of the 13 studies comparing 15 different exercise interventions, 6 were low quality (20, 22, 26, 29, 33, 45) and 7 were of high quality. Six high-quality studies presented long-term follow-up data and were used for pooling (21, 34, 37, 43 46, 53). We defined the exercise intervention with more contact hours as the standard intervention. All standard interventions used individually designed supervised exercises, 5 of them with mixed exercises, and 5 were conducted in an outpatient setting.

There was significant statistical heterogeneity in these trials (I-squared = 65.5%, $p = 0.013$). The overall effect of exercise interventions with more contact hours was not significant (OR = 1.07, 95% CI 0.67–1.72). Three trials applying exercise within a behavioural treatment approach showed some benefit (OR = 0.75, 95% CI 0.47–1.20) compared with the trials without this characteristic (OR = 1.74, 95% CI 0.71–4.30) (Fig. 4). One trial applying work-related exercise in an inpatient setting (42) showed a significant effect on work disability (OR = 0.53, 95% CI 0.30–0.93) compared with exercise not specifically designed to restore work-related physical capacity (OR = 1.25, 95% CI 0.80–1.97). None of these characteristics showed statistical significance in meta-regression analysis.

Table IV. *Odds ratios for work disability stratified by exercise characteristics in 8 high-quality randomized controlled trials comparing exercise intervention with usual care*

Exercise characteristics	Work disability (WI group included) OR (95% CI)	Work disability (WI group excluded) OR (95% CI)	Meta-regression (WI group excluded) logOR (95% CI), *p*-value
Delivery type			
Home exercises	0.38 (0.17–0.84)	0.38 (0.17–0.84)	1.74 (0.86–3.55), *p*=0.11
Supervised exercise	0.80 (0.58–1.11)	0.70 (0.58–0.85)	
Exercise dose			
Low dose (<17 hours)	0.51 (0.35–0.73)	0.51 (0.35–0.73)	1.52 (0.71–3.27), *p*=0.24
High dose (>17 hours)	1.01 (0.57–1.78)	0.76 (0.56–1.05)	
Work context			
No	0.65 (0.54–0.77)	0.65 (0.54–0.77)	0.66 (0.27–1.59), *p*=0.31
Yes	0.77 (0.21–2.85)	0.46 (0.41–1.55)	
Behavioural treatment approach			
No	0.72 (0.57–0.91)	0.72 (0.57–0.91)	0.72 (0.33–1.56), *p*=0.35
Yes	0.65 (0.39–1.10)	0.52 (0.34–0.80)	

CI: confidence interval; OR: odds ratio; WI: workplace intervention.
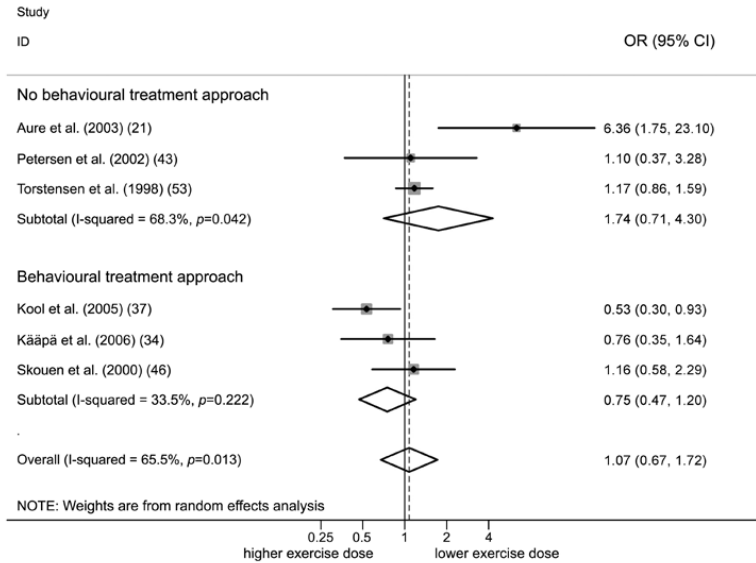
Fig. 4. Meta-analysis of 6 high-quality trials comparing lower and higher dosed exercises at long-term follow-up without and with a behavioural treatment approach.

The funnel plots that were conducted did not reveal evidence of funnel plot asymmetry.

DISCUSSION

This meta-analysis provides continuous support for the use of exercise interventions to achieve long-term benefits on work disability in patients with non-acute NSLBP. The OR of 0.66 suggests that the odds of "improvement" in work disability are in the long-term 34% lower if only usual care (rather than exercise) is given. No significant effect was observed in short- and intermediate-term follow-ups. Meta-regression showed no significant differences between different exercise types. Interestingly, home exercises seem to be at least as effective as supervised programmes. As the meta-regression is only explorative, no conclusions can be made regarding exercise types.

Our study has several strengths. The search strategy was based on the recommendations of the Cochrane Back Review Group. We planned the analysis *a priori* based on the findings of previous meta-analyses and assessed study quality based on key components of methodological quality (concealed allocation, blinded assessor, intention to treat analysis) as recommended by Juni et al. (13). Studies affected by biases have previously been shown to exaggerate treatment effects (13). We, therefore, excluded low-quality studies from meta-regression analysis in order to avoid a possible overestimation of the effect of different exercise characteristics.

A weakness of this study is the high proportion of total unexplained variance that could be attributed to study heterogeneity.

We considered this weakness thoroughly, but concluded that patients, social support and outcomes showed satisfactory homogeneity. All but one of the studies were performed in Europe, in countries with comparable social systems. All patients were diagnosed with non-acute NSLBP, were of working age and available for the job market. Despite the wide variety of used work disability outcomes, this meta-analysis is based on the pooled results of just 2 different outcome measures. We performed a stratified analysis in 3 studies providing both outcome measures and found no relevant differences in ORs, both in favour of exercise. Furthermore, using mean values and standard deviations for further statistical analysis in data with a skewed distribution is usually regarded as inappropriate. Data regarding sick days have a skewed distribution, but this was similar in both groups in treatment comparisons that reduces the risk of systematic bias (55). To address the problem of statistical heterogeneity, we performed a random effects meta-analysis. There remains the possible error of substantial variation in standard deviations across studies leading to an over- or underestimation of the ORs.

Exercise interventions did not show a significant effect on work disability at short- and intermediate-term follow-up. However, these findings are not conclusive. The mean odds ratios for short- and intermediate-term results were both below unity, but with wide CI. Therefore, a significant effect might remain undetected based on ineffectiveness, heterogeneity or limited power of the pooled studies. Possible explanations for a lack of effect at short- and intermediate-term follow-up are the required time needed to improve physical capacity, to modify pain behaviour, or to search for work. Furthermore,

the process of care has a substantial effect on work disability, as shown in a recent study comparing a graded activity programme with usual care (50). The interaction between a prior WI and graded activity, together with a delay in the start of the graded activity intervention, explained most of the delay in return to work (RTW) (50). This study introduced relevant clinical heterogeneity in this meta-analysis. All of the other trials investigated the primary treatments for this occurrence of back pain, while half of the patients in the trial from Steenstra et al. (50) had already received a WI, which has been shown to be effective on return to work (56). Herbert & Bo (57) propose that study quality can also be assessed on how interventions are administered. There were obviously problems in the implementation of the graded activity programme in the trial of Steenstra et al. (50), leading to a potentially false conclusion if the whole patient sample had been included in this meta-analysis. In view of these considerations we feel it legitimate to interpret the findings without the results of the WI group. The author's recommendation of paying special attention to the structure and process of care in implementing graded activity (50) does have clinical relevance when conducting medical interventions aiming for early RTW. An open and fast access to such interventions prevents unnecessary waiting time before a RTW can be attempted. This might also be a possible explanation for why individually designed home exercises seem to be more effective than supervised exercise interventions in reducing work disability. Home exercise may facilitate RTW, as the patients are able to continue their daily routine without spending extra time on medical intervention.

Interestingly, this meta-analysis did not show a greater effect of higher dose exercise interventions (≥ 17 contact h) compared with lower dose exercise interventions on work disability (< 17 contact h). This finding is contrary to exercise physiology postulating a dose and effect relation (58), as well as to previous findings that only intensive (> 100 h of therapy) multidisciplinary biopsychosocial rehabilitation with functional restoration improves function in patients with chronic low back pain, whereby inconclusive results were found on vocational outcomes (59). The only study included in the systematic review by Guzman et al. (59) supporting the use of functional restoration to reduce work disability (25) is not included in the performed meta-regression, as only high-quality studies were used. Moreover, in the meantime new studies with low contact hours administering home exercise have been published showing a positive effect on work disability. This might, from our point of view, be an explanation for the different findings. However, a cautionary comment must be made on the missing effect of exercise dose found in this review. As in other systematic reviews, incomplete reporting in the primary studies present important limitations and prevented the calculation of the exercise dose in home exercise programmes. It must be assumed that the actual exercise dose in home exercise interventions was higher than the calculated dose.

We have not been able to confirm the positive effects of exercises performed within a behavioural treatment approach on work disability postulated in previous reviews (8, 9, 14). In the comparison of exercises with usual care we found stronger

treatment effects for such exercises. However, this was not statistically significant in the meta-regression (Table IV). The missing confirmation might be due to the differing study inclusion criteria and the analysis performed. All previous reviews based their conclusion on a qualitative assessment, at least partly based on the evidence found by Lindström et al. (5). Hayden et al. (9) also included the results of Staal et al. (47), while Schonstein et al. 2003 (14) included the results of 3 more studies that were excluded from this analysis because of the risk of bias or missing inclusion criteria. We included in addition to the studies of Staal et al. (47) and Lindström et al. (5) the findings of 6 more studies (27, 31, 35, 47, 50, 52) that contained 9 treatment comparisons with a total of 1316 patients providing sufficient power for the meta-regression. However, it must be emphasized that the presented meta-regression analyses is only explorative and does not allow any conclusions too be drawn.

The comparison of different exercise interventions also did not reveal a significant effect of a behavioural treatment approach (Fig. 4). There might be a superior effect if exercises are performed within a behavioural treatment approach and are specifically designed to restore work-related capacity, as shown by the study of Kool et al. (38). This is also in line with Schonstein et al. (7), who hypothesize a positive effect of such a combination.

We recommend further evaluation of the combined effects of individually designed home exercises applied within a behavioural treatment approach aiming to specifically restore work-related physical capacity. Special attention must be given to an effective implementation process of exercise interventions aiming for early RTW.

ACKNOWLEDGEMENTS

REFERENCES

1. Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. Bull World Health Organ 2003; 81: 646–656.
2. Abenhaim L, Rossignol M, Valat JP, Nordin M, Avouac B, Blotman F, et al. The role of activity in the therapeutic management of back pain. Report of the International Paris Task Force on Back Pain. Spine 2000; 25 Suppl 4: 1S–33S.
3. COST B13. European guidelines for the management of acute nonspecific low back pain in primary care. Eur Spine J 2006; 15 Suppl 2: S125–S300.
4. Mayer T, Gatchel R. Functional Restoration for Spinal Disorders: The Sports Medicine Approach. Philadelphia: Lea and Febiger; 1988.
5. Lindström I, Öhlund C, Eek C, Wallin L, Peterson LE, Nachemson A. Mobility, strength, and fitness after a graded activity program for patients with subacute low back pain. A randomized prospective clinical study with a behavioral therapy approach. Spine 1992; 17: 641–652.
6. Hayden JA, van Tulder MW, Tomlinson G. Systematic review: strategies for using exercise therapy to improve outcomes in

chronic low back pain. Ann Intern Med 2005; 142: 776–785.

7. Schonstein E, Kenny DT, Keating J, Koes BW. Work conditioning, work hardening and functional restoration for workers with back and neck pain. Cochrane Database Syst Rev 2003; (1): CD001822.

8. Ostelo R, van Tulder M, Vlaeyen J, Linton S, Morley S, Assendelft W. Behavioural treatment for chronic low-back pain. Cochrane Database Syst Rev 2005: CD002014. DOI: 10.1002/14651858. CD002014.pub2(1).

9. Hayden JA, van Tulder MW, Malmivaara AV, Koes BW. Meta-analysis: exercise therapy for nonspecific low back pain. Ann Intern Med 2005; 142: 765–775.

10. Kool J, de Bie R, Oesch P, Knusel O, van den Brandt P, Bachmann S. Exercise reduces sick leave in patients with non-acute non-specific low back pain: a meta-analysis. J Rehabil Med 2004; 36: 49–62.

11. Hlobil H, Staal JB, Spoelstra M, Ariens GA, Smid T, van Mechelen W. Effectiveness of a return-to-work intervention for subacute low-back pain. Scand J Work Environ Health 2005; 31: 249–257.

12. van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine 2003; 28: 1290–1299.

13. Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. BMJ 2001; 323: 42–46.

14. Schonstein E, Kenny D, Keating J, Koes B, Herbert RD. Physical conditioning programs for workers with back and neck pain: a Cochrane systematic review. Spine 2003; 28: E391–E395.

15. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. Stat Med 2000; 19: 3127–3131.

16. Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. Psychol Bull 1995; 117: 167–178.

17. Der Simonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986; 7: 177–188.

18. Reitman RD, Emerson RH, Higgins LL, Tarbox TR. A multimodality regimen for deep venous thrombosis prophylaxis in total knee arthroplasty. J Arthroplasty 2003; 18: 161–168.

19. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. J Clin Epidemiol 2001; 54: 1046–1055.

20. Alaranta H, Rytokoski U, Rissanen A, Talo S, Ronnemaa T, Puukka P, et al. Intensive physical and psychosocial training program for patients with chronic low back pain. A controlled clinical trial. Spine 1994; 19: 1339–1349.

21. Aure OF, Nilsen JH, Vasseljen O. Manual therapy and exercise therapy in patients with chronic low back pain: a randomized, controlled trial with 1-year follow-up. Spine 2003; 28: 525–531; discussion 31–32.

22. Bendix AF, Bendix T, Ostenfeld S, Bush E, Andersen. Active treatment programs for patients with chronic low back pain: a prospective, randomized, observer-blinded study. Eur Spine J 1995; 4: 148–152.

23. Bendix AF, Bendix T, Lund C, Kirkbak S, Ostenfeld S. 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 1997; 29: 81–89.Hagen EM, Eriksen HR, Ursin H. Does early intervention with a light mobilization program reduce long-term sick leave for low back pain? Spine 2000; 25: 1973–1976.

24. Bendix AF, Bendix T, Labriola M, Boekgaard P. Functional restoration for chronic low back pain. Two-year follow-up of two randomized clinical trials. Spine 1998; 23: 717–725.

25. Bendix AF, Bendix T, Vaegter K, Lund C, Frolund L, Holm L. Multidisciplinary intensive treatment for chronic low back pain: a randomized, prospective study. Cleve Clin J Med 1996; 63: 62–69.

26. Bendix T, Bendix A, Labriola M, Haestrup C, Ebbehoj N. Functional restoration versus outpatient physical training in chronic low back pain: a randomized comparative study. Spine 2000; 25: 2494–2500.

27. Hagen EM, Eriksen HR, Ursin H. Does early intervention with a light mobilization program reduce long-term sick leave for low back pain? Spine 2000; 25: 1973–1976.

28. Hagen EM, Grasdal A, Eriksen HR. Does early intervention with a light mobilization program reduce long-term sick leave for low back pain: a 3-year follow-up study. Spine 2003; 28: 2309–2315.

29. Härkäpää K, Jarvikoski A, Mellin G, Hurri H. A controlled study on the outcome of inpatient and outpatient treatment of low back pain. Part I. Pain, disability, compliance, and reported treatment benefits three months after treatment. Scand J Rehabil Med 1989; 21: 81–89.

30. Härkäpää K, Mellin G, Jarvikoski A, Hurri H. A controlled study on the outcome of inpatient and outpatient treatment of low back pain. Part III. Long-term follow-up of pain, disability, and compliance. Scand J Rehabil Med 1990; 22: 181–188.

31. Heymans MW, de Vet HC, Bongers PM, Knol DL, Koes BW, van Mechelen W. The effectiveness of high-intensity versus low-intensity back schools in an occupational setting: a pragmatic randomized controlled trial. Spine 2006; 31: 1075–1082.

32. Hurri H. The Swedish back school in chronic low back pain. Part I. Benefits. Scand J Rehabil Med 1989; 21: 33–40.

33. Jousset N, Fanello S, Bontoux L, Dubus V, Billabert C, Vielle B, et al. Effects of functional restoration versus 3 hours per week physical therapy: a randomized controlled study. Spine 2004; 29: 487–493; discussion 494.

34. Kääpä EH, Frantsi K, Sarna S, Malmivaara A. Multidisciplinary group rehabilitation versus individual physiotherapy for chronic nonspecific low back pain: a randomized trial. Spine 2006; 31: 371–376.

35. Karjalainen K, Malmivaara A, Pohjolainen T, Hurri H, Mutanen P, Rissanen P, et al. Mini-intervention for subacute low back pain: a randomized controlled trial. Spine 2003; 28: 533–540; discussion 40–41.

36. Karjalainen K, Malmivaara A, Mutanen P, Roine R, Hurri H, Pohjolainen T. Mini-intervention for subacute low back pain: two-year follow-up and modifiers of effectiveness. Spine 2004; 29: 1069–1076.

37. Kool JP, Oesch PR, Bachmann S, Knuesel O, Dierkes JG, Russo M, et al. Increasing days at work using function-centered rehabilitation in nonacute nonspecific low back pain: a randomized controlled trial. Arch Phys Med Rehabil 2005; 86: 857–864.

38. Kool J, Bachmann S, Oesch P, Knuesel O, Ambergen T, de Bie R, et al. Function-centered rehabilitation increases work days in patients with nonacute nonspecific low back pain: 1-year results from a randomized controlled trial. Arch Phys Med Rehabil 2007; 88: 1089– 1094.

39. Lindström I, Ohlund C, Nachemson A. Physical performance, pain, pain behavior and subjective disability in patients with subacute low back pain. Scand J Rehabil Med 1995; 27: 153–160.

40. Lindström I, Öhlund C, Eek C, Wallin L, Peterson LE, Fordyce WE, et al. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an operant-conditioning behavioral approach. Phys Ther 1992; 72: 279–290.

41. Niemisto L, Lahtinen-Suopanki T, Rissanen P, Lindgren KA, Sarna S, Hurri H. A randomized trial of combined manipulation, stabilizing exercises, and physician consultation compared to physician consultation alone for chronic low back pain. Spine 2003; 28: 2185–2191.

42. Niemisto L, Rissanen P, Sarna S, Lahtinen-Suopanki T, Lindgren KA, Hurri H. Cost-effectiveness of combined manipulation, stabilizing exercises, and physician consultation compared to physician consultation alone for chronic low back pain: a prospective randomized trial with 2-year follow-up. Spine 2005; 30: 1109–1115.

43. Petersen T, Kryger P, Ekdahl C, Olsen S, Jacobsen S. The effect of McKenzie therapy as compared with that of intensive strengthening training for the treatment of patients with subacute or chronic

low back pain: a randomized controlled trial. Spine 2002; 27: 1702–1709.

44. Petersen T, Larsen K, Jacobsen S. One-year follow-up comparison of the effectiveness of McKenzie treatment and strengthening training for patients with chronic low back pain: outcome and prognostic factors. Spine 2007; 32: 2948–2956.

45. Roche G, Ponthieux A, Parot-Shinkel E, Jousset N, Bontoux L, Dubus V, et al. Comparison of a functional restoration program with active individual physical therapy for patients with chronic low back pain: a randomized controlled trial. Arch Phys Med Rehabil 2007; 88: 1229–1235.

46. Skouen JS, Grasdal AL, Haldorsen EM, Ursin H. Relative cost-effectiveness of extensive and light multidisciplinary treatment programs versus treatment as usual for patients with chronic low back pain on long-term sick leave: randomized controlled study. Spine 2002; 27: 901–909.

47. Staal JB, Hlobil H, Twisk JW, Smid T, Koke AJ, van Mechelen W. Graded activity for low back pain in occupational health care: a randomized, controlled trial. Ann Intern Med 2004; 140: 77–84.

48. Hlobil H, Staal JB, Twisk J, Köke A, Ariëns G, Smid T, et al. The effects of a graded activity intervention for low back pain in occupational health on sick leave, functional status and pain: 12-month results of a randomized controlled trial. J Occup Rehabil 2005; 15: 569–580.

49. Hlobil H, Uegaki K, Staal JB, de Bruyne MC, Smid T, van Mechelen W. Substantial sick-leave costs savings due to a graded activity intervention for workers with non-specific sub-acute low back pain. Eur Spine J 2007; 16: 919–924.

50. Steenstra IA, Anema JR, Bongers PM, de Vet HC, Knol DL, van Mechelen W. The effectiveness of graded activity for low back pain in occupational healthcare. Occup Environ Med 2006; 63: 718–725.

51. Steenstra IA, Anema JR, van Tulder MW, Bongers PM, de Vet HC, van Mechelen W. Economic evaluation of a multi-stage return to work program for workers on sick-leave due to low back pain. J Occup Rehabil 2006; 16: 557–578.

52. Storheim K, Brox JI, Holm I, Koller AK, Bo K. Intensive group training versus cognitive intervention in sub-acute low back pain: short-term results of a single-blind randomized controlled trial. J Rehabil Med 2003; 35: 132–140.

53. Torstensen TA, Ljunggren AE, Meen HD, Odland E, Mowinckel P, Geijerstam S. Efficiency and costs of medical exercise therapy, conventional physiotherapy, and self-exercise in patients with chronic low back pain. A pragmatic, randomized, single-blinded, controlled trial with 1-year follow-up. Spine 1998; 23: 2616–2624.

54. White AW. Low back pain in men receiving workmen's compensation. Can Med Assoc J 1966; 95: 50–56.

55. Boneau C. The effects of violations of assumptions underlying the t-test. Psychol Bull 1960; 57: 49–64.

56. Anema JR, Steenstra IA, Bongers PM, de Vet HC, Knol DL, Loisel P, et al. Multidisciplinary rehabilitation for subacute low back pain: graded activity or workplace intervention or both: a randomized controlled trial [with consumer summary]. Spine 2007; 32: 291–298.

57. Herbert RD, Bo K. Analysis of quality of interventions in systematic reviews. BMJ 2005; 331: 507–509.

58. Wilmore J, Costill D, Kenney W. Physiology of sport and exercise science, 4th edn. Hong Kong: Human Kinetics; 2008.

59. Guzman J, Esmail R, Karjalainen K, Malmivaara A, Irvin E, Bombardier C. Multidisciplinary rehabilitation for chronic low back pain: a systematic review. BMJ 2001; 2001; 322: 1511–1516.