Factors Associated with Functional Capacity Test Results in Patients With Non-Specific Chronic Low Back Pain

van Abbema, Renske; Lakke, Sandra E.; Reneman, Michiel; van der Schans, Cees P.; van Haastert, Corrien J. M.; Geertzen, Jan H.B; Wittink, Harriet

Published in:
Journal of Occupational Rehabilitation

DOI:
10.1007/s10926-011-9306-4

J Occup Rehabil (2011) 21:455–473
DOI 10.1007/s10926-011-9306-4

Factors Associated with Functional Capacity Test Results in Patients With Non-Specific Chronic Low Back Pain: A Systematic Review

Renske van Abbema · Sandra E. Lakke · Michiel F. Reneman · Cees P. van der Schans · Corrien J. M. van Haastert · Jan H. B. Geertzen · Harriët Wittink

Published online: 23 April 2011

Abstract *Introduction*: Functional capacity tests are standardized instruments to evaluate patients' capacities to execute work-related activities. Functional capacity test results are associated with biopsychosocial factors, making it unclear what is being measured in capacity testing. An overview of these factors was missing. The objective of this review was to investigate the level of evidence for factors that are associated with functional capacity test results in patients with non-specific chronic low back pain. *Methods:* A systematic literature review was performed identifying relevant studies from an electronic journal databases search. Candidate studies employed a cross-sectional or RCT design and were published between 1980 and October 2010. The quality of these studies was determined and level of evidence was reported for factors that were associated with capacity results in at least 3 studies. *Results:* Twenty-two studies were included. The level of evidence was reported for lifting low, lifting high, carrying, and static lifting capacity. Lifting low test results were associated with self-reported disability and specific self-efficacy but not with pain duration. There was conflicting evidence for associations of lifting low with pain intensity, fear of movement/(re)injury, depression, gender and age. Lifting high was associated with gender and specific self-efficacy, but not with pain intensity or age. There is conflicting evidence for the association of lifting high with the factors self-reported disability, pain duration and depression. Carrying was associated with self-reported disability and not with pain intensity and there is conflicting evidence for associations with specific self-efficacy, gender and age. Static lifting was associated with fear of movement/(re)injury. *Conclusions:* Much heterogeneity was observed in investigated capacity tests and candidate associated factors. There was some evidence for biological and psychological factors that are or are not associated with capacity results but there is also much conflicting evidence. High level evidence for social factors was absent.

Keywords Review · Non-specific chronic low back pain · Functional capacity

R. van Abbema (✉) · S. E. Lakke · C. P. van der Schans
Center for Applied Research and Innovation in Health Care and in Nursing, Hanze University, University of Applied Sciences, Eyssoniusplein 18, 9701 DC Groningen, The Netherlands
e-mail: r.van.abbema@pl.hanze.nl

S. E. Lakke · M. F. Reneman · C. P. van der Schans · J. H. B. Geertzen
Department of Rehabilitation Medicine, Center for Rehabilitation, University Medical Center Groningen, Groningen, The Netherlands

H. Wittink
Research Group Lifestyle and Health, University of Applied Sciences, Utrecht, The Netherlands

C. J. M. van Haastert
Policy development and innovation, Dutch Patient Consumer Federation (NPCF), Utrecht, The Netherlands

Introduction

Patients with non-specific Chronic Low Back Pain (CLBP) can be limited in their functioning because of their health condition. Functioning refers to all body functions, activities and participation as classified in 'The International Classification of Functioning, Disability and Health (ICF) [1]. Not only physical limitations determine the level of functioning in patients with non-specific CLBP,

psychosocial factors have proven to have impact as well [2, 3]. In clinical practice, assessments of functioning are performed by means of patient self assessment, clinical assessment and/or capacity tests. These assessments are important to make clinical decisions on choice of therapy, evaluation of interventions, and restriction of activities or return to work. In this study, we focused on factors that associate with capacity test results in patients with non-specific CLBP.

Capacity tests are standardized functional instruments that are used to evaluate patients' capacities to execute (work related) physical activities. There are many terms in the literature that refer to capacity tests, such as physical performance tasks, physical ability, and functional assessment tests. Work related capacity tests are, among others, referred to as Functional Capacity Evaluation (FCE), Functional Capacity Assessment or Work Capacity Evaluation. In the present study, the term capacity test is used as a consistent terminology for all tests that measure the highest probable level of functioning that a person may reach in an activity domain at a given moment in a standardized environment [1, 4].

It is not always clear what is being measured in capacity testing. Personal factors such as age, education, coping style, motivation, fear and environmental factors such as medication or assessment setting may associate with the results of a capacity test. For the interpretation of capacity test results, it is important to take notice of such factors. There have been studies in the past decades that explored the association of factors with capacity test results in patients with chronic pain. A non-systematic review on the association between psychosocial factors and capacity tests in patients with chronic pain concluded that specifically pain related fear, self-efficacy and illness behaviour were related to measures of capacity [3]. However, the relations and underlying mechanisms are complex, because many psychosocial factors are inter-correlated. Over the years, there has been further research on capacity test results in relation to self-reported disability [5, 6], cardiovascular capacity [7], pain severity [5, 7, 8], self-efficacy beliefs [2, 9, 10] and work related recovery expectations [5]. To understand the association of biopsychosocial factors with capacity test outcomes, there is a need for an overview of clinical evidence for these factors.

The objective of the present review was to determine the current level of evidence for factors that associate with capacity test results in patients with non-specific CLBP. An overview level of evidence of these factors provides useful insights for healthcare workers using capacity tests in this population and researchers investigating capacity testing in non-specific CLBP.

Method

Design and Outline

The study design is a systematic review of cross-sectional studies and clinical trials that investigated capacity tests and their potentially associated factors in patients with non-specific CLBP. For the first selection of studies, one researcher (RA) performed an electronic search for potentially relevant studies. Two reviewers (RA and SEL) independently screened titles and abstracts for the second selection. The full texts of the second selection were retrieved and assessed for inclusion by both reviewers. Selection of relevant studies was based on set inclusion and exclusion criteria. In the next stage of the review, relevant studies were assessed for methodological quality and the outcomes were analyzed to determine level of evidence.

Search Strategy

To identify relevant studies, we conducted a search of bibliographic electronic literature databases (MEDLINE, CINAHL, EMBASE and PsychINFO), using keywords, MeSH terms and free text words (supplementary Appendix A). Studies from January 1980 up to October 2010 were searched. Only full reports written in English, German or Dutch and meeting the following inclusion criteria were selected.

Inclusion Criteria

Candidate studies examined a relationship between the results of a capacity test (dependent variable) and one or more associated factors (independent variable). The study population included adults with non-specific CLBP aged from 18 up to 65 years. Studies were included when at least 75% of the population had non-specific CLBP. Non-specific CLBP was defined as back pain not attributed to recognizable specific pathology (e.g., infection, tumour, osteoporosis, ankylosing spondylitis, fracture, inflammatory process, cauda equina syndrome and pregnancy) with a duration of more than 3 months. The capacity tests in the selected studies met the definition of capacity tests according to the ICF, which was adopted by a group of scientists and clinicians in the field of capacity testing [4]. Capacity tests assess 'the highest probable level of functioning that a person may reach in a domain at a given moment in a standardized environment'. Only studies that used capacity tests measuring the activity level of participants were included. Activity is the execution of a task or action by an individual [1]

Quality Assessment

There are recommendations for reporting Meta-analysis Of Observational Studies (MOOSE) [11] and Strengthening the Reporting of Observational studies in Epidemiology (STROBE) [12, 13]. However, no clearly defined tools for assessing quality and susceptibility to bias in cross-sectional studies are available [14, 15]. We developed a checklist based on the key domains of assessing observational studies according to the STROBE checklist, the recommendations of Sanderson et al. (2007) [14], and von Elm (2007) [15] (Table 1). The 8-item checklist includes the following domains to assess: methods of selecting study participants, methods for measuring study variables, addressing design specific sources of bias, control of confounding variables and appropriate use of statistics. Two researchers (RA and SEL) independently performed quality assessment by scoring the checklist. Positive (+) was scored when an item was clearly described, negative (-) was scored when an item was not described, unclear (?) was scored when an item was not clearly described or incomplete. Primary authors were contacted to clarify items rated negative or unclear. One point was assigned to every scored positive item, half a point was assigned to every unclear item, and a total score was calculated. Studies were considered of high quality when at least 6 out of 8 items were rated positive. Studies were considered of low quality when 5 or less items were rated positive. The methodological quality of clinical trials was assessed with the PEDro scale. A PEDro score of at least 5 points (0-10) was considered to be of high quality [16]. Agreement between reviewers on the quality of included studies (+/−/?) was assessed using Cohen's kappa statistics (κ) for categorical variables and rated as poor if $\kappa \leq 0.2$; fair if $0.2 < \kappa \leq 0.4$; moderate if $0.4 < \kappa \leq 0.6$; substantial if $0.6 < \kappa \leq 0.8$; and good if $\kappa > 0.8$ [17].

Data Extraction and Analysis

For each included study, details were extracted on study population, patient characteristics, capacity tests, measurements of the potentially associated factors and the test results. All reported associations were recalculated into R^2 to realise a homogeneous analysis. Furthermore, potential confounders included in regression analyses were extracted for evaluation.

The strength of statistical significant associations between related factors and results of functional capacity test results were rated low if $0.05 \leq R^2 < 0.25$, moderate if $0.25 \leq R^2 < 0.49$ and high if $R^2 \geq 0.50$ [1, 18]. The relationships were interpreted as statistically significant when $p < 0.05$. Not significant associations or if $R^2 < 0.05$ were rated as no association. Level of evidence was reported when at least 3 studies investigated the same capacity test and potentially associated factor. High level evidence was described as consistent results in at least 2 high quality studies, moderate evidence as consistent results in at least one study of high quality, low evidence as consistent results in at least 3 low quality studies, and conflicting evidence as inconsistent results. Consistent means that at least 75% of the included studies had low, moderate, and/or high association, or at least 75% of the included studies had no association with the capacity test results. Absence of evidence was present when less than 3 studies reported on the same capacity test and biopsychosocial variable.

Results

Literature Search

The results of the search strategy are presented in Fig. 1. The literature search of databases resulted in 5534 potentially relevant studies. From the primary search, 5477

Table 1 Quality assessment checklist of cross sectional studies

Item	Number	Criteria
Study population	1	Positive if source of selection of participants is clear and a representative sample of the population intended in the study was selected.
	2	Positive if inclusion and exclusion criteria were clearly described (duration pain, age, gender, employment, co-morbidities).
Measurements	3	Positive if used capacity tests are valid and reliable.
	4	Positive if instruments for associated factors are valid and reliable.
	5	Positive if assessment therapist was blinded for other test outcomes.
Analysis	6	Positive if appropriate univariate statistical method was used to establish the relationship between the associated factors and (the) capacity test result(s) according to the appropriate measurement level.
	7	Positive if appropriate multivariate statistical methods were used to establish the relative contribution of the associated factor to (the) capacity test result(s) according to the appropriate measurement level.
	8	Positive if the intended relationship between a capacity test and an influencing factor was controlled for confounding factors.

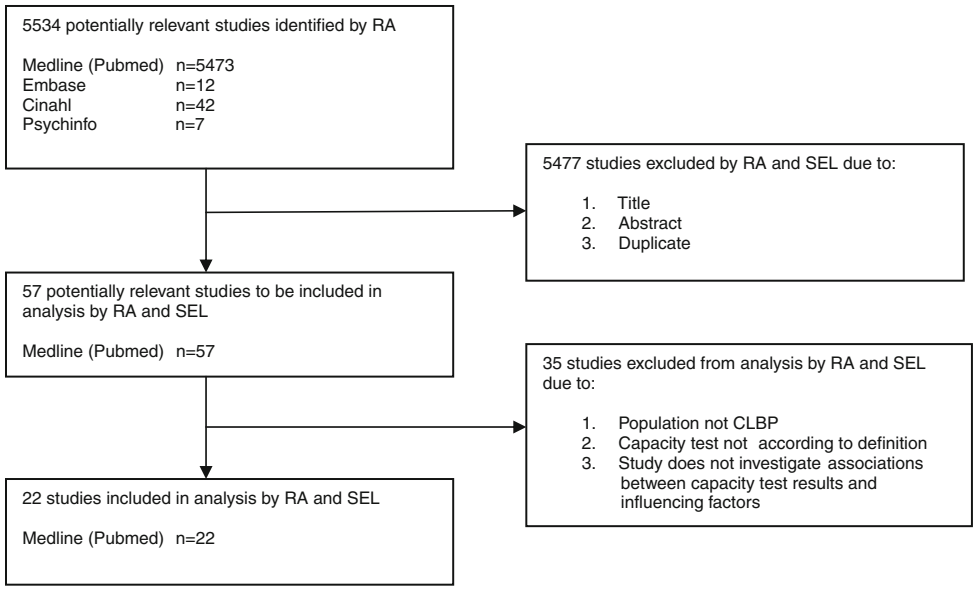

Fig. 1 Selection of relevant studies

studies were excluded on title, abstract and duplicate by 2 researchers (RA en SEL).They read full texts and individually assessed inclusion of relevant studies. These assessments were compared and discussed until consensus was reached on in/exclusion of the 57 remaining studies. As a result, another 35 studies were excluded. The main reason for exclusion was firstly not meeting the targeted population of patients with non-specific CLBP. Secondly, the capacity test used in the study did not meet the intended definition of functional capacity. For example, studies that measured isokinetic trunk strength, or studies only using self-reported measurements of functional capacity were not included in our study. Thirdly, the study did not investigate a direct relationship between capacity test results and an associated factor. For example, studies that investigated a relationship between biopsychosocial factors and outcome following assessment, like return to work, were not included. Finally a total of 22 studies were included according the set inclusion criteria [5–10, 19–33, 36].

Quality of Included Studies

Two researchers (RA en SL) scored the quality of included studies. Agreement on the quality assessment between the 2 investigators was high with a Cohen's kappa of $\kappa = 0{,}85$. The quality of the studies was rated 'high' in 19 studies [5–10, 19, 22–28, 30–32, 34, 36] and "low" in 3 studies [20, 21, 33] (Table 2).

Description of Included Studies

Table 3 presents the population of the included studies, patient's characteristics, associations between functional capacity tests and associated factors, potential confounders, and conclusions. The capacity tests that were used in the included studies measured activities such as lifting low (i.e. lifting floor to waist), lifting high (i.e. lifting waist to overhead), walking, sit to stand, crouching, pushing, pulling and stair climbing. Lifting low was the most performed capacity test. The potentially associated factors that were investigated in the included studies were factors such as depression, pain intensity, pain related fear, fear of movement re-injury, self-reported disability, age, gender, health status, job status, pain duration, aerobic capacity, general and specific self-efficacy. In specific self-efficacy questioning closely resembles the task measured, general self-efficacy measures the subjects' expectations of their capacity in general. Patients were recruited from multidisciplinary rehabilitation centres, pain management programmes or spine clinics. The mean population age in the studies ranged from 37.0 to 45.8 years.

Sixteen studies performed univariate analysis to investigate the relationships between the results of a lifting capacity test and possible influencing factors. Multivariate regression analyses were performed in 11 studies to investigate the relative contribution of associated factors or confounders to capacity test results. Five studies performed a group comparison [8, 24, 26, 28, 29]. Groups were composed based on gender [8, 26, 28], high and low fear of movement/(re)injury [29], and work status [24]. One study was a randomized controlled trial [36].

Level of Evidence

The relation between potentially associated factors and lifting low, lifting high, static lifting and carrying that was investigated in at least 3 studies was merged in Table 4 to extract the level of evidence.

Table 2 Quality assessment

	Design	1. Representative sample and clear source of selection	2. Clear inclusion/ exclusion criteria	3. Valid and reliable capacity test(s)	4. Valid and reliable instruments for associated factors	5. Assessment therapist was blinded for other test outcomes	6. Appropriate univariate statistical methods were used	7. Appropriate multivariate statistical methods were used	8. Relationships were controlled for confounders	Total out of 8 items
Alschuler et al.2007 [6]	CS	+	−	+	+	+	+	+	+	7
Asante et al. 2007 [9]	CS	+	+	+	±	+	+	+	+	7.5
Crombez et al. 1999 [19]	CS	+	?	−	+	+	+	+	+	6
Cutler et al. 2003 [22]	CS	+	+	+	+	?	+	+	+	7
Filho et al. 2002 [21]	CS	+	−	+	+	?	+	−	−	4
Geisser et al. 2000 [23]	CS	+	−	+	+	?	+	+	+	6
Gross et al. 2003 [33]	CS	+	+	+	+	−	+	−	−	5
Gross et al. 2005 [5]	CS	+	+	+	+	−	+	+	+	7
Gross et al. 2008 [36]	CT	PEDro scale								9 (0–10)
Kuijer et al. 2005 [24]	CS	+	+	+	+	+	+	+	+	8
Lackner et al. 1996 [25]	CS	−	+	+	+	+	+	+	+	7
Lackner et al.1999 [31]	CS	−	+	+	+	+	+	+	+	7
Reneman et al. 2002 [32]	CS	+	+	+	+	+	+	+	+	8
Reneman et al. 2003 [26]	CS	+	+	+	+	+	+	+	+	8
Reneman et al. 2006 [27]	CS	+	+	+	+	+	+	+	+	8
Reneman et al. 2007 [8]	CS	+	+	+	+	+	+	+	+	8
Reneman et al. 2008 [10]	CS	+	+	+	±	+	+	+	+	7.5
Schiphorst Preuper et al. 2008 [28]	CS	+	+	+	+	+	+	+	+	8
Smeets et al. 2007 [7]	CS	+	+	+	+	?	+	+	+	7
Teixeira da Cunha-Filho et al. 2010 [20]	CS	+	+	+	+	?	+	−	−	5
Vlaeyen et al. 1995 [34]	CS	+	+	−	+	+	+	+	+	7
Wittink et al. 2001 [30]	CS	+	+	+	+	+	+	+	+	8

CS cross-sectional study, *CT* controlled trial

Table 3 Description of included studies

	Study population characteristics	Factors associated with functional capacity tests						Authors' conclusions about significant associations
			Progressive isoinertial lifting evaluation (PILE)					
			Lifting low		Lifting high			
			R^2	β	R^2	β		
Alschuler et al. [6]	267 patients; 144♂/123♀ The University of Michigan Spine Program, USA Age†: 41.3 (8.6) Pain duration#: 57.8 (77.4) months	Depression (CES-D)‡	−0.06***	−0.25**	−0.06**	−0.23**		Self-reported depression and disability had low associations with PILE results. Gender had low association with the waist to shoulder lift. Age and pain index were not associated to PILE results.
		Self-reported disability (QBPDS)‡	−0.22***	–	0.15***	–		
		Pain Index (MPQ)‡	−0.01	−0.02	−0.01	0.01		
		Age	0.00	−0.01	0.00	0.01		
		Gender	0.00	0.07	0.04**	0.21*		
			Isernhagen Work Systems (IWS)-FCE					
			Lifting low		Lifting high		Carrying	
			R^2	β(adj)	R^2	β(adj)	R^2	β(adj)
Asante et al. [9]	42 patients; 29♂/13♀ Rehabilitation program; Alberta, Canada (workers compensation claimants) Age: 38.4 (10.2) Pain duration:161 (123) days	Predicted floor to waist lift	0.50*	0.68**	0.18*	–	0.37*	–	Functional self-efficacy (predicted lifting and carrying) was associated with better results on the 3 lifting tasks. Self-reported disability and pain intensity were associated (low) with test results of all three lifting tests. The physical components of the SF-36 had low association with test results of all three lifting tests. Age, gender, duration of injury and physical demands of work did not contribute to the three lifting tests.
		Predicted waist to overhead lift	0.35*	–	0.42*	0.59**	0.27*	–	
		Predicted carrying	0.49*	–	0.25*	–	0.53*	0.59**	
		Self-reported pain disability (PDI)‡	−0.12*	−0.10	−0.06	–	−0.17*	−0.01	
		Pain intensity (VAS)‡	−0.10*	0.12	−0.00	0.07	−0.13*	−0.16	
		SF-36‡(physical composition)	0.17*	0.29	0.10*	–	0.14*	0.10	
		SF-36 (mental composition)	0.00	−0.06	0.00	−0.02	0.00	−0.24	
		SF-36 (physical functioning)	0.23*	–	0.08	–	0.21*	–	
		Age		–		–		–	
		Gender		–		–		–	
		Duration of injury		–		–		–	
		Physical demands work		–		–		–	
			Behavioural approach technique (BAT): static lifting (minutes)						
			R^2		β(adj)				
Crombez et al. [19] (Study 3)	31 patients; 13♂/24♀ Rehabilitation Centre; Hoensbroeck, The Netherlands Age: 41.6 (10.7) Pain duration†: 10.1 (8.9) years	Fear of movement/(re)injury (TSK)‡	−0.24**		−0.47**				Fear of movement/(re)injury, self-reported pain disability, pain related fear and negative affect had a low association with static lifting results. Pain intensity, pain increase, catastrophizing, and age were not associated with static lifting. Radiation into the legs, fear of movement/(re)injury and being a women contributed significantly to poorer static lifting results.
		Pain disability (RDQ)‡	−0.18*		–				
		Negative affect (NEM)‡	−0.18*		–				
		Pain related fear (PASS)‡	−0.11*		–				
		Pain catastrophizing (PCS)‡	−0.26		–				
		Pain intensity (VAS)	0.01		–				
		Pain increase	–		–				
		Age	–		–				
		Gender	–		−0.48***				
		Radiation into legs	–		−0.49**				

Table 3 continued

			Dictionary of Occupational Titles (DOT) FCE						
			Climbing		Crouching		Lifting low		
			t/X^2	β	t/X^2	t/X^2	β	t/X^2	
Cutler et al. [22]	188 patients; 100♂/88♀ Multidisciplinary pain treatment center, Miami, USA Age: 40.9 (9.8) Pain intensity† (0–10) (sd): 5.9 (2.5)	Pain intensity (VAS)	3.63*	–	4.57**	−0.26*	3.43*	−0.27*	Pain intensity was associated with results of all three tests. Workers compensation and state anxiety were associated with climbing and crouching results. Depression and stress were associated to climbing. Trait anxiety was associated with crouching.
		Workers compensation	9.35*	−0.96*	13.26**	−0.84*	5.96	–	
		Depression (BDI)‡	4.25*	0.02*	2.57	–	1.54	–	
		State anxiety (STAI)‡	3.18*	–	3.02*	–	0.61	–	
		Trait anxiety (STAI)	2.47	–	2.72*	–	0.75	–	
		Stress (PSS)‡	2.64*	–	1.46	–	0.08	–	

			Sit to Stand (SS) R^2	Time to Roll (TTRL) R^2	5 min walking (5 MW) R^2	50-foot walk (50-FW) R^2	
Filho et al. [21]	51 patients; 23♂/28♀ Outpatient orthopedic spine clinic; Houston, USA Age: 45.8 (9.8) Pain duration: 95 (100.4) months	Self-reported disability (RMDQ)	0.23*	0.19*	−0.17*	0.19*	Self-reported disability had a low association with all four capacity tests. Self efficacy, pain affect, pain intensity and self-reported disability had low associations with the TTRL. Aerobic capacity had a low association with the 5 MW.
		Pain intensity (VAS)	0.01	0.12*	−0.01	0.02	
		Pain affect (VAS)	−0.07	0.12*	−0.01	0.04	
		Self-Efficacy (SES)‡	0.01	0.10*	−0.03	0.04	
		Aerobic capacity (pred. equation)	−0.07	0.00	0.11*	−0.01	

			Pile				
			Lifting low		Lifting high		
			R^2	β	R^2	β	
Geisser et al. [23]	133 patients; 75♂/58♀ University of Michigan Spine Program, USA Age: 41.7 (8.5) Pain duration: 65.3 (86.6) months	Compensation status	−0.04**	−0.06	−0.04**	−0.01	Receiving compensation, involvement in litigation, pain duration, the pain index and depression had negligible or low associations with both PILE results. The TSK-2 avoidance subscale had low associations with both PILE test results. Physiologic and perceived effort were moderately associated to both PILE test results. Gender was associated with the waist to shoulder lift. Age, gender, pain, TSK (fear), BMI and MET were not associated with lifting low test results. Age, BMI and MET were not associated with lifting high test results.
		Litigation staus	−0.07**	−0.20**	−0.05**	−0.06	
		Pain duration	0.03*	0.06	0.04**	0.07	
		Pain Index (MPQ)	−0.02	0.16	−0.04*	0.03	
		Avoidance (TSK)	−0.09***	−0.20**	−0.10***	−0.18*	
		Fear (TSK)	−0.02	0.10	0.08***	−0.05	
		Depression (CES-D)	−0.06**	0.00	−0.06**	0.00	
		Body Mass Index (BMI)	−0.01	−0.10	−0.01	−0.06	
		Metabolic Equivalent (MET)	0.00	0.09	−0.01	0.04	
		Physiologic effort (HRmax)	0.28***	0.27**	0.31***	0.33***	
		Perceived effort (Borg Scale)	0.16***	0.32***	0.17***	0.32***	
		Gender	−0.00	0.20	0.05**	0.20**	
		Age	−0.00	–	0.00	–	

Table 3 continued

			IWS-FCE Lifting low R^2	
Gross et al. [33]	321 patients; 231♂/90♀ Workers' compensation rehabilitation facility; Alberta, Canada Age:42 (9.9) Days from injury†: 737 (1361)	PDI Pain intensity (VAS)	−0.27 −0.20	Self reported pain disability was moderately associated with average maximum weight lifted in 6 lifting tests. Pain intensity had a low association with lifting capacity.

			IWS-FCE Lifting low		
			R^2	β	
Gross et al. [5]	170 patients; 121♂/49♀ Workers' compensation rehabilitation facility; Alberta, Canada Age: 41.0 (10.9) Days from injury: 450 (821) Pain intensity (0–10): 5.0 (2.0)	Pain intensity (VAS) Self-reported pain disability (PDI) Recovery expectations Support at workplace (OPP)‡ Age Gender Duration of injury	−0.18* −0.30* 0.04* 0.00 – – –	– −0.49** – – −0.24** 0.25** –	Pain intensity and self reported pain disability had low to moderate associations with lifting capacity Lifting test results were best predicted by patients perceptions of what they can and cannot do, reflected by the PDI scores and secondary by gender and age. Lifting test results were not or negligibly correlated with recovery expectations support at workplace, or duration of injury.

			FCE		
			Lifting low 1 rep max (kg)	Lifting low repetitive (time to fatigue in sec)	
Gross et al. [36]	30 patients;19♂/11♀ University Hospital Multidisciplinary Pain Centre and local community. Age: 49.4 (16.4) Pain intensity baseline (0–10): 6.0 (2.1)	Opoid administration Placebo Significance and effect size	29.4 (17.9) 25.6 (3.1) $P < 0.02$ ES 0.23 (95% CI, −0.33–0.78)	312 (251.4) 231 (199.9) $P< 0.03$ ES 0.40 (95% CI, −0.21– 0.98)	Functional capacity of lifting low was significantly different between patients under the influence of opioid and patients administered with a placebo.

			IWS-FCE material handling kg (lifting low, overhead lifting, short carry two handed, pushing and pulling)			
				Mean (SD)		
Kuijer et al. [24]	Multidisciplinary pain management programme; Groningen, The Netherlands Age: 38.5 (8.7) Duration of complaints: 75 (24.2) weeks	Men Women	Working Non-working Working Non-Working	178.3 (54.1) 171.2 (53.8) 127.0 (38.8) 114.2 (38.0)	ns ns	Functional capacity of material handling was not significantly different between working and non-working patients

Table 3 continued

			Work capacity evaluation					
			Lifting low R^2	Lifting high R^2	Carrying R^2	Static pushing R^2	Static Pulling R^2	
Lackner et al. [25]	78 patients; 49♂/36♀ Community referrals from physicians Age (range): 37(21–63) Median time since injury (range): 12.7 (2.4–252 months) weeks	FSE‡ with pain expectancy controlled	0.30***	0.14***	0.24***	0.24***	0.42***	Functional Self-Efficacy was low to moderately associated with all work capacity evaluation tasks. Gender contributed to lifting low and lifting high. Pain contributed to lifting high and carrying
		FSE with reinjury expectancy controlled	0.34***	0.18***	0.24***	0.31***	0.40***	
		Pain expectancy with FSE controlled	−0.03	−0.05	−0.08*	−0.05	0.03	
		Reinjury expectancy with FSE controlled	−0.00	−0.01	−0.03	−0.04	−0.00	
			β	β	β	β	β	
		FSE	0.21***	0.16***	0.15***	0.37***	0.79***	
		Gender	−8.9*	−9.88**	−7.45	−12.10	−18.87*	
		Pain	−2.46	−0.77***	−2.08**	−2.85	−5.00	

			WEST 2-work capacity evaluation				
			Lifting low		Lifting high		
			R^2	β	R^2	β	
Lackner et al. [31]	78 patients; 49♂/36♀ Community referrals from physicians Age (range): 37(21–63) Median time since injury (range): 12.7 (2.4–252 months) weeks	Functional Self efficacy	0.35*	0.18**	0.18**	0.16**	Functional self-efficacy and perceived pain control associated with the two lifting tasks. Perceived ability to decrease pain and anxiety were not associated with lifting low test results. Functional self-efficacy contributed to the two lifting test results. Pain intensity and gender contributed to lifting low test results.
		Perceived Pain control (CSQ)‡	0.12**	–	0.07*	–	
		Perceived ability to decrease pain	–	–	–	–	
		Anxiety (T-A PMS)‡	–	–	–	–	
		Pain Intensity	–	−2.43***	–	–	
		Gender	–	−9.58**	–	–	

			IWS-FCE total of 14 activities		
			R^2	t	
Reneman et al. [32]	64 patients; 54♂/10♀ Outpatient university rehabilitation and occupational assessment center, Groningen, The Netherlands. Age: 38.0 (8.9) Pain intensity (0–10): 5.1(2.1)	Self-reported disability (RMDQ)	0.04	–	Self reported disability measured with the OBPDS and the QBPDS were moderately associated with the IWS-FCE activities. Self reported disability measured with the RMDQ was not associated with the IWS-FCE results
		Self-reported disability (OBPDS)	−0.27*		
		Self-reported disability (QBPDS)	−0.25*		
		Gender–man		3.8**	
		–women		2.9**	

Table 3 continued

				IWS-FCE		
				Avoidance R^2	Lifting low R^2	
Reneman et al. [26]	64 patients; ♂54/♀10 Outpatient rehabilitation program, Groningen The Netherlands Age (sd):38.0 (8.9) Pain intensity (0–10): 5.1 (2.1)	Kinesiophobia (TSK) Pain intensity (NRS)‡ Pain duration	Men Women	−0.03 0.04 −0.05 −0.02	−0.01 −0.01 −0.04 −0.04	Pain intensity, kinesiophobia, gender, pain duration, previous episodes of pain and sick leave were not associated with avoidance measured with lifting low and FCE-DOT.

			IWS-FCE			
			Lifting low β	Lifting high β	Carrying β	
Reneman et al. [27]	The Netherlands: 121patients; 71♂/60♀ Outpatient rehabilitation program Age: 38.0 (9.0) Pain intensity (0–100): 51(21.4) Canada: 273 patients; 71♂/202♀ Workers compensations context Age: 41 (9.4) Pain intensity (0–100): 51 (21.4) Switzerland: 170 patients; 79♂/93♀ Inpatient rehabilitation Age: 42 (8.5) Pain intensity (0–100: 51 (21.4)	Assessment setting Self-reported disability (RMDQ) Pain intensity (VAS) Gender Age Duration of back pain	0.52*** −0.29** −0.07 0.28** −0.07* −0.11**	0.29*** −0.41** – 0.31** – −0.11**	0.55*** −0.29** – 0.23** −0.06* −0.14**	Assessment setting had moderate to high associations with lifting results; patients from the Dutch sample lifted significant more weight than patients from the Canadian and Swiss sample. Self-reported disability, gender, age and duration of back pain contributed to lifting capacity. Pain intensity was not associated with lifting capacity.

			IWS-FCE Lifting low			
			R^2	R^2	β	
Reneman et al. [8]	Outpatient rehabilitation Program; Groningen, The Netherlands study 1:79 patients; 49♂/30♀ Age ♂/♀: 37.8/37.8 Pain intensity (0–10): 4.7/5.0 study 2: 58 patients ♂/♀: 39/19 Age♂/♀: 40.4/35.6 Pain intensity (0–10): 4.5/4.9	Study 1 Pain intensity Fear of movement/(re) injury (TSK) Gender Study 2 Pain intensity Fear avoidance (FABQ)‡ activity Fear avoidance (FABQ) work Gender	Women −0.00 −0.03 – 0.01 −0.00 −0.00 –	Men −0.07 −0.01 – −0.25* −0.02 −0.14* –	 – – 0.48** −0.29* – – 0.48**	There was no association between pain intensity in study 1, pain related fear (TSK), the activity scale of the FABQ and lifting low results. Pain intensity and lifting capacity were moderately associated in men in study 2. The work subscale of the FABQ had a low association with lifting capacity in men in study 2. Gender contributed to lifting test results.

Table 3 continued

			Modified Work Well FCE						
			Lifting low		Lifting high		Carrying		
			R^2	β	R^2	β	R^2	β	
Reneman et al. [10]	92 patients; 60♂/32♀ Multidisciplinary pain management programme Groningen, The Netherlands Age: 38.5 (8.7) Pain intensity (0–10): 5.0 (2.1)	Specific SE: (prediction)	0.30*	0.53*	0.07	0.15	−0.02	−0.18	Specific SE was moderately associated with lifting low. Self reported disability had low associations with lifting low and carrying results. The physical component of SF-36 had low associations with lifting capacity. General SE, age, pain intensity, psychosocial distress and the mental subscale of the SF-36 were not associated with the lifting test results. Gender contributed to all lifting test results.
		General SE: (ALCOS)‡	−0.00	−0.20	0.02	−0.08	0.00	0.18	
		Gender	–	0.28*	–	0.51*	–	0.44*	
		Age	−0.00	–	0.00	–	−0.01	–	
		Pain intensity	0.02	–	−0.00	–	0.00	–	
		Psychosocial distress (SCL -90)‡	−0.00	–	−0.00	–	0.00	–	
		Self-reported disability (RMDQ)	0.05*	–	−0.04	–	−0.08*	–	
		Health related quality of life (SF-36):							
		–physical	0.02*	–	0.04*	–	0.06*	–	
		–mental	0.00	–	0.01	–	0.04	–	
		Self Esteem (SES)	−0.00	–	0.01	–	0.00	–	

			Work-Well FCE					
			Lifting low		Carrying		Static forward bend	
			R^2 Men	R^2 Women	R^2 Men	R^2 Women	R^2	
Schiphorst Preuper et al. [28]	92 patients; 60♂/32♀ Multidisciplinary pain management programme; Groningen, The Netherlands Age: 38.5 (8.7) Pain intensity (0–10): 5.0 (2.1)	Psychosocial distress (SCL-90-R)	−0.00	0.04	−0.01	0.00	−0.02	Only fear of movement/(re)injury had a low association with static forward bending Psychosocial distress, depression, general self-efficacy, self esteem, pain cognitions and coping were not associated with the Work-Well capacity tests.
		Depression (BDI)	−0.00	0.04	−0.01	0.00	−0.01	
		General Self-efficacy (ALCOS-SF)	−0.00	−0.00	0.01	−0.01	−0.01	
		Self esteem (SES)	0.00	−0.01	0.00	0.02	0.00	
		Fear of movement/(re) injury (TSK)	−0.00	−0.01	−0.03	−0.00	−0.06*	
		Pain cognitions (PCL-E)‡	−0.00 to 0.06	−0.04 to 0.05	−0.01 to 0.03	−0.11 to -0.00	−0.02 to 0.01	
		Coping (UCL)‡	−0.02 to 0.00	−0.10 to 0.04	−0.07 to 0.00	−0.00 to 0.01	−0.03 to 0.04	

Table 3 continued

			IWS-FCE						
			5-MW	50-FW	Sit to Stand	Loaded forward Reach (LFR)	One minute stair climbing	PILE Lifting	
Smeets et al. [7]	221 patients: 116♂/105♀ Outpatient unit of three rehabilitation centers; Brabant, The Netherlands Age : 41.6 (10.0) Duration of LBP: 56.7 (72.3) months	Model adjusted R^2	0.21	0.22	0.17	0.13	0.34	0.19	Gender was associated with 5-MW, 50-FW, LFR, PILE. Higher pain intensity was associated with 5-MW, 50-FW and stair climbing tests. Higher VO2 max was related to Sit to Stand and stair climbing. More fear of movement/(re)-injury was related to lower PILE results. Higher self-reported depression was related to lower test results on the 5-MW, Sit to Stand, stair climbing and PILE. Higher level of catastrophizing was related with more steps climbed. More internal control was related to higher test results on the 50-MW and stair climbing. Radiating pain, age and pain duration had no associations with test results.
		(only significant variables were displayed)	*β*	*β*	*β*	*β*	*β*	*β*	
		Age	–	–	–	–	–	–	
		Gender	0.16*	0.24*	–	0.22*	–	0.14*	
		Pain duration	–	–	–	–	–	–	
		Radiating pain	–	–	–	–	–	–	
		Pain intensity (VAS)	−0.20*	−0.22*	–	–	−0.29*	–	
		VO2max	–	–	0.16*	–	0.17*	–	
		Fear of movement/(re)-injury (TSK)	–	–	–	–	–	−0.23*	
		Depression (BDI)	−0.18*	–	−0.29*	–	−0.29*	−0.25*	
		Catastrophizing	–	–	–	–	0.28*	–	
		Internal control (PCL)	–	0.17*	–	–	0.17*	–	

			Sit to Stand (SS) R^2	5 Minute Walking (5 MW) R^2	50-Foot Walk (50-FW) R^2	Timed Up and Go (TUG) R^2	
Teixeira da Cunha-Filho et al. [20]	29 patients; 5♂/24♀ Program for low back treatment at the University Center of Belo Horizonte; Minas Gerais, Brazil Age: 39,4 (12,3) Pain intensity (0–10): 4.4 (2.6)	Self-reported disability (RMDQ)	0.19*	−0.15*	0.05	0.03	Self-reported disability had low associations with SS and 5-MW. Pain intensity and self-efficacy were not associated with the functional capacity tests.
		Pain intensity (VAS)	−0.00	−0.10	−0.00	−0.01	
		Self-Efficacy (SES)	0.04	−0.02	0.03	0.03	

			Behavioural Approach Technique (BAT): Static lifting (minutes) R^2	
Vlaeyen et al. [29]	Study 2: 33 patients; 8♂/25♀ Behavioral rehabilitation program; Hoensbroeck, The Netherlands Age: 42.4 (9.7) Pain duration (years): 10.3 (10.1)	Fear of movement/(re)injury	−0.19**	Fear of movement/(re)injury was associated low with static lifting capacity.

Table 3 continued

			Bruce treadmill walking test (Minutes walked) R^2	
Wittink et al. [30]	Outpatient pain management program at New England Medical Centre; Boston, USA Age: 39.9 (8.1) Pain duration: 40.6 (45.3) months	Peak VO2	0.49***	A moderate association was found between peak VO2 and minutes walked. Age, gender, mental health, and pain duration were not associated with minutes walked.
		Age	0.08	
		Gender	0.01	
		Pain intensity (NRS)	0.11	
		Pain duration	0.00	
		Mental Health (SF-36)	0.01	

* $P < 0.05$; ** $P < 0.01$;*** $P < 0.001$; *ns*=non significant

† Age/pain duration/pain intensity/day from injury: mean (SD)

– Association not calculated or not displayed in original article

‡ *CES-D* Center for Epidemiologic Studies Depression Scale, *QBPDS* Quebec Back Pain Disability Scale, *MPQ* McGill Pain Questionnaire, *PDI* Pain Disability Index, *VAS* Visual Analog Scale, *SF-36* Short Form (36) Health Survey, *RDQ* Roland Disability Questionnaire, *TSK* Tampa Scale for Kinesiophobia, *PASS* Pain Anxiety Symptoms Scale, *NEM* Negative Emotionality Scale, *PCS* Pain Catastrophizing Scale, *RMDQ* Roland Morris Disability Questionnaire, *SES* Self-efficacy Scale, *BDI* Beck Depression Inventory, *STAI* State-Trait Anxiety Inventory, *PSS* Perceived Stress Scale, *HRmax* Maximum Heart Rate, *OPP* Organizational Policies and Practices Scale, *CSQ* Coping Strategies Questionnaire; T-A *PMS* Tension-Axiety scale of the Profile of Mood States, *NRS* Numerical Rating Scale, *FABQ* Fear Avoidance Beliefs Questionnaire, *FSES* Functional Self Efficacy Scale, *ALCOS* Algemene Competentie Schaal (Dutch version of the General Self Efficacy Scale), *CSL-90-R* Symptom Checklist, *PCL(-E)* Pain Cognition List, (experimental version), *UCL* Utrecht's Coping List, *OLBPDQ* Owestry Low Back Pain Disability Questionnaire

Table 4 Evidence table

	Lifting low	Lifting high	Carrying	Static lifting
Gender male	C	POS	C	A
Age	C	NO	C	A
Pain intensity	C	NO	NO	A
Pain duration	NO	C	A	A
Self-reported disability	NEG	C	NEG	A
Specific self efficacy	POS	POS	C	A
Fear of movement/(re)-injury	C	A	A	NEG
Depression	C	C	A	A

C Conflicting evidence

POS High level evidence for positive association

NEG High level evidence for negative association

NO High level evidence for no association

A Absence of evidence

Evidence for Factors Associated With Lifting Low

Lifting Low, Gender and Age There is conflicting evidence that gender associates with lifting low test results. Four studies reported absent associations [6, 9, 23, 26] and 6 studies reported a contribution of gender after regression analysis [5, 7, 8, 10, 27, 31]. There is conflicting evidence for associations of age with lifting low test results. Lifting low was not associated with age in 4 studies [6, 9, 10, 23] but age contributed to lifting test results in 2 other studies [5, 27].

Lifting Low, Pain Intensity and Pain Duration There is conflicting evidence for an association of lifting low test results with pain intensity in patients with non-specific CLBP. The only RCT in this review reported a significant difference with a moderate effect size in lifting performance between patients who were administered an opioid and patients who were administered a placebo [36]. In 5 studies low to moderate associations were found for pain intensity [5, 8, 9, 33, 36]. After regression analysis pain intensity contributed to lifting test results in 3 studies [8, 22, 31]. In 7 studies pain intensity had no association with lifting low test results [6–8, 10, 23, 26, 27]. There is high level evidence that lifting low test results have no association with pain duration [5, 7, 9, 23, 26]. Pain duration contributed to the results of the lifting low test in only one study [27].

Lifting Low and Self-Reported Disability There is high level evidence for a low [6, 9, 10] to moderate [5, 32, 33] association of self-reported disability with lifting low test results. After regression analysis, self-reported disability contributed to lifting low in 2 studies [5, 27].

Lifting Low and Specific Self-Efficacy There is high level evidence for the association of specific self-efficacy with lifting low. Three studies reported a moderate association [10, 25, 31] and one study a high association [9]. All 4 studies reported contribution of specific self-efficacy to capacity test results after regression analysis.

Lifting Low, Fear of Movement/(Re)-Injury and Fear Avoidance Beliefs There is conflicting evidence for an association of lifting low test results with fear of movement/(re)injury. Four studies reported an absent association [8, 10, 26, 28]. In one study there was a low association with fear avoidance beliefs, but absent association of fear of movement/(re)-injury with work related activities [8]. Two studies reported contribution of fear of movement/(re)-injury after regression analysis [7, 23].

Lifting Low and Depression There is conflicting evidence for an association of lifting low test results with depression. Two studies did not find an association [22, 28]. Two studies reported a low association between depression and lifting low test results [6, 23]. Two studies reported a contribution of depression after controlling for confounders [6, 7].

Evidence for Factors Associated With Lifting High

Lifting High, Gender and Age There is high level evidence that gender was associated with lifting high. One study found no association [9], and in 5 studies gender contributed to lifting high test results [6, 10, 23, 25, 27]. There is high level evidence that age has no association with lifting high test results, because all studies relating age to lifting high found absent associations [6, 9, 10, 23, 27].

Lifting High and Specific Self Efficacy There is high level evidence that specific self-efficacy has low to moderate associations with lifting high. Two studies reported a low association [25, 31] and one study [9] reported a moderate association. Two studies found a contribution of specific self-efficacy after controlling for confounders [9, 31]. One study reported absent association between lifting high and specific self-efficacy [10].

Lifting High, Pain Intensity and Pain Duration There is high level evidence that lifting high test results have no association with pain intensity in patients with non-specific CLBP [6, 9, 10, 23, 25, 27]. Pain duration contributed in one study [27] to lifting high test results, in 2 other studies no associations were found [9, 23]. This means there is conflicting evidence for association of pain duration with lifting high test results in patients with CLPB.

Lifting High and Self-Reported Disability There is conflicting evidence of the association of lifting high test results with self-reported disability. Two studies reported no association with lifting high [9, 10], one study reported a low association [6], one study reported a moderate association [32], and one study reported a contribution of self-reported disability after multivariate regression analysis [27].

Lifting High and Depression There is conflicting evidence for an association of lifting high with depression in patients with non-specific CLBP. One study reported an absent association [28], 2 studies reported a low association between depression and lifting high test results [6, 23].

Evidence for Factors Associated With Carrying

There is high level evidence that carrying is associated with self-reported disability [9, 10, 27, 32]. There is high level evidence that carrying is not associated with pain intensity [9, 10, 25, 27]. There is conflicting evidence that carrying is associated with specific self-efficacy [9, 10, 25], gender or age [9, 10, 27].

Evidence for Factors Associated With Static Lifting

There is high level evidence that fear of movement/(re)injury has a low association with static lifting test duration [19, 28, 29, 34]. The lifting test used in these studies was specifically designed to measure avoidance in patients with chronic (low) back pain.

Other variables such as assessment setting, aerobic capacity and pain cognitions were investigated in only a few studies. Therefore, there is not enough material to supply a substantiated level of evidence.

Discussion

The objective of the present review was to provide an overview of the current status of information on factors that associate with capacity test results. There is substantial research on factors influencing capacity test results, but there is much heterogeneity in factors and kinds of capacity tests that have been investigated.

There is conflicting evidence for many factors associated to capacity test results in patients with non-specific CLBP. The high level evidence of self-reported disability and specific self-efficacy in relation to capacity test results

is an outcome of interest. It seems that patients' reports of their ability to execute activities is a factor of importance.

Similarly to our results, an earlier review in 2003 reported few psychosocial factors to be directly associated to capacity tests and other functional measures [3]. Social factors such as workers compensation, involvement in litigation, influence of the test evaluator, support from the workplace or from significant others or assessment setting are scarcely investigated in direct relation to results of functional capacity tests. Furthermore, only few studies investigated the relation between biological factors and functional capacity testing in patients with CLBP. Gender and age were related to test results but factors like muscular strength and aerobic capacity were scarcely explored. We should, therefore, conclude that there is currently absence of evidence regarding social and biological/physiological factors.

The strength of this study is the systematic approach to collect evidence from literature on the subject methodologically. This resulted in a useful overview for clinicians that use capacity tests. Researchers can benefit from this review by exploring the gaps in this research area. In the clinical setting, clinicians might use the study results in the diagnostic process when patients with non-specific CLBP have lower test results on a functional capacity test than expected.

In order to create a broad overview of related variables and get insight into the gaps in this research area, we made the choice for a fairly broad research question. As a result, interpretation of the results of all the studies that investigated capacity test results and associated factors was challenging because of the large diversity of capacity tests, potentially associated factors and diversity in measurements for each potential associated factor. This results in some points for discussion.

First, only 4 types of capacity tests were analysed for level of evidence because those tests were studied in relation to the same biopsychosocial factors in at least 3 studies. Furthermore, lifting low was measured in 3 different functional capacity tests (PILE, IWS-FCE and WEST2-Work Capacity Evalutation). We considered the possibility that biopsychosocial factors could have different associations with different capacity tests. However, in one study where this was subject of investigation; the differences in lifting between PILE and IWS-FCE could not be explained by psychosocial variables [35].

Secondly, functional capacity limiting factors could not be extracted from the reviewed studies. For example test end points were often not (clearly) operationalized and reasons for test terminations were not documented in the studies included. It is likely that this has impacted the interpretations of the primary studies and therefore also on this review.

Thirdly, many studies were not clear about, or did not mention assessment timing [5, 6, 19–24, 27, 30, 33]. Assessment timing is an important factor for interpretating the associations between biopsychosocial factors and FCE, especially those variables that may alter as a result of FCE, such as self-efficacy. However, In the 11 studies that did mention assessment timing, all predictor measures were taken prior to the FCE.

Finally, decisions on interpretation of results such as quality of included studies and level of evidence were arbitrary, but thoroughly considered. Because there is no quality assessment list available for cross sectional studies we followed guidelines from the STROBE-checklist and other recommendations on quality assessment of observational studies. Using our checklist, most studies were rated of high quality. One explanation might be that the sensitivity of our self made list was too low, which could have caused a selection bias. Because of the marked structure of reviewing there is the possibility of having excluded literature that is related to the subject of interest, but is not within our inclusion criteria.

From this review arise new areas for further research. An important next step in the research of factors influencing capacity testing is manipulating that factor in an RCT. The Gross et al. paper is one example where pain intensity was manipulated (reduced with medication) with influence on FCE test results [36]. Furthermore, we recommend other research designs to explore mechanisms behind displayed behavior, such as qualitative research on underlying motives of patients who do not reach maximal physical capacity and research on opinions of professionals working with capacity tests on what factors could influence capacity results.

Furthermore, there was a very interesting finding that did not make the final analysis because only one study performed this type of research [27]. The point of interest were social variables and has to do with the research setting. In this study, considerable differences in maximum weight handled on the various FCE items were observed between patients within a Dutch outpatient rehabilitation context, a Canadian workers' compensation context and a Swiss inpatient rehabilitation context. These differences in (financial) consequences for patients undergoing FCE, the role of evaluators and patient-evaluators interactions in different settings is still underexposed, and should be subject of further investigation.

Conclusion

Much heterogeneity was seen in investigated capacity tests and candidate associated factors. The conclusions from this review are first, that there is conflicting evidence for many factors in patients with non-specific CLBP that influence capacity test results and second, there is some high level

evidence that reported factors do or do not associate with capacity test results as follows: High level of evidence was assigned to the association between lifting low and self-reported disability and lifting low and specific self-efficacy but not for duration of pain, and to the association between lifting high and gender and specific self-efficacy, but not for pain intensity and age, and to the association between carrying and self-reported disability but not for pain intensity, and to the association between static lifting and fear of movement in patients with CLBP. Other variables such as assessment setting, aerobic capacity and pain cognitions were investigated in only a few studies. Therefore, there is not enough material to supply a substantiated level of evidence. High level evidence for social factors was absent.

Appendix 1 Search Strategies

Medline (Pubmed version), Cinahl (EBSCO host), PsycINFO (EBSCO host)

1. ("Body Regions"[Mesh] OR "Musculoskeletal System/anatomy and histology"[Mesh] OR shoulder[tw] OR elbow[tw] OR hand[tw] OR extremity[tw] OR hip[tw] OR knee[tw] OR patellofemoral[tw] OR foot[tw] OR toe*[tw] OR arm[tw] OR leg[tw] OR back[tw] OR spine[tw] OR neck[tw])
2. "Pain/diagnosis"[Mesh] OR "Pain/epidemiology"[Mesh] OR "Pain/etiology"[Mesh] OR pain[tw] OR "Occupational Diseases/diagnosis"[Mesh] OR "Occupational Diseases/epidemiology"[Mesh] OR "Occupational Diseases/etiology"[Mesh] OR "Arm Injuries/diagnosis"[Mesh] OR "Arm Injuries/epidemiology"[Mesh] OR "Arm Injuries/etiology"[Mesh] OR "Back Injuries/diagnosis"[Mesh] OR "Back Injuries/epidemiology"[Mesh] OR "Back Injuries/etiology"[Mesh] OR "Hand Injuries/diagnosis"[Mesh] OR "Hand Injuries/epidemiology"[Mesh] OR "Hand Injuries/etiology"[Mesh] OR "Hip Injuries/diagnosis"[Mesh] OR "Hip Injuries/epidemiology"[Mesh] OR "Hip Injuries/etiology"[Mesh] OR "Leg Injuries/diagnosis"[Mesh] OR "Leg Injuries/epidemiology"[Mesh] OR "Leg Injuries/etiology"[Mesh] OR "Neck Injuries/diagnosis"[Mesh] OR "Neck Injuries/epidemiology"[Mesh] OR "Neck Injuries/etiology"[Mesh] OR "Tendon Injuries/diagnosis"[Mesh] OR "Tendon Injuries/epidemiology"[Mesh] OR "Tendon Injuries/etiology"[Mesh] OR "Fibromyalgia/diagnosis" [Mesh] OR "Fibromyalgia/epidemiology"[Mesh] OR "Fatigue Syndrome, chronic/diagnosis"[Mesh] OR "Fatigue Syndrome, chronic/epidemiology"[Mesh] OR "Fatigue Syndrome, chronic/etiology"[Mesh] OR "Myofascial Pain Syndromes/diagnosis" [Mesh] OR "Myofascial Pain Syndromes/epidemiology"[Mesh] OR "Myofascial Pain Syndromes/etiology"[Mesh] NOT osteoarthritis[Mesh] NOT "Rheumatoid arthritis"[Mesh] NOT.
3. "Physical capacity"[tw] OR "Physical performance"[tw] OR "Physical ability"[tw] OR "Physical activity"[tw] OR "Physical functioning"[tw] OR "Physical test"[tw] OR "Functional test"[tw] OR "Physical measures"[tw] OR "Functional performance"[tw] OR "Functional ability"[tw] OR "Functional health status"[tw] OR "Functional limitations"[tw] OR "Functional testing"[tw] OR "Disability evaluation"[Mesh] OR "Functional capacity"[tw] OR "Behavioural performance"[tw] OR "Activity level"[tw] OR "Activity limitations"[tw] OR "Work capacity evaluation"[Mesh] OR "Functional capacity evaluation"[tw] OR "Functional capacity assessment"[tw] OR "Functional assessment"[tw] OR "Physical capacity evaluation"[tw] OR "Task performance and analysis"[Mesh] OR "Employee performance appraisal"[Mesh] OR "Physical performance test"[tw] OR "Physical ability test"[tw] OR "Assessment/rehabilitation"[tw] OR Walking[tw] OR Lifting[tw] OR "Lifting capacity"[tw] OR "Reaching task"[tw] OR "Functional reach"[tw] OR "Exercise test"[Mesh] OR "Exercise test"[tw].
4. "construct validity"[tw] OR "measurement properties"[tw] OR OR "pain measurements"[tw] OR questionnaires[Mesh] OR evaluation[tw] OR evaluating[tw] OR relation[tw] OR relationship[tw] OR contribution[tw] OR contributing[tw] OR appraisal[tw] OR determinant[tw] OR determinants[tw] OR influence[tw] OR influencing[tw] OR kinesiophobia[tw] OR "fear avoidance"[tw] OR fear[tw] OR "activity avoidance"[tw] OR avoidance[tw] OR "pain-related fear"[tw] OR "illness behaviour"[tw] OR catastrophizing[tw] OR "psychological factors"[tw] OR.
   a. "Comparative study" [Mesh] OR "Cross-sectional study"[Mesh] OR research support AND Limits: Humans, English NOT medication.
5. 1 AND 2 AND 3 AND 4.

Records Medline 5068, Cinahl 1337, Psycinfo 45

EMBASE (EMBASE.com - Elsevier. Records from EMBASE. Unique Medline is excluded)

1. (('shoulder'/exp OR 'shoulder') OR ('elbow'/exp OR 'elbow') OR ('hand'/exp OR 'hand') OR ('extremity'/exp OR 'extremity') OR ('hip'/exp OR 'hip') OR ('knee'/exp OR 'knee') OR patellofemoral OR ('foot'/exp OR 'foot') OR toe* OR ('arm'/exp OR 'arm') OR

('leg'/exp OR 'leg') OR ('back'/exp OR 'back') OR ('spine'/exp OR 'spine') OR ('neck'/exp OR 'neck') OR ('musculoskeletal system'/exp OR 'musculoskeletal system'))
2. (('pain'/exp OR 'pain') OR ('injury'/exp OR 'injury') OR ('head and neck injury'/exp) OR ('musculoskeletal injury'/exp) OR ('musculoskeletal pain'/exp) OR ('disability'/exp))
3. (('cohort analysis'/exp OR 'cohort analysis') OR ('expectancy'/exp OR 'expectancy') OR ('prevalence'/exp OR 'prevalence') OR ('probability'/exp OR 'probability') OR ('risk'/exp OR 'risk') OR ('epidemiology'/exp OR 'epidemiology') OR ('disease course'/exp OR 'disease course') OR ('prognosis'/exp OR 'prognosis') OR ('prediction'/exp OR 'prediction') OR ('epidemiological data'/exp OR 'epidemiological data') OR ('prospective study'/exp OR 'prospective study') OR ('retrospective study'/exp OR 'retrospective study') OR ('longitudinal study'/exp OR 'longitudinal study') OR ('case study'/exp OR 'case study') OR ('epidemiology'/exp OR 'epidemiology') OR (predict* OR prognos*))
4. (('meta analysis'/exp OR 'meta analysis') OR ('systematic review'/exp OR 'systematic review'))) AND [humans]/lim AND [embase]/lim AND [2000-2007]/py
5. 1 and 2 and 3 and 4

Records Embase 1487

Appendix 2

See Table 5.

Table 5 Overview associations for level of evidence

Association/ES Associated factor	Lifting low					Lifting high				
	No	Low	Moderate	High	Regression	No	Low	Moderate	High	Regression
Gender	[6, 9, 23, 24]				[5, 7, 8, 10, 25, 27, 31]	[9]	[23]			[6, 10, 23, 25, 27]
Age	[6, 7, 9, 10, 23]				[5, 27]	[6, 9, 10, 23, 27]				
Aerobic capacity VO2max	[7]									
Work status	[24, 26]					[24]				
BMI	[23]					[23]				
Pain intensity and pain index	[6, 7, 8, 10, 23, 26, 27]	[5, 9, 33]	[8, 36]		[8, 21, 31]	[6, 9, 10, 23, 25, 27]				
Pain duration	[5, 7, 9, 23, 26]				[27]	[9, 23]				[27]
Radiation into legs										
Pain expectations										
Pain cognitions	[7, 28]						[31]			
Self reported disability		[6, 9, 10]	[5, 32, 33]		[5, 27]	[9, 10]	[6]			[27]
Specific self efficacy			[10, 25, 31]	[9]	[9, 10, 31, 25]	[10]	[31, 25]	[9]		[9, 31]
General self efficacy	[10, 28]					[10]				
Fear of movement/ (re)-injury	[8, 10, 26, 28]	[23]			[7, 23]		[23]			[23]
Fear Avoidance	[8]	[8]								
Catastrophizing	[7]									
Depression	[22, 28]	[6, 23]			[6, 7]	[28]	[6, 23]			[6]
Negative affect										
Self esteem	[10, 28]					[10]				
State trait anxiety	[22, 31]					[31]				
Stress	[10, 22]					[10]				

Table 5 continued

Association/ES Associated factor	Lifting low					Lifting high				
	No	Low	Moderate	High	Regression	No	Low	Moderate	High	Regression
Recovery expectations	[5]									
Coping	[28]									
Assessment setting					[27]					[27]
Health status	[10]	[9]				[10]	[9]			
Compensation status	[22, 23]					[23]				
Litigation status		[23]			[23]		[23]			
Metabolic Equivalent (MET)	[23]					[23]				
Physiologic effort			[23]		[23]			[23]		[23]
Perceived effort			[23]		[23]			[23]		[23]
Support at workplace	[5]									

Association/ES Associated factor	Carrying					Static lifting				
	No	Low	Moderate	High	Regression	No	Low	Moderate	High	Regression
Gender	[9]				[10, 27]					[7, 19]
Age	[9, 10]	[27]				[7, 19]				
Aerobic capacity VO2max						[7]				
Work status	[24]									
BMI										
Pain intensity and pain index	[10, 25, 27]	[9]				[7, 19]				
Pain duration	[9]				[27]	[7]				
Radiation into legs										[19]
Pain expectations							[19]			
Pain cognitions	[28]					[7, 28]				
Self reported disability		[9, 10, 32]			[27]		[19]			
Specific self efficacy	[10]	[25]		[9]	[9]					
General self efficacy	[10, 28]					[28]				
Fear of movement/(re)-injury	[28]					[7]	[19, 28, 29, 34]			[19]
Fear Avoidance										
Catastrophizing						[7, 19]				
Depression						[7, 28]				
Negative affect							[19]			
Self esteem	[10, 28]					[28]				
State trait anxiety										
Stress	[10]					[28]				
Recovery expectations										
Coping	[28]					[28]				
Assessment setting					[27]					
Health status		[9, 10]								
Compensation status										
Litigation status										
Metabolic Equivalent (MET)										
Physiologic effort										
Perceived effort										
Support at workplace										

References

1. World Health Organization. ICF: International classification of functioning, disability and health. Geneva: World Health Organization; 2001.
2. Rudy TE, Lieber SJ, Boston JR, Gourley LM, Baysal E. Psychosocial predictors of physical performance in disabled individuals with chronic pain. Clin J Pain. 2003;19:18–30.
3. Geisser ME, Robinson ME, Miller QL, Bade SM. Psychosocial factors and functional capacity evaluation among persons with chronic pain. J Occup Rehabil. 2003;13:259–76.
4. Soer R, van der Schans CP, Groothoff JW, Geertzen JH, Reneman MF. Towards consensus in operational definitions in functional capacity evaluation: a Delphi Survey. J Occup Rehabil. 2008;18:389–400.
5. Gross DP, Battie MC. Factors influencing results of functional capacity evaluations in workers' compensation claimants with low back pain. Phys Ther. 2005;85:315–22.
6. Alschuler KN, Theisen-Goodvich ME, Haig AJ, Geisser ME. A comparison of the relationship between depression, perceived disability, and physical performance in persons with chronic pain. Eur J Pain. 2008;12:757–64.
7. Smeets RJ, van Geel AC, Kester AD, Knottnerus JA. Physical capacity tasks in chronic low back pain: what is the contributing role of cardiovascular capacity, pain and psychological factors? Disabil Rehabil. 2007;29:577–86.
8. Reneman MF, Schiphorst Preuper HR, Kleen M, Geertzen JH, Dijkstra PU. Are pain intensity and pain related fear related to functional capacity evaluation performances of patients with chronic low back pain? J Occup Rehabil. 2007;17:247–58.
9. Asante AK, Brintnell ES, Gross DP. Functional self-efficacy beliefs influence functional capacity evaluation. J Occup Rehabil. 2007;17:73–82.
10. Reneman MF, Geertzen JH, Groothoff JW, Brouwer S. General and specific self-efficacy reports of patients with chronic low back pain: are they related to performances in a functional capacity evaluation? J Occup Rehabil. 2008;18:183–9.
11. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. JAMA. 2000;283:2008–12.
12. Vom Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The strengthening the reporting of observational studies in epidemiology (STROBE) guidelines for reporting observational studies. Lancet. 2007;370:1453–7.
13. von Elm E, Egger M. The scandal of poor epidemiological research. BMJ. 2004;329:868–9.
14. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol. 2007;36:666–76.
15. von Elm E. Commentary: assessing the quality of observational studies-or a lesson from Mars. Int J Epidemiol. 2007;36:677–8.
16. van Tulder M, Furlan A, Bombardier C, Bouter L, The editorial board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the Cochrane collaboration back review group. Spine. 2003;28(12):1290–9.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74.
18. Munro BH, Visintainer MA, Batten Page E. Statistical methods for health care research. Philadelphia: Lippencott; 1986.
19. Crombez G, Vlaeyen JW, Heuts PH, Lysens R. Pain-related fear is more disabling than pain itself: Evidence on the role of pain-related fear in chronic back pain disability. Pain. 1999;80:329–39.
20. Teixeira da Cunha-Filho I, Lima FC, Guimaraes FR, Leite HR. Use of physical performance tests in a group of Brazilian Portuguese-speaking individuals with low back pain. Physiother Theory Pract. 2010;26:49–55.
21. Filho IT, Simmonds MJ, Protas EJ, Jones S. Back pain, physical function, and estimates of aerobic capacity: what are the relationships among methods and measures? Am J Phys Med Rehabil. 2002;81:913–20.
22. Cutler RB, Fishbain DA, Steele-Rosomoff R, Rosomoff HL. Relationships between functional capacity measures and baseline psychological measures in chronic pain patients. J Occup Rehabil. 2003;13:249–58.
23. Geisser ME, Haig AJ, Theisen ME. Activity avoidance and function in persons with chronic back pain. J Occup Rehabil. 2000;10:215–27.
24. Kuijer W, Brouwer S, Preuper HR, Groothoff JW, Geertzen JH, Dijkstra PU. Work status and chronic low back pain: exploring the international classification of functioning, disability and health. Disabil Rehabil. 2006;28:379–88.
25. Lackner JM, Carosella AM, Feuerstein M. Pain expectancies, pain, and functional self-efficacy expectancies as determinants of disability in patients with chronic low back disorders. J Consult Clin Psychol. 1996;64:212–20.
26. Reneman MF, Jorritsma W, Dijkstra SJ, Dijkstra PU. Relationship between kinesiophobia and performance in a functional capacity evaluation. J Occup Rehabil. 2003;13:277–85.
27. Reneman MF, Kool J, Oesch P, Geertzen JH, Battie MC, Gross DP. 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.
28. Schiphorst Preuper HR, Reneman MF, Boonstra AM, Dijkstra PU, Versteegen GJ, Geertzen JH, Brouwer S. Relationship between psychological factors and performance-based and self-reported disability in chronic low back pain. Eur Spine J. 2008;17:1448–56.
29. Vlaeyen JW, Kole-Snijders AM, Boeren RG, van EH. Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. Pain. 1995;62:363–72.
30. Wittink H, Rogers W, Gascon C, Sukiennik A, Cynn D, Carr DB. Relative contribution of mental health and exercise-related pain increment to treadmill test intolerance in patients with chronic low back pain. Spine. 2001;26:2368–74.
31. 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.
32. Reneman MF, Jorritsma W, Schellekens JM, Goeken LN. Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. J Occup Rehabil. 2002;12:119–29.
33. 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.
34. Vlaeyen JW, Crombez G. Fear of movement/(re)injury, avoidance and pain disability in chronic low back pain patients. Man Ther. 1999;4:187–95.
35. Soer R, Poels BJJ, Geertzen JHB, Reneman MF. A comparison of two lifting assessment approaches in patients with chronic low back pain. J Occup Rehabil. 2006;16:639–46.
36. Gross DP, Bhambhani Y, Haykowsky MJ, Rashiq S. Acute opioid administration improves work-related exercise performance in patients with chorinic back pain. J Pain. 2008;9:856–62.