

# Are performance-based measures predictive of work participation in patients with musculoskeletal disorders? A systematic review

P. P. F. M. Kuijer · V. Gouttebauge ·  
S. Brouwer · M. F. Reneman · M. H. W. Frings-Dresen

Received: 1 February 2011 / Accepted: 24 May 2011 / Published online: 10 June 2011  
© The Author(s) 2011. This article is published with open access at Springerlink.com

## Abstract

**Objective** Assessments of whether patients with musculoskeletal disorders (MSDs) can participate in work mainly consist of case history, physical examinations, and self-reports. Performance-based measures might add value in these assessments. This study answers the question: how well do performance-based measures predict work participation in patients with MSDs?

**Methods** A systematic literature search was performed to obtain longitudinal studies that used reliable performance-based measures to predict work participation in patients with MSDs. The following five sources of information were used to retrieve relevant studies: PubMed, Embase, AMA Guide to the Evaluation of Functional Ability, references of the included papers, and the expertise and personal file of the authors. A quality assessment specific for prognostic studies and an evidence synthesis were performed.

**Results** Of the 1,230 retrieved studies, eighteen fulfilled the inclusion criteria. The studies included 4,113 patients,

and the median follow-up period was 12 months. Twelve studies took possible confounders into account. Five studies were of good quality and thirteen of moderate quality. Two good-quality and all thirteen moderate-quality studies (83%) reported that performance-based measures were predictive of work participation. Two good-quality studies (11%) reported both an association and no association between performance-based measures and work participation. One good-quality study (6%) found no effect. A performance-based lifting test was used in fourteen studies and appeared to be predictive of work participation in thirteen studies.

**Conclusions** Strong evidence exists that a number of performance-based measures are predictive of work participation in patients with MSDs, especially lifting tests. Overall, the explained variance was modest.

**Keywords** Functional capacity · Low back · Upper extremity · Lower extremity · Work ability · Predictive validity

---

P. P. F. M. Kuijer (✉) · V. Gouttebauge ·  
M. H. W. Frings-Dresen  
Coronel Institute of Occupational Health, Academic Medical  
Center, University of Amsterdam, PO Box 22700,  
1100 DE Amsterdam, The Netherlands  
e-mail: p.p.kuijer@amc.uva.nl

S. Brouwer  
Department of Health Sciences, Community and Occupational  
Medicine, University Medical Center Groningen, University of  
Groningen, Groningen, The Netherlands

M. F. Reneman  
Department of Rehabilitation Medicine, Center  
for Rehabilitation, University Medical Centre Groningen,  
University of Groningen, Groningen, The Netherlands

## Introduction

The assessment of whether an employee is able to participate in work is complex (Slebus et al. 2007). According to the World Health Organizations' International Classification of Functioning, Disability, and Health (ICF), participation depends on the following five components: disease and disorder, functions and structures, activities, environmental factors, and personal factors (WHO 2001). In case of a disease or disorder, the assessment of whether or not a patient is able to work is often performed by physicians and is traditionally based on legislation, administrative rules, and the physicians' expertise (De Boer et al. 2009). These

assessments are performed for return-to-work decisions and for disability claim assessments. For most physicians, these assessments consist of a comparison between the work ability of a patient and the required demands of a job (Söderberg and Alexanderson 2005; Slebus et al. 2007). Where the work ability matches the job, a person is considered to be able to participate in work. Since there are few instruments available to support physicians in these assessments, it is not surprising that the reliability—a major indicator of an instrument's measurement quality—of these assessments performed by physicians specifically trained for these tasks varied between “poor” and “good” (Brouwer et al. 2003; Spanjer et al. 2010; Slebus et al. 2010).

For the assessment of work ability in patients with musculoskeletal disorders (MSDs), reliable questionnaires and performance-based measures are available (Wind et al. 2005). A theoretical advantage of the use of performance-based measures above questionnaires might be that the face validity is higher: After all, a client performs work-related activities in a specific environmental context (Soer et al. 2008). In line with this assumption, Wind et al. (2009a) showed that performance-based information was found to have complementary value in the assessment of the physical work ability of claimants with MSDs according to 68% of the physicians. In addition, these same physicians change their judgment of the physical work ability of claimants with MSDs in the context of disability claim procedures more often when performance-based outcomes are provided versus traditional information obtained from anamnesis and the medical file (Wind et al. 2009b). Despite these supportive findings for the use of performance-based measures in the assessment for work participation in patients with MSDs, a recent Cochrane review concluded that there is no evidence available for or against the effectiveness of performance-based measures compared with no assessment as intervention for preventing occupational re-injuries in workers with MSDs (Mahmud et al. 2010). The predictive validity of these measures for work participation, however, was not studied. Until now, it is only known that the assessment of work ability in patients with MSDs using a patient's questionnaire, a clinical examination by a physician or by performance-based measures resulted in large differences regarding the estimated work ability (Brouwer et al. 2005). The questionnaire resulted in the highest amount of work limitations and in the performance-based measures in the lowest amount. Therefore, to shed more light on the predictive validity of performance-based measures for the participation in work, a systematic review was performed to answer the following question: “How well do performance-based measures predict work participation in patients with MSDs?” As far as we know,

this review is the first on the predictive validity of performance-based tests for work participation since the review of Innes and Straker (1999). Their review demonstrated paucity in studies focussing on predictive validity. The answer to the research question is relevant because few instruments are available to support physicians in work ability assessments and performance-based measures are not often used (De Boer et al. 2009; Wind et al. 2006), probably partly due to its unknown value for work participation.

## Methods

A systematic review of the literature was performed. The following five sources of information were used to retrieve relevant studies: PubMed (until October 21, 2010), Embase (until October 21, 2010), reference list of Chapter 21 of the American Medical Association Guide to the Evaluation of Functional Ability (Genovese and Galper 2009), references of the included papers were also checked for other potentially relevant papers, and relevant papers suggested by the authors based on their expertise and their personal file. The search terms for PubMed and Embase are listed in “Appendix A” and were based on the PubMed prognosis filter and the search terms for work as suggested by Schaafsma et al. (2006).

After checking for duplicates, the following inclusion criteria were applied to the title and abstract by two reviewers (PK and VG or MFD):

- The paper is a primary study;
- The population of interest are employees with MSDs;
- The study design is a prospective or retrospective cohort study or an intervention study (in the latter case, the data of the group tested with a performance-based measure were used);
- The paper describes a reliable physical test of performance;
- The outcome measure is work participation such as in return to work, or being employed, or a surrogate like the termination of a disability claim;
- The result of a physical test of performance is statistically related to the outcome measure;
- The paper is written in English, Dutch, German, French, or Italian.

If title and abstract did not provide enough information to decide whether the inclusion criteria were met, the full paper was checked. Next, the inclusion criteria were applied to the full paper. When doubts existed about whether a paper fulfilled the inclusion criteria, one other researcher (VG or MFD) was consulted and a decision was made based on consensus. Finally, the references of the

included papers were also checked for other potentially relevant papers.

### Quality description

The quality description of the selected studies was based on an established criteria list for assessing the validity of prognostic studies, as recommended by Altman (2001) and modified by Scholten-Peeters et al. (2003) and Cornelius et al. (2010). This list consisted of 16 items, each having yes/no/don't know answer options. This modified criteria list is presented in “Appendix B”. The quality of all included studies was independently scored by two reviewers (PK, VG). If the study complied with the criterion, the item was rated with one point. If the study did not comply with the criterion or when the information was not described or unclear, then the item was rated with zero points. In case of disagreement, the two reviewers came to a decision through mutual agreement. For the total quality score, all points of each study were added together (maximum score is 16 points). Studies achieving a score of at least 13 points ( $\geq 81\%$ ) were considered to be of good quality, at least 9 (56%) and a maximum of 12 points (75%) of moderate quality, and those with 8 points (50%) or less of low quality.

### Data extraction

Data were extracted by the first author using a standardized form (PK). The following information was extracted as follows: primary author, year of publication, country, study design (cohort (retrospective or prospective) or intervention), characteristics of the population (i.e., number of employees, age and type of MSD), description of the treatment, description of the reliable performance-based test, the confounders taken into account, and the main result of the study regarding the performance-based test and work participation, and a summary of whether the test was significantly related to work participation (yes, no). A distinction was made between studies with good, moderate, and poor quality based on the quality description.

### Evidence synthesis

For the best evidence synthesis, we used the following rules adapted from Van Tulder et al. (2003) and De Croon et al. (2004): (1) if there are four or more studies, the statistically significant findings of 75% or more of the studies in the same direction were taken into account; (2) if there are three studies, the statistically significant findings of at least two studies in the same direction were taken into account; (3) if there are two studies, the statistically significant findings of both studies in the same direction were

taken into account; (4) if there is one study, the statistically significant finding was taken into account. Otherwise, the evidence is “conflicting” regarding the relation between a performance-based measure and work participation. In addition, using the methodological quality scores, the corresponding level of evidence was scored as strong where the result is based on at least two or more good-quality studies, moderate in case of one good-quality study, and limited in all other cases.

## Results

### Search strategy

The search strategy resulted in 588 studies in PubMed and 642 studies in Embase. A total of 167 duplicate studies were found in these two databases. After applying the inclusion criteria to the remaining 1,063 studies, 17 studies remained. Chapter 21 “The scientific status of functional capacity evaluation” of the American Medical Association Guide to the Evaluation of Functional Ability did not result in an additional study. Neither did the experts suggest any additional studies that fulfilled the inclusion criteria. Finally, checking the references of the included studies resulted in one more study, making a total of 18 studies from eight countries: Canada, China, Germany, the Netherlands, Norway, Switzerland, and the United States of America.

### Quality of the studies

The two raters agreed on a total of 261 of the 288 items (91%) for the 18 studies, with a mean difference of 1.5 per paper (SD 1.7, range 0–4). After reaching consensus, five (28%) of the 18 studies were of good quality and the remaining thirteen (72%) of moderate quality (Table 1). The mean quality score was 12 (SD = 2, range 9–14). The four quality criteria that received the least number of points across all studies were as follows: (1) the participants were not recruited during the same uniform period in time after for instance sick leave (1 out of 18 points), (2) no description of the relevant characteristics of the completers and the drop outs (8 out of 18 points), (3) no multivariate analysis was performed taking into account possible confounders (9 out of 18 points), and (4) the treatment was not described and/or standardized (9 out of 18 points).

### Characteristics of the studies

The 18 studies reported on 4,113 participants (median = 147, IQR = 152, range 30–650) (Table 2). Ten

**Table 1** Quality description of the included studies according to the criteria of study population (inception cohort, description of source population, description of inclusion/exclusion criteria), follow-up (at least 12 months, drop outs, description of completers and drop outs, design of the study), treatment (standardized), prognostic factors (relevant, valid, presented), outcome (relevant, valid, presented), analysis (univariate, multivariate), and the quality score (good  $\geq 13$ ,  $9 \leq$  moderate  $\leq 12$ ) (See also “Appendix B” for definitions)

Primary author year of publication	Study population			Follow-up			Treatment			Prognostic factors			Outcome			Analysis			Quality		
	Incept	Pop	Incl	Year	%Out	Descrp	Dsgn	Stnd	Rlvnt	Valid	Pres	Rlvnt	Valid	Pres	Rlvnt	Valid	Pres	Uni	Multi	Sum	
<i>Good quality</i>																					
Gross et al. (2004)	0	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	13
Gross and Battié (2004)	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	14
Gross and Battié (2005)	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	14
Gross and Battié (2006)	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	14
Streibel et al. (2009)	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	13
<i>Moderate quality</i>																					
Bachman et al. (2003)	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	12
Branton et al. (2010)	0	1	1	1	1	1	1	0	1	1	0	1	1	0	1	1	0	1	1	1	12
Cheng and Cheng (2010)	0	1	1	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	12
Fishbain et al. (1999)	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	11
Goutteborge et al. (2009a)	1	1	1	1	1	1	1	0	1	0	0	0	1	1	1	1	1	1	0	0	11
Gross et al. (2006)	0	1	1	1	0	0	1	0	0	0	0	0	1	1	1	1	1	1	1	1	9
Hazard et al. (1991)	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	12
Kool et al. (2002)	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	12
Lechner et al. (2008)	0	1	1	0	1	1	1	1	0	0	0	0	1	1	1	1	1	0	0	0	9
Matheson et al. (2002)	0	1	1	0	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	11
Mayer et al. (1986)	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	10
Strand et al. (2001)	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	12
Vowles et al. (2004)	0	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	10
Sum score	1	17	17	13	12	8	18	9	15	15	13	15	18	18	18	18	17	11	11	9	

studies reported on patients with low back pain, six studies in patients with musculoskeletal disorders (MSDs) in general, and in one study on patients with upper extremity disorders. In one study, the type or region of the MSDs was not specified. In at least 78% of the studies (14/18), the MSDs were described as chronic. Seventeen of the 18 studies took place in a rehabilitation setting and one in an occupational setting. The median follow-up period of the studies is 12 months (IQR = 3, range 3–30 months). Type of treatment was described in 50% (9/18) of the studies. The other studies only described the care provider or gave no description. In 67% of the studies (12/18), confounders were taken into account to establish the relation between performance-based measures and work participation. The median number of confounders taken into account was 3 (SD = 5, range 0–14). The confounders varied between disease characteristics like pain intensity, pain-related disability or depression, personal characteristics like age, work-related recovery expectations, or being a breadwinner, and work characteristics like physical work demand level, pre-injury annual salary, or organizational policies and practices.

#### Performance-based tests and work participation

Thirteen out of the 18 studies used a so-called functional capacity evaluation (FCE): nine studies used the Workwell System (formerly Isernhagen Work Systems), one used the BT Work Simulator, one the ErgoKit, one the Dictionary of Occupational Titles residual FCE, and one the Physical Work Performance Evaluation (Table 2). In five of these thirteen studies, a limited number of tests of the total FCE were used. The other five studies used tests or combinations of like a step test, a lift test, or a trunk strength tester. Two studies combined the results of the performance-based test with non-performance-based outcomes like pain and Waddell signs (Bachmann et al. 2003; Kool et al. 2002).

Four of the five good-quality studies (80%) reported that a better result on a performance-based measure was predictive of work participation: one study on return to work and three studies on suspension of benefits and claim closure (Table 2). Three of these good-quality studies found no effect on sustained return to work. One good-quality study found no effect on work participation in terms of sustained return to work. All thirteen studies (100%) of moderate quality reported that performance-based measures were predictive of work participation: seven studies in terms of being employed, or (sustainable) return to work, four studies on being unemployed or non-return to work, and two studies on days to benefit suspension or claim closure.

## Discussion

### Methodological considerations

Selection bias and publication bias are two concerns worthy of attention when performing a systematic review. To overcome selection bias, we used five sources of information: two databases, the American Medical Association Guide to the Evaluation of Functional Ability (Genovese and Galper 2009), references of the included papers, and relevant papers suggested by the authors. The sensitivity of our search strategy for the databases was supported by the fact that checking the references of the included studies for other potentially relevant papers resulted in only one extra study. Moreover, the authors, who have published several papers on performance-based measures, could not add other studies. Regarding publication bias, this review found three studies (Gross and Battié 2004, 2005, 2006) that reported that performance-based measures of the Workwell System were not predictive of sustained return to work in patients with chronic low back pain and with upper extremity disorders. However, more studies from the same performance-based measures (Workwell System) and in similar and different patient populations reported also on a significant predictive value for work participation in terms of return to work (Matheson et al. 2002; Vowles et al. 2004, Streibelt et al. 2009) and in terms of temporary disability suspension and claim closure (Gross et al. 2004, 2006; Gross and Battié 2005; Branton et al. 2010). Therefore, there appears to be no publication bias regarding the most described performance-based measure. To prevent publication bias resulting in a higher level of evidence due to studies of less than good quality, the evidence synthesis was formulated in such a way that regardless of the number of studies of moderate or poor quality, the qualification remained “limited”. This stringent evidence synthesis was also used to do justice to the heterogeneity of the included studies regarding not only the different performance-based tests and outcome measures for work participation but also for differences regarding chronic and non-chronic patients with MSDs in different body regions, rehabilitation and occupational setting, and treatment and non-treatment studies.

Performance-based tests can be performed in patients with severe MSDs (pain intensity 7 out of 10 or higher). Patients with severe MSDs were indeed included in the studies. Of course, regardless of pain intensity, if a person is not willing to participate, then the reliability and the validity of the results should be reconsidered. In the included studies, participants were able to perform the tests and no comments were made about unwillingness to perform a test. In test practice, however, patients' willingness

**Table 2** The extracted data from the included studies: primary author, year of publication, country, study design (cohort or intervention (retrospective or prospective)), characteristics of the population (i.e., number of employees, age and type of MSD), the treatment given, description of the reliable performance test, the confounders taken into account, the main outcome for work participation, and a summary of whether the test protocol is significantly related to work participation (yes, no, unclear)

Primary author year of publication Country	Design	Population	Treatment	Performance test	Confounders	Work participation	Predictive (yes, no, unclear)
<i>Good quality</i>							
Gross et al. (2004) Canada	Retrospective cohort 12 months	$N = 114$ patients with chronic low back pain, mean age = 41 years (SD 10), 84 men and 30 women  $N = 132$ patients with chronic low back pain, mean age = 40 years (SD 9), 94 men and 38 women	Care provided at the major Workers' Compensation Board-Alberta rehabilitation facility	Isernhagen FCE	Age, Gender, Diagnosis, Employment status, Days from injury to FCE, Pain score on disability index, Pain Visual Analog Scale, Clinician recommendation regarding fitness or readiness to work following FCE administration, Job physical demands, Pre-injury annual salary, Number of health care visits for low back pain, Number of low back claims	<i>Time to total temporary disability suspension (TTD)</i> Higher number of failed FCE tasks was related to delayed TTD (HRR = 0.91 95% CI 0.86–0.96, $n = 114$ ; HRR = 0.92 95% CI 0.87–0.97, $n = 132$ ) Higher levels on floor-to-waist lift resulted in sooner TTD (HRR = 1.48 95% CI 1.14–1.92, $n = 114$ ; HRR = 1.43 95% CI 1.09–1.89, $n = 132$ ) Pass floor-to-waist lift resulted in sooner TTD (HRR = 2.83 95% CI 1.49–5.35, $n = 114$ ; HRR = 3.74 95% CI 1.81–7.71, $n = 132$ )  <i>Time to claim closure (TCC)</i> Higher number of failed FCE tasks was related to delayed TCC (HRR = 0.92 95% CI 0.88–0.98, $n = 114$ ; HRR = 0.92 95% CI 0.870.97, $n = 132$ ) Higher levels on floor-to-waist lift resulted in sooner TCC (HRR = 1.17 95% CI 0.91–1.50, $n = 114$ ; HRR = 1.29 95% CI 1.02–1.64, $n = 132$ ) Pass floor-to-waist lift resulted in sooner TCC (HRR = 2.18 95% CI 1.26–3.77, $n = 114$ ; HRR = 4.01 95% CI 2.01–7.64, $n = 114$ )	Yes

**Table 2** continued

Primary author year of publication Country	Design	Population	Treatment	Performance test	Confounders	Work participation	Predictive (yes, no, unclear)
Gross and Battié (2004) Canada	Retrospective cohort 12 months	N = 226 patients with chronic low back pain, mean age = 41 years (SD 9), 160 men and 66 women	Care provided throughout the Workers' Compensation Board-Alberta health care provider network	Isernhagen Work System FCE	Age, Gender, Diagnosis, Employment status, Days between FCE and time to total temporary disability suspension and time to claim closure, Days from injury to FCE, Pain score on disability index, Pain Visual Analog Scale, Clinician recommendation regarding fitness or readiness to work following FCE administration, Job physical demands, Pre-injury annual salary, Number of health care visits for low back pain, Number of low back claims	<i>Sustained return-to-work (SRTW)</i> Number of failed FCE tasks was not related to SRTW (OR = 0.94 95% CI 0.87–1.02) Levels on floor-to-waist lift was not related SRTW (OR = 0.92 95% CI 0.62–1.38) Pass floor-to-waist lift was not related to SRTW (OR = 1.19 95% CI 0.46–3.05)	No
Gross and Battié (2005) Canada	Prospective cohort 12 months	N = 130 claimants with low back pain, mean age = 42 years (SD 11), 82 men and 48 women	Care provided at the Workers' Compensation Board-Alberta rehabilitation facility	Isernhagen Work System FCE	Work-related recovery expectations, Organizational policies and practices, Injury duration at time of FCE	<i>Days until suspension of time-loss benefits</i> Fewer failed tasks (HRR = 0.91 95% CI 0.87–0.96) and higher floor-to-waist lift (HRR = 1.55 95% CI 1.28–1.89) were associated with faster suspension of benefits  <i>Claim closure</i> Fewer failed tasks (HRR = 0.93 95% CI 0.89–0.98) and higher floor-to-waist lift (HRR = 1.42 95% CI 1.12–1.80) were associated with faster claim closure  <i>Sustained return-to-work (SRTW)</i> Fewer failed tasks (OR = 0.95 95% CI 0.89–1.03) and higher floor-to-waist lift (OR = 0.91 95% CI 0.63–1.33) were not significantly associated with future SRTW	Yes

Table 2 continued

Primary author year of publication Country	Design	Population	Treatment	Performance test	Confounders	Work participation	Predictive (yes, no, unclear)
Gross and Battié (2006) Canada	Prospective cohort 12 months	N = 336 claimants with upper extremity disorders, mean age = 45 years (SD 11), 239 men and 97 women	Care provided at the major Workers' Compensation Board-Alberta outpatient rehabilitation facility	Isernhagen Work System FCE	Age, Gender, Employment status, Days between FCE and time to total temporary disability suspension and time to claim closure, Days from injury to FCE, Pain score on disability index, Pain Visual Analog Scale, Clinician recommendation regarding fitness or readiness to work following FCE administration, Job physical demands, Pre-injury annual salary, Number of health care visits for the compensable condition, Total number of previous claims, Number of previous upper extremity claims	<p><i>Days until suspension of time-loss benefits</i></p> <p>Higher weight lifted on the waist-to-overhead lift (HRR = 1.51 95% CI 1.29–1.87) and on floor-to-waist lift (HRR = 1.21 95% CI 1.06–1.38) were associated with faster suspension of benefits</p> <p><i>Claim closure</i></p> <p>Higher weight lifted on the waist-to-overhead lift (HRR = 1.81 95% CI 1.49–2.20) and on the floor-to-waist lift (HRR = 1.29 95% CI 1.13–1.49) were associated with faster claim closure</p> <p><i>Sustained return-to-work (SRTW)</i></p> <p>Waist-to-overhead lift OR = 0.87 95% CI 0.60–1.27 and floor-to-waist lift (OR = 1.05 95% CI 0.70–1.17) were not associated with future SRTW</p>	Yes
Streibel et al (2009) Germany	Prospective cohort 12 months	N = 145, patients with imminent or prevailing occupational disability due to musculoskeletal disorders, mean age = 48 years (SD 9), 114 men and 31 women	Multidisciplinary rehabilitation program	Isernhagen Work System FCE	Patient's expected disability in the job, Employment status at admission, Number of weeks on sick leave prior to admission	<p><i>Non-Return-to-work (RTW)</i></p> <p>All FCE information showed significant relations to non-RTW (<math>r = 0.28-0.43</math>, <math>p &lt; 0.05</math>)</p> <p>Higher maximum functional capacity (OR = 0.22 95% CI 0.07–0.67)</p> <p>More failed test (OR = 1.10 95% CI 1.01–1.19)</p> <p>Recommended work ability &gt; 6 h a day based on actual FCE performance compared to the last job performed (OR = 0.24 95% CI 0.07–0.85)</p> <p>Using the prediction rule of more than 5 failed tests defined non RTW in the best manner: 76.9% of the patients could be predicted correctly regarding RTW in the 1-year follow-up (sensitivity: 69.7%, specificity: 80.0%).</p>	Yes



Table 2 continued

Primary author of publication Country	Design	Population	Treatment	Performance test	Confounders	Work participation	Predictive (yes, no, unclear)
<i>Moderate quality</i>							
Bachman et al. (2003) Switzerland	Prospective cohort 12 months	N = 115 patients with more than 3 months musculoskeletal pain, mean age = 42 years (SD 9), 92 men and 23 women	Structured therapy program with daily walking and strength training, and sports therapy	3-min step-test on a 30 cm high platform with a frequency of 24 steps per minute Laying on one's back and lifting a weight of 3 kg in each hand for 2 min	Nationality, Having no job at entry, Lifting more than 25 kg at work, Sick leave > 6 months	<i>Unemployed (vs. Employed)</i> Failing both performance tests (or one of these test in combination with a high pain score (9 or 10 on a scale from 0 to 10) or having more than 3 Waddell signs) resulted in a sensitivity 22% and a specificity 78% for unemployment	Yes
Branton et al. (2010) Canada	Prospective cohort 12 months	N = 147 claimants in a workers' compensation rehabilitation facility with one MSD and no occupational disease, mean age = 44 years (SD 11), 101 men and 46 women	Care provided at the Workers' Compensation Board of Alberta's rehabilitation facility	Short-form FCE (Isershagen Workwell System) <i>Trunk</i> 15-min stand, Floor-to-waist lift, 1-min crouch, 2-min kneel, 5-min rotation <i>Lower extremity</i> 15-min stand, Floor-to-waist lift, 1-min crouch, 2-min kneel, Stepladder/stairs <i>Upper extremity</i> 15-min stand, Waist-to-overhead lift, Elevated work, Crawling, Handgrip, Hand coordination	Age, Gender, Injury duration, Having a job and an employer to which to return, Occupation classification, Salary, Number of prior disability claims, Number of health care visits, Pain score on disability index, Pain Visual Analog Scale	<i>Days to benefit suspension</i> Pass all FCE test resulted in hazard ratio = 5.4 (95% CI 2.7–10.9) <i>Claim closure</i> Pass all FCE test resulted in hazard ratio = 5.8 (95% CI 3.5–9.6) for claim closure	Yes  Yes
Cheng and Cheng (2010) China	Retrospective cohort 3 months	N = 645 patients with non-specific low back pain, mean age = 42 years (SD 10), 390 men and 255 women	Care provided at designated work rehabilitation centers in Hong Kong	BTE work simulator, including torso lifting, arm lifting, high-near lifting, bi- and unilateral horizontal pushing and pulling, bilateral carrying, stooping and bending	Age, Gender, Days from injury to work, Being a breadwinner, Educational level, Compensability, Occupational categories, Physical work demand level	<i>Employed (vs. Unemployed)</i> Pass all FCE tasks resulted in positive prediction of 80% Fail all FCE tasks resulted in negative prediction of 62%	Yes
Fishbain et al. (1999) United States of America	Prospective cohort 30 months	N = 185 patients with chronic low back pain, mean age = ? years (SD ?), ? men and ? women	Chronic pain patient treatment facility	Dictionary of Occupational Titles-Residual FCE	Pain level	<i>Employed (vs. Unemployed)</i> Pass 8 DOT job measures (stooping, climbing, balancing, crouching, feeling shapes, handling left and right, lifting, carrying), and a pain level of less than 5.4, then patient had a 75% chance of being employed at 30 months (sensitivity: 75%, specificity 76%)	Yes

Table 2 continued

Primary author of publication Country	Design	Population	Treatment	Performance test	Confounders	Work participation	Predictive (yes, no, unclear)
Goutteborge et al. (2009a) Netherlands	Prospective cohort 12 months	$N = 60$ construction workers 6 weeks on sick leave due to MSDs, mean age = 42 years (SD 9), 60 men	Care provided at the largest occupational health and safety service in the Dutch construction industry	ErgoKit FCE lifting tests	No	<i>Time to sustainable return-to-work</i> Carrying and Lower lifting strength test were significant ( $p \leq 0.03$ ) although weak (HR = 1.03; HR = 1.05) predictors of the number of days on sick leave until sustainable return-to-work	Yes
Gross et al. (2006) Canada	Prospective cohort 12 months	Three cohorts ( $n = 183$ , $n = 138$ , $n = 228$ ) of claimants with low back disorders, mean age = ? years (SD ?), ? men and ? women	Care provided at the major Workers' Compensation Board-Alberta rehabilitation facility	Isernhagen Work System FCE—short form consisting of passing or failing three tests: floor-to-waist lift, crouching and standing	?	<i>Days until suspension of time-loss benefits</i> Pass three FCE tests was associated with faster suspension of benefits in all three cohorts (HRR = 4.70 95% CI 2.70–8.21; HRR = 2.86 95% CI 1.60–5.11; HRR = 1.89 95% CI 1.07–3.32)	Yes
Hazard et al. (1991) United States of America	Prospective cohort 12 months	$N = 258$ patients with chronic low back pain, mean age = 37 years (SD 9), 173 men and 85 women	Functional restoration program	Floor-to-waist lift	?	<i>Employed (vs. Unemployed)</i> Employed lifted higher weight at discharge than unemployed at 12 months (30 kg versus 27 kg, $p = 0.024$ )	Yes
Kool et al. (2002) Switzerland	Prospective cohort 12 months	$N = 99$ patients with chronic low back pain mean age = 42 years (SD 9), 84 men and 15 women	Interdisciplinary rehabilitation including strength and endurance training, exercise therapy, back school, relaxation, and passive treatment and, depending on personal needs, psychological interventions	3 min step-test on a 30 cm high platform with a frequency of 24 steps per minute Laying on one's back and lifting a weight of 3 kg in each hand for 2 min	Physical work load, Time of work, Unemployment, Nationality	<i>Non-Return-to-work</i> Failing both performance tests (or one of these tests in combination with a high pain score (9 or 10 on a scale from 0 to 10) or having more than 3 Waddell signs) resulted in a sensitivity 0.45, positive predictive value 0.97 and a specificity 0.95 for unemployment	Yes
Lechner et al. (2008) United States of America	Prospective cohort 6 months	$N = 30$ patients with injuries of the lower extremities, upper extremities or spine, mean age = 41 years (SD 11), 26 men and 4 women	Industrial rehabilitation program	Physical Work Performance Evaluation	?	<i>Return-to-work according to recommendation</i> (Percentage (%)) Full (86%) Modified (64%) Not (100%) Kappa = 0.7	Yes
Matheson et al. (2002) United States of America	Retrospective cohort 7 months	$N = 650$ clients of clinics affiliated with Isernhagen Work System FCE, mean age = 42 years (SD 10), ? men and ? women	Care provided by 25 Clinics in 16 States in the United States of America and one province in Canada affiliated with the Isernhagen Work System	Isernhagen Work System FCE, Floor-to-waist lift, Waist-to-overhead lift, Horizontal lift, Grip force	Age, Gender, Time of work	<i>Return-to-work (RTW)</i> Higher weight lifted on the floor-to-waist lift was associated with an improved likelihood of RTW ( $\chi^2 = 4.81$ , $p = 0.028$ )	Yes

**Table 2** continued

Primary author year of publication Country	Design	Population	Treatment	Performance test	Confounders	Work participation	Predictive (yes, no, unclear)
Mayer et al. (1986) United States of America	Prospective cohort 5 months	N = 66 chronic low back pain patients, mean age = 36 years (SD ?), 42 men and 24 women	Comprehensive treatment program based on functional capacity measures	Isometric and multispeed isokinetic dynamic trunk strength utilizing cybex trunk strength tester	?	<i>Return-to-work (RTW)</i> Positive change on trunk strength was associated with an improved likelihood of RTW compared to those who showed no or negative change ( $p < 0.001$ )	Yes
Strand et al. (2001) Norway	Prospective intervention study (RCT) 12 months	N = 81 patients with low back pain, mean age = 45 years (SD 10), 33 men and 48 women	Multidisciplinary rehabilitation program for 4 weeks	Five tests of physical performance: Pick-up test, Sock test, Roll-up test, Fingertip-to-floor test, lift test	?	<i>Non-Return-to-work(RTW)</i> A lower score for the pick-up test (score 0: OR = 1, score 1; OR = 4.7 95% CI 1.7–13.0, score 2,3: OR = 22.5 95% CI 2.6–196.1) and the lift test (>15 lifts: OR = 1, 1–15 lifts: OR = 5.3 95% CI 1.6–16.8, 0 lift: OR = 13.3 95% CI 3.5–50.8) was consistently related to non-RTW	Yes
Vowles et al. (2004) United States of America	Prospective cohort 6 months	N = 138, patients with chronic musculoskeletal complaints, mean age = 41 years (SD 8), 81 men and 57 women	Interdisciplinary treatment program based on a sports medicine approach to rehabilitation	Isernhagen Work System FCE, Floor-to-waist lift and Waist-to-shoulder lift	Age, Gender, Education, Pain duration, Pain anxiety symptoms, Depression, Pain intensity, Pain-related disability	<i>Non-Return-to-work</i> Lower amounts of floor-to-waist lift was correlated with less likely to return to work ( $r = -0.21, p < 0.05$ )	Yes

A distinction was made between studies with good and moderate quality  
? not reported/unknown

to perform to full capacity is seldom a matter of 100 or 0% but almost always somewhere in between. None of the studies reported to have controlled for level of effort. When looking at these tests as measures of behavior, it is plausible that physically submaximal effort has occurred, which is consistent with the definition of FCE and also observed in a systematic review by van Abbema et al. (2011).

#### Performance-based measures and work participation

The use of performance-based measures to guide decisions on work participation (pre- and periodic work screens, return-to-work, and disability claim assessments) is still under debate, at least in the Netherlands (Wind et al. 2006). This is not only due to the time-consuming nature of some of these assessments but also to its perceived limited evidence for predictive value regarding work participation. Regarding the time-consuming nature, this study also showed that a number of tests were predictive of work participation: lifting tests (Gross et al. 2004; Gross and Battié 2005, 2006; Gouttebauge et al. 2009a; Hazard et al. 1991; Matheson et al. 2002; Strand et al. 2001; Vowles et al. 2004), a 3-min step test and a lifting test (Bachman et al. 2003; Kool et al. 2002), a short-form FCE consisting of tests specific for the region of complaints (Gross and Battié 2006; Branton et al. 2010), and a trunk strength test (Mayer et al. 1986). A performance-based lifting test was most often used and appeared to be predictive of work participation in 13 of these 14 studies—especially a lifting test from floor-to-waist level in patients with chronic low back pain. An explanation might be that lifting reflects a large number of physical strenuous activities such as gripping, holding, bending, and of course lifting and lowering. Besides, van Abbema et al. (2011) showed that a “low lifting test” was not related to pain duration and showed conflicting evidence for associations with pain intensity, fear of movement/(re)injury, depression, gender, and age. Thereby, these lifting tests assess more than “just” physical components. Moreover, lifting is an important predictor of work ability in patients with MSDs (Martimo et al. 2007; Van Abbema et al. 2011). Additionally, it is plausible that “shared behaviors” occur between the tests, in which case the added value of extra tests decreases. The selection of the lifting tests appears in line with the three-step model as suggested by Gouttebauge et al. (2010) to assess physical work ability in workers with MSDs more efficiently using a limited number of tests.

Regarding its predictive value, this study showed that strong evidence exists that a number of performance-based measures are predictive of work participation for patients with chronic MSDs, irrespective whether it concerns complaints of the upper extremity, lower extremity, or low back. All patients in the included studies were considered able to

perform these reliable tests, and no comments were made that patients were unwilling to perform these tests. Of course, one has to bear in mind that the results of the performance-based measures are often used in clinical decision making regarding work participation. Moreover, patients are often not blinded to the outcome of the test itself (Reneman and Soer 2010). Gross and Battié (2004, 2006) and Gross et al. (2004) adjusted their outcome for the recommendation of the physician and Streibelt et al. (2009) for the expectation of the patient. Nevertheless, they still found that a number of performance-based tests were predictive of work participation. It seems worthwhile to establish how physicians and patients take into account the results of the performance-based tests and other instruments in their decision making regarding work participation.

Finally, the studies in this review used outcome measures in terms of future work participation and/or future non-work participation. Although not all studies presented relevant statistics, it seemed that the predictive strength of performance-based measures is higher for non-work participation than for work participation. For instance, for non-work participation, the predictive quality varied between poor (Vowles et al. 2004; Streibelt et al. 2009), moderate (Bachman et al. 2003; Streibelt et al. 2009), and good (Kool et al. 2002). For work participation, the predictive quality was mostly poor (Gross et al. 2004, 2006; Gross and Battié 2006; Gouttebauge et al. 2009a).

#### Future directions

A number of performance-based measures are predictive of work participation. Moreover, these measures differ from other relevant constructs such as pain intensity (Gross and Battié 2005; Gouttebauge et al. 2009b), self-efficacy (Reneman et al. 2008), self-reported disability (Brouwer et al. 2005; Gross and Battié 2005; Schiphorst Preuper et al. 2008; Gouttebauge et al. 2009b), and self-reported work status (Gross and Battié 2005). Also, the present study showed that potential confounders like pain intensity, work-related recovery expectations, and organizational policies and practises did not diminish the predictive validity of performance-based measures on work participation (see Table 2 “Confounders”). However, the predictive strength of performance-based measures is in general modest. Work participation is a multidimensional construct according to the ICF (WHO 2001). One cannot expect that a single instrument is able to assess such a multidimensional construct. Seen in this perspective, the conclusion of this review that the predictive validity of performance-based measures for work participation is “modest” may not be unexpected.

One way to improve the predictive strength might be combining performance- and non-performance-based

measures that assess different constructs of work participation. Bachman et al. (2003) and Kool et al. (2002) combined performance-based measures with high pain scores (9 or 10 on a scale from 0 to 10) or having more than 3 Waddell signs. Vowles et al. (2004) reported that patient age and level of depression were factors best able to predict work participation. This suggests that a combination of reliable and valid measures of different constructs might improve the ability to predict work participation. Another strategy might be the following. Seventeen of the 18 studies took place in a rehabilitation setting. Generally speaking, this means that the performance-based measures are not specific for the physical demands of the future work of a patient. One study described performance-based measures resembling the physically demanding job of construction workers (Gouttebarga et al. 2009a). One study used a job demands analysis to establish a job-specific FCE (Cheng and Cheng 2010). By doing this, the minimal performance criterion that is required to perform the job is also specified. This might overcome the misconception that a better performance is always a better predictor for work participation. This information might especially be relevant for decisions regarding work participation in patients with MSDs working in physically demanding jobs (blue collar work) (Bos et al. 2002).

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Appendix A

See Table 3.

**Table 3** Search terms used in PubMed and Embase for ‘Performance-based measures’, ‘Work participation’, and ‘Predict’

Performance-based measures
performance test, functional ability, pushing, lifting
Work participation
occupations, work, vocation, job, employment
Predict
evaluation, validity, follow-up studies, prognosis, predict, course

## Appendix B

See Table 4.

**Table 4** Criteria for the quality assessment

Study population	<p>A Inception cohort</p> <ul style="list-style-type: none"> <li>• One point if patients were identified at an early uniform point in the course of their disability e.g., uniform period after first day of sick leave</li> <li>• Zero point if it was not clear if an inception cohort was used.</li> </ul> <p>B Description of source population</p> <ul style="list-style-type: none"> <li>• One point if the source population was described in terms of place of recruitment (for example: Groningen, the Netherlands), time-period of recruitment and sampling frame of source population (for example: occupational health service, organization for social security)</li> <li>• Zero point if <math>\leq 2</math> features of source population were given.</li> </ul> <p>C Description of relevant inclusion and exclusion criteria</p> <ul style="list-style-type: none"> <li>• One point if <math>&gt;2</math> criteria were formulated</li> <li>• Zero point if <math>\leq 2</math> criteria were formulated.</li> </ul>
Follow-up	<p>D Follow-up at least 12 months</p> <ul style="list-style-type: none"> <li>• One point if the follow-up period was at least 12 months and data were provided for this moment in time.</li> </ul> <p>E Drop outs/loss to follow-up <math>&lt;20\%</math></p> <ul style="list-style-type: none"> <li>• One point if total number of drop outs/loss to follow-up <math>&lt;20\%</math> at 12 months.</li> </ul> <p>F Information completers versus loss to follow-up/drop outs</p> <ul style="list-style-type: none"> <li>• One point if sociodemographic information was presented for completers and those lost to follow-up/drop outs at baseline or no loss to follow-up/drop outs. Reasons for loss to follow-up/drop outs have to be unrelated to the outcome. Loss to follow-up/drop outs: all patients of the assembled cohort minus the number of patients at the main moment of measurement for the main outcome measure, divided by the total number of patients of the assembled cohort.</li> </ul> <p>G Prospective data collection</p> <ul style="list-style-type: none"> <li>• One point if a prospective design was used or a historical cohort when the prognostic factors were measured before the outcome was determined</li> <li>• Zero point if a historical cohort was used, considering prognostic factors at time zero which were not related to the primary research question for which the cohort was created or in case of an ambispective design.</li> </ul>
Treatment	<p>H Treatment in cohort was fully described/standardized</p> <ul style="list-style-type: none"> <li>• One point if treatment subsequent to inclusion into cohort was fully described and standardized, or in the case that no treatment was given, or if multivariate correction for treatment was performed in analysis</li> <li>• Zero point if different treatment was given and if it was not clear how the outcome was influenced by it, or if it was not clear whether any treatment was given.</li> </ul>

**Table 4** continued

Prognostic factors	<p>I Clinically relevant potential prognostic factors</p> <ul style="list-style-type: none"> <li>• One point if in addition to socio-demographic factors (age, gender) at least one other factor of the following was described at baseline: <ul style="list-style-type: none"> <li>– health-related factors (e.g., comorbidity like depression, pain anxiety symptoms, pain intensity)</li> <li>– personal factors (e.g., avoidance behavior, coping, employment, education, income)</li> <li>– external factors (e.g., physical work demands, employer characteristics, social support, health care system, social security system, social benefit).</li> </ul> </li> </ul> <p>J Standardized or valid measurements</p> <ul style="list-style-type: none"> <li>• One point if at least one of the factors of I, excluding age and gender, were reported in a standardized or valid way (for example: questionnaire, structured interview, register, patient-status of occupational/insurance physician).</li> </ul> <p>K Data presentation of most important prognostic factors</p> <ul style="list-style-type: none"> <li>• One point if frequencies, or percentages, or mean (and standard deviation/confidence interval), or median (and Inter Quartile Range) were reported for the three most important factors of I, namely age, gender and at least one other factor, for the most important follow-up measurements.</li> </ul>
Outcome	<p>L Clinically relevant outcome measures</p> <ul style="list-style-type: none"> <li>• One point if at least one of the following outcome criteria for change was reported: work disability, return to work.</li> </ul> <p>M Standardized or valid measurements</p> <ul style="list-style-type: none"> <li>• One point if one or more of the main outcome measures of L were reported in a standardized or valid way (for example: questionnaire, structured interview, registration, patient status of occupational/insurance physician).</li> </ul> <p>N Data presentation of most important outcome measures</p> <ul style="list-style-type: none"> <li>• One point if frequencies, or percentages, or mean (and standard deviation/confidence interval), or median (and Inter Quartile Range) were reported for one or more of the main outcomes for the most important follow-up measurements.</li> </ul>
Analysis	<p>O Appropriate univariate crude estimates</p> <ul style="list-style-type: none"> <li>• One point if univariate crude estimates (RR, OR, HRR) between prognostic factors separately and outcome were presented</li> <li>• Zero point if only <i>p</i>-values or wrong association values (Spearman, Pearson, sensitivity) were given, or if no tests were performed at all.</li> </ul> <p>P Appropriate multivariate analysis techniques</p> <ul style="list-style-type: none"> <li>• One point if logistic regression analysis was used, or survival analysis for dichotomous outcomes, or linear regression analysis for continuous outcomes</li> <li>• Zero point if no multivariate techniques were performed at all.</li> </ul>

## References

- Altman DG (2001) Systematic reviews of evaluations of prognostic variables. *BMJ* 323(7306):224–228
- Bachman S, Oesch PR, Kool JP, Persili S, Knüsel O (2003) Treatment of patients with chronic low back pain in a functional restoration program: work related function parameters, pain parameters and the working status after 12 months. *Phys Med Rehab Kuror* 13:263–270
- Bos J, Kuijer PPFM, Frings-Dresen MHW (2002) Definition and assessment of specific occupational demands concerning lifting, pushing and pulling based on a systematic literature search. *Occup Environ Med* 59:800–806
- Branton EN, Arnold KM, Appelt SR, Hodges MM, Battie MC, Gross DP (2010) A short-form functional capacity evaluation predicts time to recovery but not sustained return-to-work. *J Occup Rehabil* 20(3):387–393
- Brouwer S, Dijkstra PU, Gerrits EHV, Schellekens JMH, Groothoff JH, Geertzen JHB et al (2003) Intra- en inter-beoordelaarsbetrouwbaarheid ‘FIS-Belastbaarheidspatroon’ en ‘Functionele mogelijkhedenlijst’ [Intra- and interrater reliability ‘FIS-capacity profile’ and ‘Functional ability list’]. *Tijdschr Bedrijfs Verzekeringsgeneeskde* 11:360–367
- Brouwer S, Dijkstra PU, Stewart RE, Göeken LN, Groothoff JW, Geertzen JH (2005) Comparing self-report, clinical examination and functional testing in the assessment of work-related limitations in patients with chronic low back pain. *Disabil Rehabil* 27(17):999–1005
- Cheng AS, Cheng SW (2010) The predictive validity of job-specific functional capacity evaluation on the employment status of patients with nonspecific low back pain. *J Occup Environ Med* 52:719–724
- Cornelius LR, van der Klink JJ, Groothoff JW, Brouwer S (2011) Prognostic factors of long term disability due to mental disorders: a systematic review. *J Occup Rehabil* 21(2):259–274. doi: 10.1007/s10926-010-9261-5
- De Boer WE, Bruinvels DJ, Rijkenberg AM, Donceel P, Anema JR (2009) Evidence-based guidelines in the evaluation of work disability: an international survey and a comparison of quality of development. *BMC Public Health* 18(9):349–357
- De Croon EM, Sluiter JK, Nijssen TF, Dijkmans BA, Lankhorst GJ, Frings-Dresen MH (2004) Predictive factors of work disability in rheumatoid arthritis: a systematic literature review. *Ann Rheum Dis* 63(11):1362–1367
- Fishbain DA, Cutler RB, Rosomoff H, Khalil T, Abdel-Moty E, Steele-Rosomoff R (1999) Validity of the dictionary of occupational titles residual functional capacity battery. *Clin J Pain* 15:102–110
- Genovese E, Galper JS (2009) Guide to the evaluation of functional ability: how to request, interpret, and apply functional capacity evaluations. American Medical Association, USA
- Gouttebauge V, Kuijer PPFM, Wind H, Sluiter JK, Frings-Dresen MHW (2009a) Criterion-related validity of functional capacity evaluation lifting tests on future work disability risk and return to work in the construction industry. *Occup Environ Med* 66:657–663
- Gouttebauge V, Wind H, Kuijer PPFM, Sluiter JK, Frings-Dresen MHW (2009b) Construct validity of functional capacity evaluation lifting tests in construction workers on sick leave as a result of musculoskeletal disorders. *Arch Phys Med Rehabil* 90(2):302–308
- Gouttebauge V, Wind H, Kuijer PPFM, Sluiter JK, Frings-Dresen MHW (2010) How to assess physical work-ability with

- functional capacity evaluation methods in a more specific and efficient way? *Work* 37:111–115
- Gross DP, Battié MC (2004) The prognostic value of functional capacity evaluation in patients with chronic low back pain: part 2—sustained recovery. *Spine* 29(8):920–924
- Gross DP, Battié MC (2005) Functional capacity evaluation performance does not predict sustained return to work in claimants with chronic back pain. *J Occup Rehabil* 15(3):285–294
- Gross DP, Battié MC (2006) Does functional capacity evaluation predict recovery in workers' compensation claimants with upper extremity disorders? *Occup Environ Med* 63(6):404–410
- Gross DP, Battié MC, Cassidy JD (2004) The prognostic value of functional capacity evaluation in patients with chronic low back pain: part 1—timely return to work. *Spine* 29(8):914–919
- Gross DP, Battié MC, Asante A (2006) Development and validation of a short-form functional capacity evaluation for use in claimants with low back disorders. *J Occup Rehabil* 16(1):53–62
- Hazard RG, Bendix A, Fenwick JW (1991) Disability exaggeration as a predictor of functional restoration outcomes for patients with chronic low-back pain. *Spine* 16(9):1062–1067
- Innes E, Straker L (1999) Validity of work-related assessments. *Work* 13:125–152
- Kool JP, Oesch PR, de Bie RA (2002) Predictive tests for non-return to work in patients with chronic low back pain. *Eur Spine J* 11(3):258–266
- Lechner DE, Page JJ, Sheffield G (2008) Predictive validity of a functional capacity evaluation: the physical work performance evaluation. *Work* 31:21–25
- Mahmud N, Schonstein E, Schaafsma F, Lehtola MM, Fassier JB, Verbeek JH, Reneman MF (2010) Functional capacity evaluations for the prevention of occupational re-injuries in injured workers. *Cochrane Database Syst Rev* 7(7):CD007290
- Martimo KP, Varonen H, Husman K, Viikari-Juntura E (2007) Factors associated with self-assessed work ability. *Occup Med* 57(5):380–382
- Matheson LN, Isernhagen SJ, Hart DL (2002) Relationships among lifting ability, grip force, and return to work. *Phys Ther* 82:249–256
- Mayer TG, Gatchel RJ, Kishino N, Keeley J, Mayer H, Capra P, Mooney V (1986) A prospective short-term study of chronic low back pain patients utilizing novel objective functional measurement. *Pain* 25:53–68
- Reneman MF, Soer R (2010) Was predictive validity of a job-specific FCE established? *J Occup Environ Med* 52(12):1145
- Reneman MF, Geertzen JH, Groothoff JW, Brouwer S (2008) 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* 18(2):183–189
- Schaafsma F, Hulshof C, Verbeek J, Bos J, Dyserinck H, van Dijk F (2006) Developing search strategies in medline on the occupational origin of diseases. *Am J Ind Med* 49(2):127–137
- Schiphorst Preuper HR, Reneman MF, Boonstra AM, Dijkstra PU, Versteegen GJ, Geertzen JHB et al (2008) The relationship between psychological factors and performance based and self reported disability measures in patients with chronic low back pain. *Eur Spine J* 17:1448–1456
- Scholten-Peeters GG, Verhagen AP, Bekkering GE, van der Windt DA, Barnsley L, Oostendorp RA et al (2003) Prognostic factors of whiplash-associated disorders: a systematic review of prospective cohort studies. *Pain* 104(1–2):303–322
- Slebus FG, Sluiter JK, Kuijjer PPFM, Willems JHBM, Frings-Dresen MHW (2007) Work-ability evaluation: a piece of cake or a hard nut to crack? *Disabil Rehabil* 29(16):1295–1300
- Slebus FG, Kuijjer PPFM, Willems JHBM, Frings-Dresen MHW, Sluiter JK (2010) Work ability assessment in prolonged depressive illness (short report). *Occup Med* 60:307–309
- Söderberg E, Alexanderson K (2005) Sickness certificates as a basis for decisions regarding entitlement to sickness insurance benefits. *Scand J Public Health* 33(4):314–320
- Soer R, van der Schans CP, Groothoff JW, Geertzen JH, Reneman MF (2008) Towards consensus in operational definitions in functional capacity evaluation: a Delphi Survey. *J Occup Rehabil* 18(4):389–400
- Spanjer J, Krol B, Brouwer S, Popping R, Groothoff JW, van der Klink JJ (2010) Reliability and validity of the disability assessment structured interview (DASI): a tool for assessing functional limitations in claimants. *J Occup Rehabil* 20(1):33–40
- Strand LI, Ljunggren AE, Haldorsen EM, Espehaug B (2001) The impact of physical function and pain on work status at 1-year follow-up in patients with back pain. *Spine* 26:800–808
- Streibel M, Blume C, Thren K, Reneman MF, Mueller-Fahmow W (2009) Value of functional capacity evaluation information in a clinical setting for predicting return to work. *Arch Phys Med Rehabil* 90(3):429–434
- Van Abbema R, Lakke SE, Reneman MF, van der Schans CP, van Haastert CJ, Geertzen JH, Wittink H (2011) Factors associated with functional capacity test results in patients with non-specific chronic low back pain: a systematic review. *J Occup Rehabil* [Epub ahead of print]
- Van Tulder M, Furlan A, Bombardier C, Bouter L (2003) Updated method guidelines for systematic reviews in the Cochrane collaboration back review group. *Spine* 28(12):1290–1299
- Vowles KE, Gross RT, Sorrell JT (2004) Predicting work status following interdisciplinary treatment for chronic pain. *Eur J Pain* 8(4):351–358
- WHO (2001) International classification of functioning, disability and health. World Health Organization, Geneva, Switzerland
- Wind H, Gouttebauge V, Kuijjer PPFM, Frings-Dresen MHW (2005) Assessment of functional capacity evaluation of the musculoskeletal system in the context of work, daily living and sport: a systematic review. *J Occup Rehabil* 15(2):253–272
- Wind H, Gouttebauge V, Kuijjer PPFM, Frings-Dresen MHW (2006) The utility of functional capacity evaluation: the opinion of physicians and other experts in the field of return to work and disability claims. *Int Arch Occup Environ Health* 79:528–534
- Wind H, Gouttebauge V, Kuijjer PPFM, Sluiter JK, Frings-Dresen MHW (2009a) Complementary value of functional capacity evaluation for physicians in assessing the physical work ability of workers with musculoskeletal disorders. *Int Arch Occup Environ Health* 82(4):435–443
- Wind H, Gouttebauge V, Kuijjer PPFM, Sluiter JK, Frings-Dresen MHW (2009b) Effect of functional capacity evaluation on the judgment of physicians about physical work ability in the context of disability claims. *Int Arch Occup Environ Health* 82:1087–1096