

## Symptom Increase Following a Functional Capacity Evaluation in Patients with Chronic Low Back Pain: An Explorative Study of Safety

Michiel F. Reneman · Wietske Kuijer · Sandra Brouwer ·  
H. R. Schiphorst Preuper · Johan W. Groothoff ·  
Jan H. B. Geertzen · Pieter U. Dijkstra

Published online: 16 May 2006  
© Springer Science+Business Media, Inc. 2006

**Abstract** *Introduction:* This study was performed to study intensity and duration of symptom increase following an FCE and to explore safety of an FCE. *Methods:* Included were 92 patients with chronic low back pain (CLBP), mean age 38.5 years, mean self-reported disability 12.5 (Roland Morris Disability Questionnaire). All patients underwent an FCE. Symptom increase was measured with a 2-item questionnaire. Operational definition for safety: no formal complaint filed and symptom increase to occur only temporarily. *Results:* No formal complaints were filed ( $n = 92$ ). In total, 54 patients returned the questionnaire (59%; ‘responders’). Of the responders, 76% reported increased symptom intensity after an FCE, ranging from ‘little increase’ to ‘severe increase’. Symptoms of all responders returned to pre-FCE level. Duration of symptom increase of the responders ranged from 1 day to 3 weeks. Symptom increase resided to pre-FCE level within 1 week in 93% of the responders. Symptom increase was weakly related to self-reported disability ( $r = 0.38$ ,  $p < 0.05$ ). Except for gender, differences between responders and non-responders were non-significant. *Conclusion:* A temporary increase in symptom intensity following an FCE is common. Within the operational definitions of safety used in this study, assessment of functional capacity of patients with CLBP appears safe.

**Keywords** Chronic low back pain · Safety · Lifting · Injury · Occupational rehabilitation · Insurance medicine

---

M. F. Reneman (✉) · W. Kuijer · S. Brouwer · H. R. Schiphorst Preuper · J. H. B. Geertzen · P. U. Dijkstra  
Center for Rehabilitation, University Medical Center Groningen, University of Groningen,  
P.O. Box 30.002, 9750 RA, The Netherlands  
e-mail: m.reneman@cvr.umcg.nl

M. F. Reneman · J. W. Groothoff  
Center for Occupational Health, University Medical Center Groningen, University of Groningen,  
The Netherlands

M. F. Reneman · W. Kuijer · S. Brouwer · H. R. Schiphorst Preuper · J. W. Groothoff · J. H. B. Geertzen ·  
P. U. Dijkstra  
Northern Center for Health Care Research, University of Groningen, The Netherlands

P. U. Dijkstra  
Department of Oral and Maxillofacial Surgery, University Medical Center Groningen, University of  
Groningen, The Netherlands

## Introduction

Functional Capacity Evaluations (FCEs) are batteries of tests to measure the ability of a person to perform work-related activities. FCEs are used in occupational rehabilitation, return to work determinations, in disability determinations and in medico-legal issues. FCEs should have proof of the ability to assess a person's ability to perform activities safely, reliably, validly, and practically [1]. During the past years, research has been published that confirms reliability and validity of the Isernhagen WorkWell Systems (IWS) FCE<sup>1</sup> in healthy adults and patients with chronic low back pain (CLBP) [2–8]. 'Safety' has been researched only to determine whether operational definitions of safety could be applied reliably by evaluators during FCEs [8–12]. These operational definitions of safety include biomechanical, metabolic and psychophysical aspects [10,13] to ensure that patients remain within defined limits of strength and heart rate, and that patients do not perform beyond the point that they feel unsafe or unwilling to proceed. However, whether applying these operational definitions of safety leads to a safe assessment of functional capacity has not been investigated for this, or any other FCE [13].

It is a common clinical observation that patients with CLBP report a temporary increase of symptoms after an FCE. Increased symptom intensity is not per se regarded unusual or unsafe in FCE testing. This clinical observation was confirmed in a small pilot study sample ( $n = 10$  patients with chronic pain) published recently in this journal [13]. However, no other studies on this subject are published. The primary aim of this study was to study intensity and duration of symptom increase after an FCE in a sample of patients with CLBP, and to explore the relationship of increased symptoms with age, self-reported disability, pain intensity, lifting performance, heart rate, and work status.

The secondary aim of this study, linked to the increase of symptoms, was to explore the safety of the FCE. The term 'safety' is often used in FCE literature [13], but never defined. The American Physical Therapy Association's (APTA) task force on objective functional measurements, however, has defined safety. Their definition was "given the known characteristics of the evaluatee, the procedure should not be expected to lead to new injury" [14, p. 682]. This definition, however, cannot be tested because of the wording used ('...should not be expected to...') and because of the absence of a definition for 'injury' in this definition. Therefore, a new operational definition of safety was used in this study: an FCE was considered safe when no formal complaints of injury were filed by the patients, and when increased symptoms returned to or below their pre-FCE level. A temporary increase in symptom intensity was not considered unsafe. References as to what is considered 'temporary' are unavailable. Therefore, a time limit was not included in the operational definition. Special attention in this paper is given to the lifting capacity test, because lifting is one of the most stressful items of the FCE, it is considered a risk factor for low back pain [15, 16], and is deemed potentially unsafe. Additionally, lifting provides an indication of total FCE performance [17].

## Materials and methods

### Subjects

Included in this study were 92 patients with CLBP (duration of more than 3 months or non-specific recurrent low back pain), who were admitted to an outpatient occupational rehabilitation

<sup>1</sup> Isernhagen WorkWell Systems Inc., Duluth, Minnesota, USA.

program. Excluded for the program were patients with acute low back pain, specific diagnoses related to low back pain (for example disk herniations) and patients with co-morbidity (additional diagnoses unrelated to low back pain). For the FCE specifically, patients whose blood pressures and heart rate in rest exceeded 100 mmHg (diastolic), 159 mmHg (systolic) and 90 beats per minute respectively were excluded [18]. In- and exclusion for the program and FCE was performed by a physiatrist. The data of this study were collected as part of a larger study program, LOBADIS (low back pain and disability) [19].

### General procedures

Prior to the FCE, the patients filled out a questionnaire to obtain demographic information (age, gender, duration CLBP, work status). Self-reported disability due to CLBP was measured with the Roland Morris Disability Questionnaire (RDQ). Scores range from 0 to 24, with 0 indicating no self-reported disability and 24 indicating severe self-reported disability. Reliability and validity of the RDQ is good [20–22]. Pain intensity was measured with a 100 mm Visual Analogue Scale. The score can range from 0 to 100 mm, with 0 indicating no pain, and 100 mm indicating the worst pain imaginable. Patients underwent a standardized near full length FCE, consisting of 28 tests that were all performed in a 2.5 to 3.0 h session on one day. Tests included were: tests to measure material handling capacities (i.e. lifting and carrying), tests to measure static postural tolerance (i.e. kneeling and forward bending) and tests to measure dynamic activities (i.e. walking and reaching). Detailed information of the procedures is described elsewhere [2, 3]. Sitting and standing tolerance tests, each tests lasting 30 min, were not tested. Test-retest reliability of the material handling tests of the FCE is good in patients with CLBP [2, 4, 5]. Test-retest reliability of 17 of 21 additional tests of the FCE is acceptable in patients with CLBP [2]. Recovery between tests: patients were to start a next test only when his heart rate was below 70% of his predicted maximum ( $[220 - \text{age}] \times 0.70$ ). Lifting capacity was assessed by means of a standardized lifting task consisting of lowering and lifting a receptacle with incremental weights from a table (74 cm) to the floor and vice versa. The patient's maximum was reached in 4 to 5 increments. The maximal amount lifted safely 5 times within 90 s was recorded. Lifting was tested as the first item in the FCE.

### Safety procedures

Medical history and examination, performed by physiatrists prior to the FCE, ensured that inclusion and exclusion criteria were met. All FCEs were led by a licensed physical therapist certified into the procedures of the IWS FCE. During the FCE three standardized safety procedures for termination of a test were adhered to. The reason for termination of a test was documented. Two or 3 reasons could occur simultaneously.

1. Patient endpoint. All patients signed a consent statement stating that they were informed both by the evaluator and in writing that undergoing an FCE could lead to a temporary increase of symptoms and general muscle soreness and that they could terminate testing whenever they felt unsafe or unwilling to proceed. Patients were also informed that test termination based on symptoms was solely left to the patient (thus without interpretation or inquiry of the evaluator).
2. Cardiac endpoint. A heart rate ceiling of 85% of a patient's age related maximum heart rate ( $[220 - \text{age}] \times 0.85$ ) was set on a heart rate monitor. Testing was terminated when the patient's heart rate met or exceeded this criterion [18].

3. Evaluator endpoint. The evaluator could terminate testing when the patient was not in full control of himself or the load that he handled. Patients were allowed to use body mechanics of their own preference, for example a semi-squat lifting technique or a trunk lift during lifting tests [23]. The inter- and intrarater reliability of this determination is good [6, 8].

After completion of the FCE, all patients were handed two questions on paper (questions translated for this manuscript from Dutch into English). Question 1: To what extent have your symptoms increased compared to yesterday (day of the FCE)? Answering possibilities were: my symptoms have decreased; no increase; little increase; moderate increase; strong increase; very strong increase. This question was to be answered 1 day after the FCE. Question 2: How many days did it last until the symptoms returned to the original level? This question was to be answered whenever this occurred. No time limit was attached to question 2. Patients who had not responded one year post FCE were classified as 'non-responders'. No attempts were made to contact non-responders. The questionnaire could be returned by mail free of charge to a researcher independent from the evaluator.

#### Operational definitions of safety

An FCE was considered safe when no complaints of injury were filed by the patients and when increased symptoms returned to or below their pre-FCE level. As part of regular institutional admission procedures, all patients were informed on their rights and responsibilities, including the ability to file a formal written complaint if and when they had felt mistreated in any way. An institutional committee, functioning independent from clinicians and researchers, conducted operation of these rights.

#### Data analyses

Descriptive statistics were used for patient characteristics, the results of the lifting capacity test and the responses on the questionnaire. Students' *t*-tests for independent samples were used to test differences between responders and non-responders. Depending on measurement level and distribution of the data, Pearsons' correlation or Spearman's rank correlation coefficients were calculated to express the strength of the relationship between variables. Correlation coefficients were interpreted as follows: 0.25 or less little if any relationship, 0.26–0.49 weak relationship, 0.50–0.69 moderate relationship, 0.70–0.89 strong relationship, 0.90–1.00 very strong relationship [24]. Test results for males and females were analyzed separately. A *p*-value less than 0.05 was considered statistically significant for all tests.

## Results

#### Patient characteristics and test performance

Characteristics of the study sample and test performances are presented in Table 1. All patients who were on modified work or off work due to CLBP received wage replacement benefits, as regulated within the Dutch Social Security system. The characteristics of those who returned the questionnaire ( $n = 54$ , 59%; 'responders') and those who did not return the questionnaire ( $n = 38$ , 41%; 'non-responders') are presented separately. Non-responders were significantly more frequent male. Other differences between responders and non-responders were non-significant. Males performed significantly better than females on the lifting capacity test. Percent of predicted

**Table 1** Characteristics of the patients and FCE performances ( $n = 92^a$ )

	Responders ( $n = 54$ )			Non-responders ( $n = 38$ )		
	<i>n</i>	% or Mean	SD	<i>n</i>	% or Mean	SD
Male/female ( <i>n</i> /%)	31/23	57/43	–	29/9	76/24	–
Work status: normal/modified/off work ( <i>n</i> /%)	22/16/14	42/31/27	–	5/15/16	14/42/44	–
Age (years)	54	38.4	9.2	38	38.5	8.1
Duration CLBP (weeks)	54	157	218	38	142	259
Self-reported disability (scale 0–24) <sup>b</sup>	54	11.8	5.0	38	13.4	4.4
Pain intensity (scale 0–100) <sup>c</sup>	54	46	23	38	52	21
Lifting test maximum performance males (kg)	31	32.1	17.2	29	33.0	13.4
Lifting test maximum performance females (kg)	23	18.6	7.5	9	19.4	9.0
Heart rate maximum performance males (bpm)	31	126.1	18.8	28	127.8	16.4
Heart rate maximum performance females (bpm)	22	132.4	23.3	9	123.4	21.8

bpm: beats per minute.

<sup>a</sup>Incomplete datasets of 2 responders and 2 non-responders.

<sup>b</sup>Roland Morris Disability Scale.

<sup>c</sup>Visual Analogue Scale.

maximum heart rate: male responders 69%, male non-responders 70%, female responders 73%, and female non-responders 68% (calculation: mean predicted maximum heart rate (220 – mean age) divided by mean observed maximum heart rate).

The lifting capacity test was terminated in the following frequencies (2 or 3 endpoints may have occurred simultaneously): responders ( $n = 54$ ): patient endpoint occurred in 89% ( $n = 48$ ) of the cases, cardiac endpoint occurred in 15% ( $n = 8$ ) of the cases, and evaluator endpoint occurred in 13% ( $n = 7$ ) of the cases. Non-responders ( $n = 38$ ): patient endpoint occurred in 82% ( $n = 31$ ) of the cases, cardiac endpoint occurred in 11% ( $n = 4$ ) of the cases, and evaluator endpoint occurred in 16% ( $n = 6$ ) of the cases.

### Symptom increase

The complaint committee received no complaints from any of the 92 patients up to one year after the FCE was completed. Differences in symptom intensity following the FCE are presented in Table 2. Most responders reported an increase in symptoms (76%;  $n = 41$ ), however some reported no difference or a decrease in symptoms. Symptom increase, when applicable, was temporary for all responders. The question about the duration of symptom intensity was left blank or was indicated as not applicable by 20.4% of the patients ( $n = 11$ ). All of these patients had not experienced an increase in symptoms. Of the remaining 43 patients the mean duration of symptom increase to return to pre-FCE level was 3.4 days (SD 3.4 days), minimum 1 day and maximum 21 days. Increased symptom intensity lasted 1 week or less in 93% of the patients ( $n = 40$ ).

The relationships of increased symptoms with age, self-reported disability, pain intensity, lifting performance, heart rate, and work status were non-significant (Table 3). The relationships

**Table 2** The intensity of symptom increase 1 day following FCE ( $n = 54$ )

	%	<i>n</i>
Symptoms decreased	4	2
No difference	20	11
Little increase	19	10
Moderate increase	17	9
Strong increase	26	14
Very strong increase	15	8

between the intensity and duration of increased symptoms on the one hand and self-reported disability (RDQ score) on the other hand were significant but weak. Repeating these analyses with inclusion of the non-responders (duration symptoms set at 365 days), then the duration of symptom increase associates non-significantly with all variables except work status (Spearman's  $\rho = 0.26, p = 0.03$ ).

## Discussion

No complaint was filed. Most patients reported symptom increase following an FCE. In all respondents, the symptoms returned to a pre-FCE level. An FCE was considered safe when the patients filed no formal complaints of injury, and when increased symptoms returned to or below their pre-FCE. Applying the results of this study to our operational definition of safety, it is interpreted that the FCE is a safe instrument to measure the functional capacity of patients with CLBP.

Although difficult to compare due to differences in methodology, the symptom increase in this study appears to be larger in comparison to the results of a small sample ( $n = 10$ ) described previously [13]. No criteria are available as to what is considered a 'temporary' increase, or how to interpret intensity and duration of symptom increase. To determine whether intensity and duration of increased symptoms are acceptable remains a clinical decision at this point. In our operational definition, a temporary increase of symptoms is not considered unsafe, and the results of this study apply to this definition. The results may not apply to other (implicit) definitions of injury or safety. However, we are unaware of definitions of safety and injury that is consented upon.

**Table 3** Correlation coefficients between intensity and duration of increased symptoms and age, self-reported disability, pain intensity, lifting performance, heart rate and work status ( $n = 53$ )

	Intensity increase <sup>a</sup>	Age <sup>b</sup>	Self-reported disability <sup>b</sup>	Pain intensity	Lifting performance <sup>b</sup>	Heart rate <sup>b</sup>	Work status <sup>a</sup>
Intensity increase		-0.19	0.38*	0.23	-0.22	0.08	0.06
Duration increase	0.50*	0.23	0.37*	0.30	-0.18	-0.23	0.08

<sup>a</sup>Pearson correlation.

<sup>b</sup>Spearman rank correlation.

\* $p < 0.05$ .

 Springer

It may be debated whether the operational definition of safety used in this study was correct. With regard to safety during FCEs or lifting assessments, the statements of other researchers are similar. Safety (or unsafety) is referred to as ‘prevention of further injury’ [25], ‘significant complications or injury’ [26] or ‘report of injuries’ [18]. Our operational definition of safety resembles those. The next step was to assess whether the criteria for safety were met. Even though no complaints were filed, injuries may have occurred but left unreported. The (perceived) weight of the procedure and/or the (perceived) threshold to file a complaint, and a (perceived) power imbalance between patient and clinician may have prevented patients to report injuries, regardless of the patient’s legal rights to do so. Alternatively, increased pain intensity or the occurrence of an injury may not have been interpreted as mistreatment. Both may have occurred in this study. We judge this as unlikely to have happened in the group of responders, because their symptoms had resided to pre-FCE level. Because all patients were asked to return the questionnaire when symptoms had resided to the original level, the non-responders may still have had increased symptoms more than a year after completion of the FCE. In practice we regard this as very unlikely, because none of the non-responders filed a complaint, even though their condition would theoretically be worse for at least one year. Additionally, a non-responder bias was controlled for. Besides a difference in gender, which cannot be explained, no systematic differences were found between the characteristics and test performances of the responders and the non-responders. However, non-responders may be different from responders on some unknown (and unmeasured) factor that is more closely associated with symptom change following FCE.

As part of a different study, patients were asked to undergo another FCE after completion of their rehabilitation program. Some patients refused to undergo this post-treatment FCE. Clinical information, which has not been gathered systematically, suggests that their reason for refusal was their experience of symptom increase after the first FCE. Other researchers have reported findings of a few patients refusing to undergo a second test session because they did not feel capable of participation in manual handling activities due to pain exacerbation [5]. In a study where patients with CLBP were retested on the day following the FCE, patients lifted and carried on average 6 to 9% better than the day before [4]. However, 8 to 21% patients performed worse on the second day. Whether these three observations are related to a large ‘normal’ variation in patient performances or to (un-)safety of an FCE is uncertain, and should be subject of future research. Future studies should also consider larger sample sizes to enable multivariate analyses, for example to predict symptom increase. The sample size of 54 responders and the poor strength of the univariate relationships prevented performance of multivariate analyses.

It may be debated whether a temporary increase in symptoms is related to a temporary loss of function. In different studies, pain intensity related weakly [4, 6] or moderately [8] to lifting and FCE performance, and increased pain intensity did not prevent patients to perform similarly on a consecutive day [4], or two weeks later [2]. As demonstrated in Table 3, pain intensity prior to the FCE related non-significantly to differences in pain intensity following the FCE. If pain were related to tissue damage, it could be hypothesized that patients with high pain levels prior to the FCE would perform less than patients with low pain levels, and vice versa. In a post hoc analysis we found no support for this hypothesis (Pearson’s correlation between pain intensity and lifting performance of  $r = 0.08$ ). Pain intensity is recognized as unrelated to tissue damage in chronic non-specific pain syndromes, including CLBP [27]. These results suggest that tissue damage is unlikely. However, because tissue damage was not assessed objectively, this cannot be concluded with certainty. The temporary increase of pain intensity may also be interpreted as delayed onset muscle soreness after the performance of high intensity activities [28, 29]. This type of soreness may last from 1 to 10 days and is considered a normal response to unusual exertion [28, 29]. As judged by both the mean amount of weight lifted and the percent of maximum heart rate (approximately 70%), performances may indeed qualify as ‘unusual’

and as ‘high intensity’. However, post-FCE responses were at best only weakly related to the performance parameters. Thus, within the context of a chronic pain syndrome, we interpret the self-reported pain increase as a non-specific response to unusual high intensity exertion, rather than the onset of ‘specific’ injuries such as strains or sprains to the low back region. As such, the results appear to confirm clinical terminology: ‘it may hurt you, but it won’t harm you’.

While this study was performed as part of a larger study program, regular clinical FCE procedures were adhered to. With regards to the FCE, ‘extra’ for the patients for this research was the questionnaire following the FCE. This could be regarded as a strength of this study, because the study results represent regular clinical procedures and this research may thus be generalized to everyday practice, as confirmed by a comparison with patient characteristics and study results described in other studies [6, 7]. A weakness of this study was that we did not control for pain medication use in this exploratory study. Perhaps patients started to use or increased pain medication in response to testing and may thus be misclassified as safe. After completion of this study, we regard it as very unlikely that the FCE has led to injuries, but, although we have reasoned that injuries probably did not occur, we cannot be completely sure. Future studies should consider objectively reassessing all patients after completion of the FCE for objective signs of injury.

The nature of this study was exploratory. Because of the absence of a consented operational definition of safety in literature, arbitrary choices were made about operationally defining safety. Within this definition, the results suggest that the FCE may be safe in patients with CLBP. This study may serve as groundwork for future research. Future studies should consider defining the term injury, assess this objectively and should not rely on self-reports only. Future studies should use validated measures for repeated measurements of pain intensity (VAS or NRS scales, instead of the current scales for follow up measurements), control for pain medication use and consider larger sample sizes to enable multivariate analyses to predict symptom increase. Future studies should differentiate between a normal non-specific response to unusual high intensity exertion and the onset of ‘specific’ injuries. Additionally, it is of great importance for the progression of this field that future studies should operationally define ‘safety’.

## Conclusion

A temporary increase in symptom intensity after completion of an FCE is common. Within the operational definitions of safety used in this study, assessment of functional capacity of patients with CLBP appears safe.

**Acknowledgement** The data of this study were collected as part of the low back pain disability (LOBADIS) research program, supported by Zorgonderzoek Nederland (ZON) grant number 96-06-006. The authors wish to acknowledge the support of the physiatrists during this study.

## References

1. Hart DL, Isernhagen SJ, Matheson LN. Clinical commentary: guidelines for functional capacity evaluation of people with medical conditions. *JOSPT* 1993;18:682–6.
2. Brouwer S, Reneman MF, Dijkstra PU, Groothoff JW, Schellekens JMH, Göeken LNH. Test-retest reliability of the Isernhagen Work Systems functional capacity evaluation in patients with chronic low back pain. *J Occup Rehabil* 2003;13:207–18.
3. Reneman MF, Brouwer S, Speelman-Meinema A, Dijkstra PU, Geertzen JHB, Groothoff JW. Test-retest reliability of the Isernhagen Work Systems functional capacity evaluation in healthy adults. *J Occup Rehabil* 2004;14:295–305.



4. Reneman MF, Dijkstra PU, Westmaas M, Göeken LNH. Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. *J Occup Rehabil* 2002;12:269–76.
5. Gross DP, Battie MC. Reliability of safe maximum lifting determinations of a functional capacity evaluation. *Phys Ther* 2002;82:364–71.
6. Reneman MF, Jorritsma W, Dijkstra SJ, Dijkstra PU. The relationship between kinesiophobia and performance in a Functional Capacity Evaluation. *J Occup Rehabil* 2003;13:277–85.
7. Reneman MF, Jorritsma W, Schellekens JMH, Göeken LNH. Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic non-specific low back pain. *J Occup Rehabil* 2002;12:119–30.
8. Gross DP, Battie MC. The construct validity of a kinesiophysical functional capacity evaluation administered within a workers' compensation environment. *J Occup Rehabil* 2003;13:287–95.
9. Smith RL. Therapists' ability to identify safe maximum lifting in low back pain patients during functional capacity evaluation. *JOSPT* 1994;19:277–81.
10. Gardener L, McKenna K. Reliability of occupational therapists in determining safe, maximal lifting capacity. *Aus Occup Ther J* 1999;46:110–9.
11. Isernhagen SJ, Hart DL, Matheson LM. Reliability of independent observer judgements of level of lift effort in a kinesiophysical Functional Capacity Evaluation. *WORK* 1999;12:145–50.
12. Reneman MF, Jaegers SMHJ, Westmaas M, Göeken LNH. The reliability of determining effort level of lifting and carrying in a functional capacity evaluation. *WORK* 2002;18:23–7.
13. Gibson L, Strong J. Safety issues in functional capacity evaluation: findings from a trial of a new approach for evaluating clients with chronic back pain. *J Occup Rehabil* 2005;15:237–51.
14. Hart DL, Berlin S, Brager PE, Caruso M, Hejduk JF, Howar JM, Snyder KP, Susi JL, Wah MD. Development of clinical standards in industrial rehabilitation. *JOSPT* 1994;19:232–41.
15. Hoogendoorn WE, van Poppel MNM, Bongers PM, Koes BW, Bouter LM. Physical load during work and leisure time as risk factors for low back pain: a systematic review. *Scand Work Environ Health* 1999;25:387–403.
16. Hoogendoorn WE, Bongers PM, de Vet HCW, Douwes M, Koes BW, Miedema MC, Ariens GA, Bouter LM. Flexion and rotation of the trunk and lifting at work are risk factors for low back pain: results of a prospective cohort study. *Spine* 2000;25:3087–92.
17. Gross DP, Battie MC. The prognostic value of functional capacity evaluation in patients with chronic low back pain: part 1. Timely return to work. *Spine* 2004;14:914–9.
18. Matheson LN, Mooney V, Grant JE, Affleck M, Hall H, Melles T, Lichter RL, McIntosh G. A test to measure lift capacity of physically impaired adults; Part I: development and reliability testing. *Spine* 1995;19:2119–29.
19. Kuijer W, Brouwer S, Schiphorst Preuper HR, Dijkstra PU, Groothoff JW, Geertzen JHB. Work status and chronic low back pain. Exploring the international classification of functioning, disability and health. *Disabil Rehabil* 2005 (in press);27.
20. Nachemson A, Vingard E. Assessment of patients with back and neck pain: a best-evidence synthesis. In: Nachemson A, Jonsson E, editors. *Neck and back pain: the scientific evidence of causes, diagnostics, and treatment*. Philadelphia: Lippincott Williams & Wilkins; 2000. pp. 57–79.
21. Roland M, Morris R. A study of the natural history of back pain. Part 1: development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983;8:141–4.
22. Brouwer S, Kuijer W, Dijkstra PU, Göeken LNH, Groothoff JW, Geertzen JHB. Reliability and stability of the Roland Morris Disability Questionnaire: intra class correlation and limits of agreement. *Disabil Rehabil* 2004;26:162–5.
23. van Dieen JH, Hoozemans MJ, Toussaint HM. Stoop or squat: a review of biomechanical studies on lifting technique. *Clin Biomech* 1999;14:685–96.
24. Munro BH, Visintainer MA, Batten Page E. *Statistical methods for healthcare research*. Philadelphia: JB Lippincott Company; 1986.
25. King PM, Tuckwell N, Barrett TE. A critical review of Functional Capacity Evaluations. *Phys Therapy* 1998;78:852–66.
26. Newton M, Waddell G. Trunk strength testing with iso-machines. Part 1: review of a decade of scientific evidence. *Spine* 1993;18:801–11.
27. Fordyce WE. *Back pain in the workplace; management of disability in non-specific conditions*. USA: IASP Press; 1995.
28. Powers SK, Howley ET. *Exercise physiology. Theory and application to fitness and performance*. 4th ed. New York: McGraw-Hill; 2001.
29. Coudreuse JM, Dupont P, Nicol C. Delayed post effort muscle soreness. *Ann Readapt Med Phys* 2004;47:290–8.