

Effect of Functional Capacity Evaluation information on the judgment of physicians about physical work ability in the context of disability claims

Haije Wind · Vincent Goutteborge ·
P. Paul F. M. Kuijer · Judith K. Sluiter ·
Monique H. W. Frings-Dresen

Received: 3 June 2008 / Accepted: 29 April 2009 / Published online: 21 May 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract

Purpose To test whether Functional Capacity Evaluation (FCE) information lead insurance physicians (IPs) to change their judgment about the physical work ability of claimants with musculoskeletal disorders (MSDs).

Methods Twenty-seven IPs scored twice the physical work ability of two claimants for 12 specified activities, using a visual analogue scale. One claimant performed an FCE, the other served as a control. Outcome measure was the difference between experimental and control group in number of shifts in the physical work ability for the total of 12 specified activities.

Results The IPs changed their judgment about the work ability 141 times when using FCE information compared to 102 times when not using this information (P -value = 0.001), both in the direction of more and less ability.

Conclusions The IPs change their judgment of the physical work ability of claimants with MSDs in the context of disability claim procedures more often when FCE information is provided.

Keywords Functional Capacity Evaluation · Disability claim · Physical work ability · Musculoskeletal disorders

Introduction

The assessment of work ability in the context of long-term disability claim procedures is a complex matter, and the physicians who perform these assessments do not have many instruments to help them in this endeavour. Many people are subject to work-related illnesses or injuries, which may lead to long-term disability. In many countries, it is the statutory responsibility of physicians to assess the work ability of persons claiming disability benefit. It has been found that physicians are often unfamiliar with disability criteria and have little confidence in their ability to determine who is disabled and who is not (Zinn and Furutani 1996). The variability of impairment ratings among physicians is large and sometimes inconsistent with scientific evidence (Patel et al. 2003; Carey et al. 1988; Rainville et al. 2005).

An important category of disorders presented to physicians in the context of assessing work ability for disability claims is that of musculoskeletal disorders (MSDs). MSDs are one of the major causes of disability, and the burden of MSDs will increase in an ageing society (Brooks 2006). The direct and indirect costs of chronic disability associated with these disorders in the USA and Canada is enormous (Baldwin 2004).

There are only few instruments available to physicians engaged in the assessment of physical work ability that are both reliable and valid (Wind et al. 2005). Some questionnaires have been found to have a high level of validity and reliability. Several studies on the reliability and validity of a number of functional tests, in particular, Functional Capacity Evaluation (FCE), have been performed in recent years (Goutteborge et al. 2005, 2006; Reneman et al. 2002; Brouwer et al. 2003; Gross and Battié 2002, 2003). FCE packages are batteries of tests designed to assess the physical ability of

H. Wind (✉) · V. Goutteborge · P. P. F. M. Kuijer · J. K. Sluiter ·
M. H. W. Frings-Dresen
Academic Medical Center, Coronel Institute of Occupational
Health, University of Amsterdam, P.O. Box 22700,
1100 DE Amsterdam, The Netherlands
e-mail: h.wind@amc.uva.nl

persons—especially (ex-)workers with MSDs—to perform work-related activities (Hart et al. 1993). The physical work capacity determined by an FCE assessment can be compared to the physical job requirements of the patient's occupation or to physical job requirements in general.

In the Netherlands, the ability of a patient to return to his former job or to undertake a new job is assessed by trained, certified insurance physicians (IPs) after 24 months of sick leave. IPs rely heavily on information received from claimants in such work-ability assessments (de Bont et al. 2002; Knepper 2002). Assessing the physical work ability by IPs is like a diagnostic process, in which the work ability is the target and not the medical diagnose. As FCE information might be relevant for the judgment of the IP on the physical work ability, FCE could be added as an instrument in this process. The aim of the present study is to explore the effect of FCE information on the judgment of IPs in the context of disability claim assessments of claimants with MSDs. The research question is as follows:

- Does information derived from FCE assessments lead IPs to change their judgment of the physical work ability of claimants with MSDs?

Methods

A pre/post-test controlled experiment within subjects was used to answer the research question. To study the extent to which FCE information caused IPs to change their judgment of the physical work ability of a group of subjects with MSDs in the context of a disability claims procedure, IPs assessed the work ability twice, in an experimental group, where the claimants underwent FCE assessments after the first assessment, and in a control group, where claimants did not undergo FCE assessments. The medical Ethical Committee of the Academic Medical Center of the University of Amsterdam has approved this study.

Participants

Insurance physicians

In the Netherlands, statutory assessments of long-term disability claims are performed by IPs in the service of the Institute for Employee Benefit Schemes (UWV). The UWV is a semi-governmental organization that employs 566 IPs. One hundred IPs, selected at random, were invited to participate in the study. Fifty-four of these IPs complied with the inclusion criterion: they performed work-ability assessments on long-term disability claimants, and were prepared to take part in the study. The response rate was 54%. They all signed an informed consent form.

Claimants

Two claimants with MSDs of each IP, who were both seen in the context of a long-term disability claims procedure, were included in the study. Claimants could come either for a first disability claim assessment or for a disability re-assessment procedure, i.e. they were currently receiving a full or partial disability pension and were re-assessed pursuant to statutory requirements. Blinded for the IPs, the first claimant signed an informed consent form and underwent an FCE assessment. A second claimant served as a control. The results of the FCE assessments had no influence on the IP's statutory assessment of the claimant.

FCE assessment

The FCE assessment used in this study was the Ergo Kit (EK FCE). This FCE assessment relies on a battery of standardized tests reflecting work-related activities. A certified rater performed the 55 tests on each subject, following a standard protocol. The whole procedure took approximately 3 h. If a medical contra-indication for an FCE assessment existed, e.g. heart failure or recent surgery, the claimant was excluded from the study. Reliability of EK FCE lifting tests was found to be satisfactory in subjects with and without low-back pain (Gouttebauge et al. 2005, 2006). Other tests of the EK FCE were not studied on reliability aspects, except for the manipulation test. Content validity of the EK FCE is thought to be good, considering that the test procedures are fully described in a manual, and that they are standardized, as well as the procedure of drawing up a report. Moreover, the tested activities are work-related and are derived, like the tested activities from other FCE assessment methods, from activities mentioned in the Dictionary of Occupational Titles (DOT) (US Department of Labor 1991).

Procedure

The work ability of each claimant was assessed by the IP in accordance with the statutory rules. IPs provided information about the study to claimants with MSDs who were applying for a disability benefit or continuation of a disability benefit, and who complied with the inclusion criteria. The procedure is elucidated in Fig. 1.

The claimants were divided into two groups. The experimental group underwent an FCE assessment, while the second group served as a control group. As soon as an informed consent had been received from a claimant in the experimental group, an appointment for an FCE assessment was made with the EK team. The FCE assessment always took place after the statutory assessment of the disability claim. The claimants in the experimental group were tested

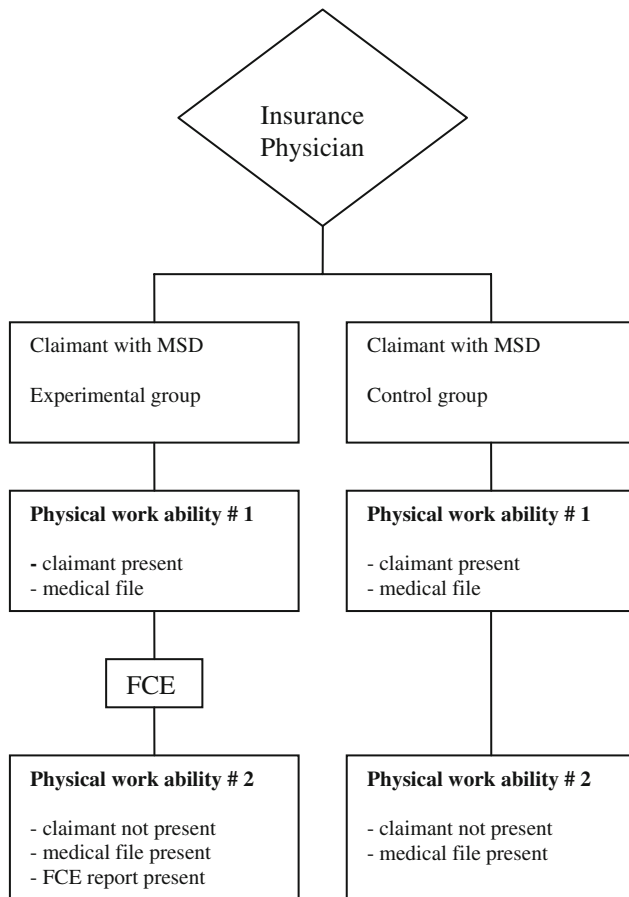


Fig. 1 Flow diagram of the procedure used in the study

in accordance with a standard FCE EK protocol by 13 certified raters at 13 locations throughout the Netherlands. A report of the EK FCE assessments performed was added to the claimant's file and a copy was sent to the claimant.

Then the physical work ability of both claimants was judged twice by the same IP in the context of long-term disability assessments. As said, half of this group of claimants underwent FCE assessments, while the other half of the claimants formed the control group. The first claimant handled by a given IP who indicated willingness to participate in the study was assigned to the group that underwent an FCE assessment, without the knowledge of the IP. The second claimant of that IP was assigned to the group that underwent no FCE assessment. In both cases, each IP assessed the work ability of each claimant twice: in the experimental group without (pre) and with (post) the information from the FCE assessment in connection with the information in the patient's file and in the control group, based only on the information in the patient's file (pre and post). At the first assessment claimants were always present, and usually the IP performed a physical examination of the claimant, although the statutory rules do not prescribe this. At the second assessment the claimants were not

present; in the latter case, the IP reviewed the claimant's case on the basis of the information available in the file. The IPs were blinded for their first judgment during the review of the claimants work ability, both in the experimental and in the control group. For the second judgment, the file of the control claimants was offered to the IP, after the FCE report had been presented to the IP with the file of the claimant that underwent the FCE assessment.

Outcomes

The characteristics of the IP, such as gender, age, years of experience with work-ability assessment and familiarity with FCE were noted, as were the characteristics of the claimants, such as gender, age and location of disorder. The IPs were asked what information was used for the first and second assessment in both groups of claimants. The time interval between the IP's first assessment and the FCE assessment for each claimant was recorded.

Visual analogue scales (VAS) were used to record the results of the assessment of the physical work ability by the IP. Although VAS scales are mostly used in studies of self-reports on pain, already in 1977 they were used in a study about the functional capacity in rheumatoid arthritis patients (Scott and Huskisson 1977). Also in other studies VAS scales were used, such as, in assessing functional disability and ability to perform physical activities (Durüoz 1996; Knop et al. 2001; Kwa et al. 1996; Post et al. 2006). Furthermore, VAS scales were used in studies on quality of life and functional scores (Krief and Huguet 2005; Matheson et al. 2006). We also performed a pilot study in which we studied the feasibility of the VAS to assess the judgment of IPs in disability claims. According to the participating IPs, the VAS was a feasible method of assessing the level of physical work ability in claimants with MSDs. The following 12 activities were rated on a VAS: walking, sitting, standing, lifting or carrying, dynamic movements of the trunk, static bending of the trunk, reaching, movements above shoulder height, kneeling or crouching and 3 activities related to hand and finger movements (repetitive hand movements, specific hand movements and pinch or grip strength). These activities were selected from several questionnaires as being valid and useful for assessment of the physical work ability of subjects with MSDs. Questionnaires were taken only for the selection of activities and not tests, because no physical tests were found to have the same clinimetric quality (Wind et al. 2005). All the selected activities are part of the FCE test, and the test results are described in the FCE report. The selected activities are also part of the functional ability list (FAL), which is the instrument currently used routinely by IPs to classify physical work ability in the context of disability claims.

The VAS score ranged from 0 to 10 and was represented by a horizontal line, length of 10 cm. The lower limit (0) was defined as complete lack of physical work ability for the activity in question compared to the situation before the claimant became disabled. The upper limit (10) was defined as no loss of physical work ability for that activity compared to the situation before onset of disability. The main outcome measure is a shift of more than 1.2 cm in the VAS score for work ability as determined for one of the 12 physical activities between the first and second assessment carried out by each IP. A change of more than 1.2 cm between the two VAS scores for a given claimant was regarded as representing an intentional change in the IP's judgment of the physical work ability. This assumption was based on the outcome of the previous mentioned unpublished feasibility study. In that study, 6 IPs assessed the physical work ability of claimants with MSDs in the context of disability claims and re-assessed the physical work ability after 2 weeks, based on the information in the claimants file. They scored the physical work ability using a VAS for the same 12 activities as used in the present study. The shift between the first and second judgment was on an average of 0.7 cm (SD 0.5). Therefore, a shift of <1.2 cm is regarded as not intentional (average + 1 SD) and thus, not clinically relevant. Moreover, in previous studies in which VAS were used, shifts between 9 and 13 mm were considered to be clinically relevant (Kelly 1998; Gallagher et al. 2001; Bodian et al. 2001; Ehrich et al. 2000). In these studies, the VAS was used on an individual level and analysed on a group level, which is also the procedure in the present study.

Data analysis

The age of the IPs and of the claimants in the two groups, and the number of years' experience the IPs had in work-ability assessment, were given as a mean value with the standard deviation. Other characteristics were noted as numbers and percentages.

A shift of more than 1.2 cm in the judgment of the IPs was considered a difference between first and second assessment. The McNemar Chi-square test for paired samples was used to test the significance of the effect of FCE information on IPs' judgment of physical work ability (Altman 1991). Tests were performed for the 12 activities as a whole, as well as for the separate activities. The Bonferroni correction was applied, as a result of which a *P*-value smaller than 0.004 was considered to be statistically significant.

The relation between the results of the FCE assessment and the shift in judgment of the IPs was first studied by classifying the results of the FCE assessment for each activity into our separate classes. These classes were: 0–33%

(class 1), 34–50% (class 2), 51–66% (class 3) and 67–100% (class 4). These classes represent the ability to perform that activity during a whole day (higher number means better abilities). In addition, some strenuous activities, such as kneeling, movements above shoulder height, dynamic movements of the trunk, and reaching, cannot be performed during the whole day according to the Ergo Kit FCE. The maximum ability for these strenuous activities is set at 66% for the whole day and these classes were recalculated starting from 0 to 66% into four classes. Lifting and grip and pinch force are presented in the FCE report in kilograms and classified into norm scores by the test leader. The outcome and classes were: not possible, very low (class 1), low (class 2), average (class 3), high and very high (class 4). Second, the outcomes of 11 out of the 12 activities (static bend work postures is not summarized in the FCE report) were compared to the first VAS score by the IP. To this end, the VAS was divided proportionally into four categories as in the FCE classification. The categories were: 0–3.3 cm (class 1), 3.4–5.0 cm (class 2), 5.1–6.6 cm (class 3) and 6.7–10 cm (class 4). The classification for each activity in the four classes based on the first VAS score of the IP and the FCE result were compared. When the classes were the same, the expectation was that the IP would not alter his score on the second VAS during the second judgment. In the case of the FCE result showing either a lower or a higher class than the IP judgment, the expectation was that the IP would lower or raise his score on the VAS for that activity during the second judgment, i.e. a shift of more than 1.2 cm. The judgment was noted as 'corresponding' in the cases of no discrepancy in classes between the first VAS score and FCE result, or when a lower FCE classification was followed by a lower classification by the IP on the second VAS score. Likewise, when the FCE classification was higher and the IP followed this classification by a raised judgment on the second VAS score, this was noted as 'corresponding'. Finally, we calculated the total numbers of corresponding outcomes. Hereby, we noted the numbers of corresponding outcomes in which the IP did not change his judgment, and the numbers of corresponding outcomes in which the IP raised or lowered his judgment on the second VAS. In all these cases, the second VAS score of the IP was in line with the result of the FCE assessment. The other cases, in which the second VAS score of the IP was not in line with the FCE assessment, were noted as 'not-corresponding'. For these 'not-corresponding' outcomes, also the direction of the difference between the expected second VAS score and the actual second VAS score was noted.

By using this method, it was possible to compare a total number of 297 activities (27 IPs and 11 activities). The scoring and analysis were performed independently by the first two authors (HW and VG). Any disagreements that remained after discussion were resolved by consulting a

third researcher. The statistical analyses were carried out using SPSS version 13.

Results

Insurance physicians

Fifty-four IPs were willing to participate in the study and signed an informed consent form, response rate of 54%. The mean age \pm standard deviation (SD) of the IPs was 47 ± 7 years, and 56% of the IPs were male. They had 15 ± 7 years of experience in work-ability assessments. Fifteen of the IPs were familiar with FCE assessments. From 27 IPs, claimants entered the study. From the other 27 IPs, no claimants were included. These two groups of IPs did not significantly differ from each other in age, gender, and work experience. Only the Chi-square test for familiarity with FCE of the IP and the participation of claimants from that IP in the study showed a significant difference, viz. that claimants from IPs who were, preceding the study, familiar with FCE participated more often than claimants from IPs who were not familiar with FCE.

In the group of IPs from whom patients were included in the study, there was no difference in the mean number of changed judgments between the first and second assessment of the physical work ability between the IPs who were familiar with FCE and the IPs who were not familiar with FCE.

Claimants

Fifty-four claimants (27 pairs from 27 IPs) indicated their willingness to participate in the study and signed an informed consent form during the study period, which extended from November 2005 to February 2007. The mean time between the disability claim assessment and the FCE assessments in the experimental group was 45 days (SD 24). The mean time between the first disability claim assessment and the re-assessment in the experimental group was 103 days (SD 43, range 39–184 days) and in the control group was 106 days (SD 99, range 16–339 days). The high SD in the latter group is primarily caused by five exceptional long time intervals of more than 184 days. The characteristics of the claimants are described in Table 1. The claimants in the experimental and the control group did not statistically differ on age, gender and the location of disorders. Seventeen claimants came for a first disability claim assessment and 37 claimants came for a disability re-assessment.

In the experimental group, the FCE report was the only new information added to the claimant's file during the second judgment of the physical work ability. In the control

Table 1 Characteristics of claimants in the experimental and control group: gender, age, and location of disorder, together with number of other sources of information used in second assessment

	Experimental group (<i>N</i> = 27)	Control group (<i>N</i> = 27)
Male (No.; percentage)	11 (41)	10 (37)
Female (No.; percentage)	16 (59)	17 (63)
Age in years (mean; standard deviation)	46 (1)	43 (2)
Location of disorder (No., %)		
Upper extremity	3 (11)	1 (4)
Lower extremity (No., %)	2 (7)	8 (30)
Back and neck (No., %)	15 (52)	9 (33)
Combination (No., %)	8 (30)	9 (33)

group, new information in two files was added, i.e. the report of a colleague IP and the letter of a treating specialist about the treatment.

The IPs could indicate the level of ability to perform the activity on the VAS scales between 0 and 10, in which a higher level stands for a better ability to perform the activity. Because of the difference in location of disorders of the claimants, there was a great variety in outcomes on the VAS scales, both in the experimental and in the control group. When a level of 5 cm or lower is taken as an indication of a more serious impairment, both in the experimental and in the control group, lifting/carrying was the activity that was judged as most limited. In the control group, the mean ability to stand was also limited. On average, the shift in judgment between the first and second assessment varied between -1.1 to 1.0 cm for the experimental group and -0.3 and 0.9 cm for the control group. The results of the first judgment (mean; SD) and the shift in judgment (mean; SD) as well as the direction of the shift, in terms of more (positive) or less (negative) physical work ability, are presented in Table 2.

Work ability judgment

Whether the provision of FCE information caused IPs to change their judgment or not of the physical work ability of claimants for the 12 specified activities by at least 1.2 cm on the VAS is presented in Table 3. In this table, a shift in judgment of more or less than 1.2 cm on the VAS for each activity during the second judgment compared to the first judgment in the experimental and in the control group is presented. The provision of FCE information caused IPs to change their judgment of the physical work ability of claimants for the totality of 12 activities significantly more often than in the control group (P -value = 0.001). No significant differences were found between the two groups for the single activities.

Table 2 Mean score and standard deviation (SD) on the VAS scores (first judgment) about the physical work ability for the 12 activities in the experimental and control group and the mean shift in judgment and SD based on the difference between the first and second judgment

	First judgment		Shift in judgment	
	Experimental	Control	Experimental	Control
Walking	7.0 (2.5)	5.6 (2.7)	0 (1.5)	0.3 (1.4)
Sitting	7.9 (2.1)	8.1 (1.7)	1.0 (1.9)	0.1 (1.3)
Standing	6.0 (2.5)	4.9 (2.9)	0.2 (2.5)	0 (1.6)
Lifting/carrying	5.0 (2.1)	4.7 (2.5)	0.1 (1.9)	0 (2.0)
Dynamic moving trunk	7.0 (2.5)	6.7 (2.8)	-0.4 (1.8)	0.4 (2.1)
Static bending trunk	6.4 (2.6)	6.5 (2.9)	-0.7 (2.6)	-0.2 (1.7)
Reaching	8.4 (1.9)	8.3 (2.0)	-0.9 (1.9)	-0.1 (1.6)
Moving above shoulder height	6.7 (3.2)	7.5 (2.7)	-0.7 (2.0)	-0.3 (1.8)
Kneeling/crouching	6.7 (3.1)	5.1 (3.2)	-1.1 (2.4)	0.9 (2.5)
Repetitive movements hands	8.3 (2.6)	8.8 (2.0)	-0.1 (1.4)	0.2 (1.8)
Specific movements hands	9.0 (2.1)	9.5 (1.2)	-0.3 (2.4)	0.2 (1.0)
Pinch/grip strength	8.9 (2.2)	9.1 (2.0)	-0.5 (1.7)	-0.3 (1.3)

Table 3 Number out of 27 insurance physicians in the experimental and in the control group with a changed or an unchanged judgment according to the cut-off point of 1.2 cm on the VAS for the total of 12

activities and for each activity separately for the second judgment compared to the first judgment

	Experimental group		Control group		McNemar χ^2 test
	Changed	Unchanged	Changed	Unchanged	
Total of activities	141	183	102	222	0.001*
Walking	13	14	9	18	0.69
Sitting	6	21	10	17	0.13
Standing	15	12	9	18	0.80
Lifting/carrying	14	13	10	17	0.15
Dynamic moving trunk	14	13	11	16	0.79
Static bending trunk	16	11	10	17	0.27
Reaching	12	15	6	21	0.15
Moving above shoulder height	14	13	9	18	0.23
Kneeling/crouching	13	14	13	14	1.00
Repetitive movements hands	7	20	7	20	1.00
Specific movements hands	8	19	3	24	0.13
Pinch/grip strength	9	18	5	22	0.29

The *P*-value of the McNemar χ^2 test for the comparison between both groups is also displayed (* significant)

The mean number of activities for which IPs changed their judgment to the above-mentioned extent in the experimental group was 4 (SD 2), compared with 5 (SD 2) in the control group. In the experimental group, 56% of the number of activities remained unchanged, for 27% of the activities the judgment about work ability was lowered and for 17% of the activities the judgment was raised. In the control group, 69% of the number of activities remained unchanged, 14% was lowered and 17% was raised.

The comparison between the results of the second VAS score and the results in the FCE report and the first VAS

score, showed that the second VAS scores were in majority in accordance with the results of the FCE assessment. In 186 out of the total 297 times (63%) the IPs scored in line with the FCE result. Of these 186 consistent scores, the IP's judgment and the FCE result were the same for 93 activities and therefore no change took place. For 56 activities, the IPs lowered their judgment of work ability in line with the FCE result that showed that the patient performed lower than the IP had judged at the first assessment. For 37 activities, the IPs raised their judgment of work ability in line with the FCE result that showed higher results than rated at

the first judgment. The judgment about walking, moving above shoulder height and dynamic moving of the trunk was most frequently lowered in line with the FCE results. For 111 activities (37%), the IPs did not follow the outcome of the FCE assessment. They maintained their judgment in 73 cases despite the result of the FCE assessment. In 23 cases the IP lowered, and in 15 cases the IP raised the work ability for that activity in contrast to the outcome of the FCE assessment. The activity pinch/grip strength showed the largest difference between expected second VAS scores and FCE results. Reaching and kneeling were the activities for which the IPs most often lowered their judgment in contrast to the FCE result. The two researchers agreed for 98% on the scoring and analysis of the comparison between the results of the second VAS score to the results in the FCE report and the first VAS score. Differences seemed random and consensus was reached regarding these differences.

Discussion

This study, based on a pre–post experimental design within subjects, evaluated the effect of FCE information on IPs' judgment of the physical work ability of disability benefit claimants with MSDs. For the totality of activities, the FCE information leads to a significant shift in the assessment of the physical work ability. Besides, for 11 out of the 12 activities the judgment of the IPs is for 62% of the activities in line with the FCE report.

The first aspect to consider is whether the VAS is a suitable means of recording physical work-ability assessments made by IPs. Many studies have shown that VAS scales are indeed a reliable means of representing judgments (Zanoli et al. 2001; Anagnostis et al. 2003). VAS scales are not only used in pain studies but also in other studies, such as assessing about the ability to perform activities or the level of disability where requested (Scott and Huskisson 1977; Durüoz 1996; Knop et al. 2001; Kwa et al. 1996; Post et al. 2006; Krief and Huguet 2005; Matheson et al. 2006). It is the statutory duty of the IP to consider all the available information about the claimant's medical situation and ability to perform various tasks, and to decide on this basis whether he is fit to work, or is fully or partially disabled. There is no reference criterion that indicates whether this judgment is accurate. One argument in favour of the use of the VAS is that it may be more sensitive to changes in assessments than the functional ability list (FAL). The FAL, rates physical work ability on an ordinal scale in 2, 3 or 4 categories, and will probably not reflect relatively small changes. We have chosen 1.2 cm as a relevant shift in judgment between the two assessments by the IP based on the results of our pilot study (average + 1 SD). Moreover,

shifts between 9 and 13 mm are considered to be clinically relevant (Kelly 1998; Gallagher et al. 2001; Bodian et al. 2001; Ehrich et al. 2000). With our choice of 12 mm we follow these values. By dichotomizing the outcome of the VAS, information is lost, namely the insight in the amount of shift in judgment of IPs. This could be a disadvantage, however, the research question was about whether IPs intentionally changed their judgment and not about the amount of change.

The second topic for consideration is the suitability of FCE as a source of supplementary information in work-ability assessments. While suggestions have been made previously to include FCE information in the disability screening process, we believe that the present study is the first one to actually measure the influence of this information on the judgment of IPs in a claim procedure (Lyth 2001; Liang et al. 1991). The study of Oesch et al. should be mentioned in this context (Oesch et al. 2006). The setting of their study was the assessment of work capacity for decisions about medical fitness for work. The use of FCE assessments in that study improved the quality of medical fitness for work certificates after rehabilitation. The focus on a rehabilitation intervention is the main difference with the present study in which the assessment of physical work ability is the main outcome and not the evaluation of a rehabilitation programme. The similarity between both studies is the influence of FCE information on the judgment of IPs for work ability. This study was designed to allow the effect of FCE information on IPs' judgment of physical work ability to be studied in its natural setting—with the proviso that, in contrast to normal diagnostic routine, the IPs taking part in the present study could not refer claimants for an FCE assessment themselves. They were unaware whether claimants were participating in the study during the first work-ability assessment. No specific direction in terms of more or less physical work ability was found for the change in judgment between the initial and the second assessment: for some activities, the assessment tended to change from a higher to a lower ability, while for other activities the change tended to be in the reverse direction. This contrasts with the findings obtained in the study of Brouwer et al. (2005), stating that the results of FCE assessment showed a higher level of physical work ability in patients with low-back pain compared to the IP judgment.

The majority of judgments (186 out of 297) of IPs about the activities was in line with the FCE results. Because in half of these cases (93) the result of the first IP judgment as scored on the VAS was in accordance with the FCE result, it could be expected that the second VAS score would likewise be in accordance with both FCE result and first VAS score. However, in the other 93 cases the FCE result was not in accordance with the first VAS score, in contrast to what was hypothesized. It implicates that there can be a

shift in judgement about the physical work ability without new information being added. This stresses the importance of using an experimental and control group in evaluating the effect of new information in disability claim assessments. In the cases that IPs altered their judgment in the direction of the FCE results, the direction of the alteration was more often (56 out of 93) towards less work ability than towards more work ability (37 out of 93). When there was a difference between the judgment of the IP and the results in the FCE report, IPs most frequently did not alter their judgments (73 out of 111). A relatively small part of the IPs (6 out of 27) are responsible for a large proportion of the differences between IP judgments and FCE report outcomes. This finding might justify the conclusion that the majority of IPs in this study are susceptible to FCE information.

Concerning the difference in number of changes between the control and experimental groups, the explanation could also be a dissimilarity between the two claimant groups. While the control group had appreciably fewer disorders of the upper extremities, the disorders at the other locations were fairly evenly spread. In the experimental group, disorders of the back and neck and combined disorders occurred most frequently. Disorders of the lower back and combined disorders might affect several physical activities, which may explain why a wide-spectrum set of tests like FCE provides information that can lead IPs to change their judgment on a range of different activities. This may also explain the small differences in mean shift in judgment between the experimental and control group. Although there seems to be an inequality regarding the location of disorders in the two groups, the size of it was not such that it has led to statistical differences between both groups and therefore, dissimilarity between the two claimant groups cannot be explained by this difference. Moreover, to overcome bias due to differences in patients and IPs on the one hand we used a within subjects design and on the other hand the shift between the first and the second judgment.

The time between the initial assessment of physical work ability by the IP and the FCE assessments (45 days on average) determines the period between the two assessments carried out by the IP on each claimant. In our opinion, this relatively long time gap does not invalidate the results of the study. The claimants who undergo the FCE assessments have been disabled for a long time. The initial assessment takes place after 2 years of sick leave—and even longer in the case of those claimants who come for re-assessment after having received disability benefit for some time. It seems implausible that their physical work ability will change considerably between the initial assessment and the FCE assessments. In addition, the long period between the two judgments has the advantage that during the FCE assessments the claimant has no recollection of the initial

assessment by the IP. The period between the first and second judgment by the IP is of less importance both in the experimental and control group, because the review is based solely on inspection of the claimant's file without any actual physical examination of the claimant. It is noteworthy that IPs in the control group altered their judgment for 102 out of 324 judgments. Only in two cases in the control group new information was presented. This emphasizes the importance of intra-rater reliability studies for the present disability assessment. As far as we know, these studies do not exist for the current practise in the Netherlands. However, the assessment of physical work ability in the context of disability claim procedures is a complex process, characterized by considerable uncertainty about the accuracy of the outcome and hence leaving ample room for changes in judgment. Information derived from FCE assessments is of a different nature than the other information that IPs use in assessing the physical work ability of workers with MSDs in disability claim procedures, which is largely anecdotal and provided by the claimant himself. The advantage of FCE information might be that it is performance-based.

This study shows that the provision of FCE information caused IPs to change their judgment of the physical work ability of disability claimants with MSDs. Physical work ability is not only important in situations of disability claim procedures, like in this study, but also in RTW and rehabilitation programmes. Although return to work of the disabled worker is the main goal in these programmes, it is not the main goal in disability claim procedures. However, it is frequently the consequence of the disability claim procedure whereby the results of the disability claim assessment are intended to be the starting point for the return to work process.

The reliability of all the tests of the EK FCE is not known. This probably has no effect on the present results because of the pre/post-test controlled experiment within IPS and that not the actual physical work ability is at stake but the effect of FCE information on the judgment of IPs. Before the EK FCE can be used as an instrument in disability claim assessments, conditions of reliability and validity have to be satisfied. Another aspect of this study needs to be taken into account. The participation of the claimants had no influence on the statutory disability claim assessment. Considering the alterations in IP's judgments, it is imaginable that after implementation of the FCE in the claim procedure the results of the FCE assessment do have consequences for the claimants. This knowledge might affect the performance of claimants in FCE assessments.

We have seen that professionals do take information from an FCE assessment seriously enough to alter their judgment about the physical work ability in disability claim assessments of workers with MSDs. There is no reason to suppose that IPs would react differently to the FCE outcome when

they would have received this information in an actual disability claim assessment. It is though imaginable that when the level of performance is below what could be expected from that patient, and the FCE results are lower than what the IP thought to be possible, that the IP will be less willing to follow the FCE results. For now, the finding that physicians take the information seriously supports the complementary value of FCE information in the assessment of disability claimants with MSDs.

What we still do not know is whether the IP assessment of work ability in the context of disability claims is improved by adding FCE information to this judgment. One of the reasons is that no referent standard exists for physical work ability in claimants who do not have worked for more than 2 years. Future studies should also focus on what specific information in the FCE report made IPs alter their judgment, or why they did not alter their judgment when the FCE results might give cause to an alteration. This and other questions, like what patients are pre-eminently fit for these types of FCE assessments according to the IPs, are of interest before implementing FCE assessments as a standard routine in disability claim assessments. The results of these studies could be used for a follow-up study about the design of FCE methods, leading to perhaps shorter, less costly and more specific assessments.

Conclusions

Provision of FCE information results in IPs to change their judgment of the physical work ability of claimants with MSDs more often in the context of disability claim procedures. Change in judgment was in majority in line with the FCE results, both in the direction of more and less physical work ability. Therefore, FCE would seem to be a valuable new instrument to support IPs in judging the physical work ability of claimants.

Acknowledgments This study was financially supported by a grant of the SIG (Stichting Instituut GAK), The Netherlands. Grant number: STIG-GV/02020021.

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.

References

Altman DG (1991) Comparing groups—categorical data. In: Altman DG (ed) Practical statistics for medical research. Chapman and

- Hall, Boca Raton, London, New York, Washington DC, pp 229–272
- Anagnostis C, Mayer TG, Gatchel RJ, Proctor TJ (2003) The million visual analog scale: its utility for predicting tertiary rehabilitation outcomes. *Spine* 28:1051–1060. doi:10.1097/00007632-200305150-00018
- Baldwin M (2004) Reducing the costs of work-related musculoskeletal disorders: targeting strategies to chronic disability cases. *J Electromyogr Kinesiol* 14:33–41. doi:10.1016/j.jelekin.2003.09.013
- Bodian CA, Freedman G, Hossain S, Eisenkraft JB, Beilin Y (2001) The visual analogue scale for pain. *Anesthesiology* 95:1356–1361. doi:10.1097/0000542-200112000-00013
- Brooks PM (2006) The burden of musculoskeletal disease—a global perspective. *Clin Rheumatol* 25:778–781. doi:10.1007/s10067-006-0240-3
- Brouwer S, Reneman MF, Dijkstra PU, Groothoff JW, Schellekens JM, Goëken LNH (2003) Test-retest reliability of the Isernhagen work systems functional capacity evaluation in patients with chronic low back pain. *J Occup Rehabil* 13:207–218. doi:10.1023/A:1026264519996
- Brouwer S, Dijkstra PU, Stewart RE, Goëken LNH, 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* 27:999–1005. doi:10.1080/09638280500052823
- Carey TS, Hadler NM, Gillings D, Stinnett S, Wallstein T (1988) Medical disability assessment of the back pain patient for the social security administration: the weighting of presenting clinical features. *J Clin Epidemiol* 41:691–697. doi:10.1016/0895-4356(88)90121-7
- de Bont A, Brink van den JC, Berendsen L, Boonk M (2002) Limited control of information for work disability evaluation. *Ned Tijdschr Geneeskd* 146:27–30. De beperkte controle van de informatie voor de arbeidsongeschiktheidsbeoordeling (in Dutch)
- Duruöz MT, Poiraudéau S, Fermanian J, Menkes CJ, Amor B, Dougados M, Revel M (1996) Development and validation of a rheumatoid hand functional disability scale that assesses functional handicap. *J Rheumatol* 23(7):1167–1172
- Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N (2000) Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. *J Rheumatol* 27(11):2635–2641
- Gallagher EJ, Liebman M, Buijter PE (2001) Prospective validation of clinically important changes in pain severity measured on a visual analogue scale. *Ann Emerg Med* 38(6):633–638. doi:10.1067/mem.2001.118863
- Gouttebauge V, Wind H, Kuijer PPFM, Sluiter JK, Frings-Dresen MHW (2005) Intra- and interrater reliability of the Ergo-Kit functional capacity evaluation method in adults without musculoskeletal complaints. *Arch Phys Med Rehabil* 86:2354–2360. doi:10.1016/j.apmr.2005.06.004
- Gouttebauge V, Wind H, Kuijer PPFM, Sluiter JK, Frings-Dresen MHW (2006) Reliability and agreement of 5 Ergo-Kit functional capacity evaluation lifting tests in subjects with low back pain. *Arch Phys Med Rehabil* 87:1365–1370. doi:10.1016/j.apmr.2006.05.028
- Gross DP, Battié MC (2002) Reliability of safe maximum lifting determinations of a functional capacity evaluation. *Phys Ther* 82:364–371
- Gross DP, Battié MC (2003) The construct validity of a kinesio-physical functional capacity evaluation administered within a workers' compensation environment. *J Occup Rehabil* 12:287–295. doi:10.1023/A:1026276822721
- Hart DL, Isernhagen SJ, Matheson LN (1993) Guidelines for functional capacity evaluation of people with medical conditions. *J Orthop Sports Phys Ther* 18:682–686

- Kelly AM (1998) Does the clinically significant difference in visual analog scale pain scores vary with gender, age, or cause of pain? *Am Emerg Med* 5:1086–1090
- Knepper S (2002) Significance of medical data in work disability evaluation. *Ned Tijdschr Geneesk* 146:6–8. De betekenis van medische gegevens bij de beoordeling van arbeidsongeschiktheid (in Dutch)
- Knop C, Oeser M, Bastian L, Lange U, Zdichavsky M, Blauth M (2001) Development and validation of the visual analogue scale (VAS) spine score. *Unfallchirurg* 104:488–497. Entwicklung und Validierung des VAS-Wirbelsäulenscores (in German). doi:10.1007/s001130170111
- Krief OP, Huguet D (2005) Shoulder pain and disability: comparison with MR findings. *AJR* 186:1234–1239. doi:10.2214/AJR.04.1766
- Kwa VIH, Limburg M, De Haan RJ (1996) The role of cognitive impairment in the quality of life after ischaemic stroke. *J Neurol* 243:599–604. doi:10.1007/BF00900948
- Liang MH, Dalroy LH, Larson MG, Partridge AJ, Abeles M, Taylor C, Fossel AH (1991) Evaluation of social security disability in claimants with rheumatic diseases. *Ann Intern Med* 115:26–31
- Lyth JR (2001) Disability management and functional capacity evaluation: a dynamic resource. *Work* 16:13–22
- Matheson LN, Melhorn JM, Mayer TG, Theodore BR, Gatchel RJ (2006) Reliability of a visual analogue version of the QuickDash. *J Bone Joint Surg* 88-A(8):1782–1787. doi:10.2106/JBJS.F.00406
- Oesch PR, Kool JP, Bachmann S, Devereaux J (2006) The influence of a functional capacity evaluation on fitness for work certificates in patients with non-specific chronic low back pain. *Work* 26:259–271
- Patel B, Buschbacher R, Crawford J (2003) National variability in permanent partial impairment ratings. *Am J Phys Med Rehabil* 82:302–306. doi:10.1097/00002060-200304000-00009
- Post RB, Keizer HJE, Leferink VJM, Van der Sluis CK (2006) Functional outcome 5 years after non-operative treatment of type A spinal fractures. *Eur Spine J* 15:472–478. doi:10.1007/s00586-005-0887-5
- Rainville J, Pransky G, Indahl A, Mayer EK (2005) The physician as disability advisor for patients with musculoskeletal complaints. *Spine* 30:2579–2584. doi:10.1097/01.brs.0000186589.69382.1d
- Reneman MF, Dijkstra PU, Westmaas M, Goëken LNH (2002) Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. *J Occup Rehabil* 12:269–276. doi:10.1023/A:1020274624791
- Scott PJ, Huskisson EC (1977) Measurement of functional capacity with visual analogue scales. *Rheumatol Rehabil* 16(4):257–259
- United States Department of Labor (1991) Dictionary of occupational titles, 4th edn. US Government Printing Office. Washington, DC
- Wind H, Gouttebauge V, Kuijer PPFM, Sluiter JK, Frings-Dresen MHW (2005) Assessment of functional capacity of the musculoskeletal system in the context of work, daily living, and sport: a systematic review. *J Occup Rehabil* 15:253–272. doi:10.1007/s10926-005-1223-y
- Zanoli G, Stromqvist B, Jonsson B (2001) Visual analog scales for interpretation of back and leg pain intensity in patients operated for degenerative lumbar spine disorders. *Spine* 26:2375–2380. doi:10.1097/00007632-200111010-00015
- Zinn W, Furutani N (1996) Physician perspectives on the ethical aspects of disability determination. *J Gen Intern Med* 11:525–532. doi:10.1007/BF02599599