Criterion-related validity of functional capacity evaluation lifting tests on future work disability risk and return to work in the construction industry

Gouttebarge, V.; Kuijer, P.P.F.M.; Wind, H.; van Duivenbooden, C.; Sluiter, J.K.; Frings-Dresen, M.H.W.

DOI
10.1136/oem.2008.042903

Publication date
2009

Document Version
Final published version

Published in
Occupational and Environmental Medicine

Criterion-related validity of functional capacity evaluation lifting tests on future work disability risk and return to work in the construction industry

V Gouttebarge, P P F M Kuijer, H Wind, et al.

Occup Environ Med 2009 66: 657-663 originally published online May 24, 2009
doi: 10.1136/oem.2008.042903

Criterion-related validity of functional capacity evaluation lifting tests on future work disability risk and return to work in the construction industry

V Gouttebarge,[1] P P F M Kuijer,[1] H Wind,[1] C van Duivenbooden,[2] J K Sluiter,[1] M H W Frings-Dresen[1]

[1] Academic Medical Center, University of Amsterdam, Department: Coronel Institute of Occupational Health, Amsterdam, The Netherlands; [2] Arbouw, Ceintuurbaan 2, Harderwijk, The Netherlands

Correspondence to: Vincent Gouttebarge, Academic Medical Center, University of Amsterdam, Coronel Institute of Occupational Health, PO Box 22700, 1100 DE Amsterdam, The Netherlands; v.gouttebarge@amc.nl

Accepted 26 February 2009
Published Online First 24 May 2009

ABSTRACT

Objectives: To assess the criterion-related validity of the five Ergo-Kit (EK) functional capacity evaluation (FCE) lifting tests in construction workers on sick leave due to musculoskeletal disorders (MSDs).

Methods: Six weeks, 6 months and 1 year after the first sick leave day due to MSDs, construction workers underwent two isometric and three dynamic EK FCE lifting tests, and completed the Instrument for Disability Risk (IDR) for future work disability risk. Concurrent and predictive validity were assessed by the associations between the scores of the EK FCE lifting tests and the IDR outcomes (Pearson Correlation coefficients (r) and associated proportions of variance (PV) and area under receiver operating characteristic curve (AUC)). Predictive validity of the EK FCE lifting tests on the total number of days on sick leave until full durable return to work (RTW) was also evaluated (Cox regression analysis).

Results: Concurrent validity with future work disability risk was poor for the two isometric EK FCE lifting tests ($-0.15 \leqslant r \leqslant 0.04$) and moderate for the three dynamic EK FCE lifting tests ($-0.47 \leqslant r \leqslant -0.31$). Only the carrying lifting strength test showed moderate and acceptable predictive validity on future work disability risk ($r = -0.39$; AUC = 0.72). Cox regression analyses revealed that two out of the five EK FCE lifting tests predicted durable RTW significantly, but only weakly.

Conclusions: Criterion-related validity with future work disability risk was poor for the two isometric EK lifting tests and moderate for the three dynamic lifting tests, especially the carrying lifting strength test. Predictive validity on durable RTW was poor, although weakly significant in two dynamic EK FCE tests, of which one was the carrying lifting strength test.

What this paper adds

- Concurrent validity with future work disability risk is poor for the isometric Ergo-Kit functional capacity evaluation (EK FCE) lifting tests and moderate for the dynamic EK FCE lifting tests for workers in the construction industry.
- The carrying lifting strength test presents a moderate level of predictive validity on future work disability risk in construction workers on sick leave due to musculoskeletal disorders (MSDs).
- The carrying and lower lifting strength tests show significant but very weak predictive validity on durable return to work in construction workers on sick leave due to MSDs.
- As the carrying lifting strength test presents only moderate criterion-related validity with future work disability risk in the construction industry, this test cannot be used solo by occupational professionals working in health and safety services for jobs exposed to manual material handling such as firefighters, garbage collectors and movers.
- It seems necessary to evaluate whether information coming from the carrying lifting strength test, in combination with information provided by anamnesis, clinical examination and self-reported questionnaires, could have an added value for the judgement and decision-making process of occupational and insurance physicians in their assessment of physical work ability.

Edwin Smith's Surgical Papyrus, roughly written in 1700 BC, is the world's earliest known document that acknowledged signs of work-related musculoskeletal disorders (MSDs) in construction workers, which arose from the imposing Egyptian pyramids construction projects.[1] In construction industries all over the world, MSDs are the primary reason for long-term sickness absence and related work disability, and the incidence of MSDs is strongly associated with manual material handling, especially lifting.[2 3] In 2005, for the construction industry of the USA, overexertion when lifting caused 42% of the work-related MSDs with associated days away from work, while lifting was responsible for 21% of work compensation due to MSDs.[4 5]

In order to reduce sick leave and work compensation costs due to MSDs, occupational and insurance physicians need to assess the physical ability or inability to work ("physical work ability") of an injured worker, in particular, the ability to perform safe lifting among construction workers. In the Netherlands, physicians working either in return to work (RTW) or disability claims do not possess many instruments to assess physical work ability but they have a positive view on the utility of complementary information derived from the functional capacity evaluation (FCE).[6] FCE was designed to offer comprehensive performance-based assessments to measure the current physical work ability of workers with or without MSDs.[7–9] The Ergo-Kit (EK) is an FCE method that relies on a battery of standardised tests that assess work-related activities, such as standing, walking, lifting, carrying and reaching.[10] As lifting ability is one of the most important components of heavy physical work, especially in the construction industry, the EK FCE lifting tests in particular could be seen as useful tools to provide relevant information for the assessment of physical work ability in the construction industry.

Information provided by any clinical instrument cannot be trusted and used if its measurement quality, that is, reproducibility and validity have not been positively evaluated. After that the EK FCE lifting tests were found reproducible in participants with and without MSDs, validity should now be evaluated.[11 12] Validation of instruments is challenging and is the main topic of interest when it comes to the evaluation of the quality of an instrument's measurements, that is, its clinimetric properties. Without the assessment of validity, it cannot be claimed that what is purportedly being measured is what is truly being measured.[7 13 14] Therefore, before one can administer the EK FCE lifting tests in occupational healthcare settings in the construction industry, the validity of the tests must be assessed. Among the different validity types, criterion-related validity is especially relevant for functional assessments.[15–17] Criterion-related validity, subdivided into concurrent and predictive validity, describes how the evaluated test relates to another existing instrument measuring the same concept (or partially the same concept), ideally a gold standard showed to be reproducible and valid.[15–17] Concurrent validity refers to the relation between the two instruments concurrently, meaning nearly at the same time, while predictive validity refers to the relation between two instruments, where the existing instrument is measured later on time.[15–17] When no gold standard is available, as in the case of the assessment of physical work ability,[18 19] a well-grounded reference test (also referred to as a silver standard) measuring an affiliated relevant concept and accepted in practice is commonly used as an alternative.[20 21] In the Netherlands, the Instrument for Disability Risk (IDR) is an established and accepted instrument for identifying construction workers at risk for work disability over a 2-year period.[22 23] The IDR is a questionnaire assessing the status of four risk factors of future work disability in construction industry: age, sickness absence, musculoskeletal complaints and work ability (based on the Work Ability Index).[24] The IDR is appropriate as a reference test because it is a well-grounded instrument that is accepted and used in the construction industry and an instrument that measures future work disability risk, an affiliated concept of physical work ability. Furthermore, the EK FCE tests were found to provide occupational professionals with complementary information that was useful when they made judgements of workers' physical work ability to aid the RTW process.[25–27] Hence, the time until durable RTW (ie, the number of days on sick leave until full durable RTW) seems another relevant affiliated concept that could be used in a validity study of the EK FCE lifting tests.

Thus, the three aims of the present study were to assess (1) the concurrent validity of the EK FCE lifting tests and future work disability risk in construction workers; (2) the predictive validity of the EK FCE lifting tests on future work disability risk in construction workers on sick leave due to MSDs; and (3) the predictive validity of the EK lifting tests on time until durable RTW in construction workers on sick leave due to MSDs.

MATERIALS AND METHODS

Design

A longitudinal within-subject design with a 1-year follow-up period was conducted.

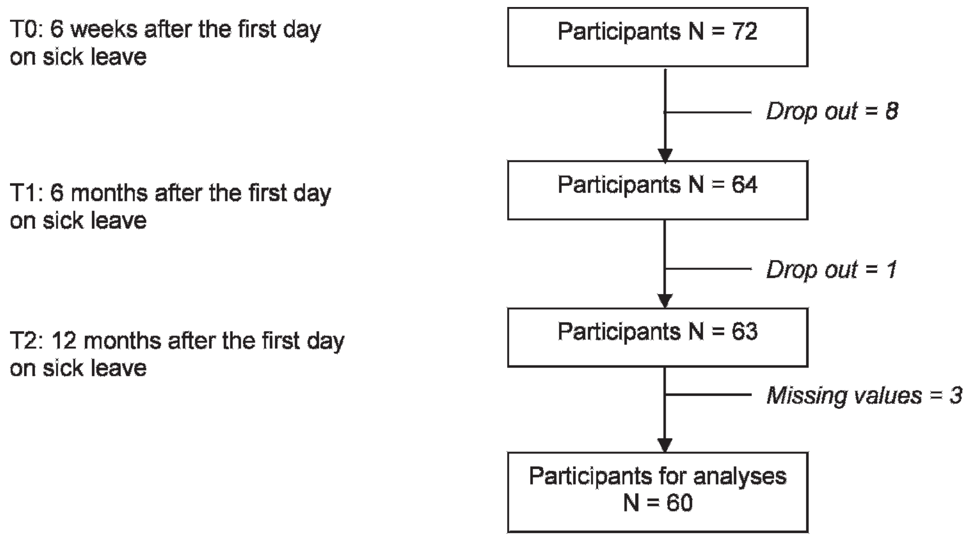
Participants and recruitment procedures

From a nationwide list obtained from the largest occupational health and safety service in the Dutch construction sector, construction workers on sick leave for 3–4 weeks were contacted by phone by the first author. If a worker expressed interest to participate, detailed written information on the study procedure was sent and signed statements of informed consent were obtained. A sample size calculation was performed for our research questions ((1) 2-tailed t test with $\alpha = 0.05$ and power $= 0.80$; (2) confidence level of 0.95, correlation coefficient set at 0.50 and limit at 0.30), and indicated that a minimum of 50 subjects were required at the end of our 1-year follow-up period. To take dropouts during follow-up into account, we strived to include 75 participants at baseline, based on the following inclusion criteria: (1) performing heavy physical work in the construction industry; (2) age between 18 and 55 years; and (3) on sick leave for the last 6 weeks (SD 1 week) due to MSDs. The registration of performing heavy physical work was carried out by the occupational physicians from the patient file and based on the job specific classification provided by the Dutch construction industry organisation Arbouw in terms of physical work demands. Participants were free to withdraw from the study at any time.

Table 1 Ergo-Kit (EK) test descriptions and outcomes[10]

EK tests	Description	Outcome
Back-torso lift test (Btlt)	Use of a "back and leg dynamometer" fixed on a platform, a chain and a handle. Handle is set at patella height for Btlt and at elbow height for Slt. Maximal pulling during 4 s, 2 tries per test.	
Shoulder lift test (Slt)		Maximal isometric lift capacity (kg)
Carrying lifting strength test (Clst)	Use of a stand with two vertically adjustable shelves, a box with different weights and a step (20 cm). Following standardised procedure, weight is added to the box (2.5, 5, 7.5 or 10 kg), depending on the subject's coordination in the task, subject's perception of the weight of the box, and subject's complaints. 4–6 carries 5 m for Clst, 4–6 lifts from knuckle height to step for Llst and 4–6 lifts from knuckle to acromion height for Ulst.	
Lower lifting strength test (Llst)		Maximal safe weight for lifting (kg)
Upper lifting strength test (Ulst)		

Figure 1 Flowchart of the participants and loss to follow-up.



EK FCE lifting tests

Two isometric lifting tests, back-torso lift test (Btlt) and shoulder lift test (Slt), and three dynamic lifting tests, carrying lifting strength test (Clst), lower lifting strength test (Llst) and upper lifting strength test (Ulst), were selected for this study. Table 1 presents the descriptions and outcomes of the five EK FCE lifting tests. According to the standardised procedures,[10] subjects did not receive their own results after their assessment on the EK FCE lifting tests, which was also guaranteed by the use of variable weights (added throughout the test procedures) that are not recognisable by the subjects. The assessment of the five EK FCE lifting tests by certified raters took approximately 30 minutes.

Instrument for disability risk

In the present study, the IDR was selected as the reference test. In the Netherlands, the construction industry has developed this construction-industry-specific instrument to identify workers at risk for work disability over a 2-year period.[22 23] Assessing four risk factors for work disability in the construction industry (ie, age, work ability, sickness absence and musculoskeletal complaints), the IDR score is calculated from responses to nine questions (see Appendix). The IDR provides two types of outcomes: (1) a binomial outcome, having an increased risk for work disability or not; and (2) a risk of work disability (percentage). A percentage of 38 or more has been chosen in expert consensus meetings as the cut-off point for an increased risk of work disability in the years to come.[22 23]

Return to work

In the present study, time to durable RTW was defined as the duration of work absenteeism due to MSDs in calendar days from 6 weeks after the first day on sick leave until the first day of returning fully to the worker's own work or other work for a period of at least 4 weeks.[28] As RTW was registered throughout the 1-year follow-up period by the occupational health and safety service in the construction industry, number of days until durable RTW was established by medical records.

Study procedures

Six weeks (baseline, t_0), 6 months (t_1) and 1 year (t_2) after the first sick leave day, subjects were assessed on five EK FCE lifting tests and were asked to complete the IDR, during the occupational physician consultation at t_0, t_1 and t_2 at home. To guarantee that the time interval between the two assessments (ie, the EK FCE lifting tests and IDR) was as short as possible, participants who did not return the IDR questionnaire within 3 days after their assessment on the EK FCE lifting tests were again contacted by phone. This study was performed in accordance with the Helsinki Declaration (1964) and received approval from the Medical Ethics Committee of the Academic Medical Center in Amsterdam, the Netherlands.

Data analyses

Only the participants included at baseline and who completed the three assessments without any missing value(s) during the 1-year follow-up period were included in the statistical analyses

Table 2 Means, standard deviations and ranges of age, height, bodyweight and outcomes of the Ergo-Kit functional capacity evaluation lifting tests and Instrument for Disability Risk (IDR) at t_0 (baseline), t_1 and t_2 (n = 60)

	Baseline t_0			t_1			t_2		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Age (years)	42	9	18–55						
Height (cm)	182	8	168–198						
Bodyweight (kg)	86	13	59–125						
Back-torso lift test (kg)	86.4	32.2	21.0–152.5	94.2	31.2	25.7–175.5	91.9	32.4	32.5–170.0
Shoulder lift test (kg)	43.9	18.5	2.5–88.5	51.4	19.0	10.0–93.3	51.4	17.7	17.5–85.0
Carrying lifting strength test (kg)	36.2	13.3	10.0–75.0	35.2	10.1	15.0–75.0	33.5	7.9	17.5–47.5
Lower lifting strength test (kg)	32.8	13.2	0.0–75.0	34.4	11.6	2.5–75.0	33.8	8.4	12.5–47.5
Upper lifting strength test (kg)	22.1	8.6	5.0–50.0	24.8	6.4	10.0–40.0	24.8	6.5	10.0–40.0
IDR (%)	42.7	16.6	9.0–65.0	36.8	22.6	9.0–79.0	34.5	24.5	9.0–79.0

Table 3 Correlations (r) and proportions of variance (PV) between the outcomes of the Instrument for Disability Risk (IDR) and the Ergo-Kit functional capacity evaluation lifting tests at t_0 (baseline), t_1 and t_2 (n = 60)

Variables	r (t_0)	PV (t_0)	r (t_1)	PV (t_1)	r (t_2)	PV (t_2)
Back-torso lift test	−0.15	2.25	−0.15	2.25	0.02	0.04
Shoulder lift test	−0.07	0.49	−0.05	0.25	0.04	0.16
Carrying lifting strength test	−0.17	2.89	−0.47**	22.09	−0.33**	10.89
Lower lifting strength test	−0.17	2.89	−0.36**	12.96	−0.31*	9.61
Upper lifting strength test	−0.12	1.44	−0.42**	17.64	−0.23	5.29

*p<0.05; **p<0.01.

Table 4 Predictive validity of the Ergo-Kit functional capacity evaluation lifting tests at t_0 (baseline) on the Instrument for Disability Risk (IDR) at t_1 and t_2: correlations (r), proportions of variance (PV) and area under the curve (AUC) (n = 60)

	IDR at t_1			IDR at t_2		
Variables	r	PV	AUC	r	PV	AUC
Back-torso lift test at t_0	−0.14	1.96	0.50	−0.10	1.00	0.53
Shoulder lift test at t_0	−0.04	0.16	0.45	−0.09	0.81	0.52
Carrying lifting strength test at t_0	−0.39**	15.21	0.72	−0.32*	10.24	0.68
Lower lifting strength test at t_0	−0.29*	8.41	0.67	−0.19	3.61	0.60
Upper lifting strength test at t_0	−0.19	3.61	0.62	−0.22	4.84	0.58

*p<0.05; **p<0.01.

(SPSS V.14.0 for Windows). At t_0, t_1 and t_2, descriptive statistics were calculated for each of the EK FCE lifting tests and the IDR outcome.

Concurrent validity was determined by assessing the relationship at t_0, t_1 and t_2 between the five EK FCE lifting tests scores and the IDR outcomes. Predictive validity of the EK FCE lifting tests on future work disability risk was evaluated by assessing the associations between the five EK FCE lifting tests scores at t_0 and the IDR outcomes at both t_1 and t_2. For both concurrent and predictive validity, Pearson's correlation coefficients (r) and associated proportions of variance ($PV = 100 \times r^2$) were calculated.[29] For concurrent and predictive validity, correlations >0.50 are considered as good, 0.30–0.50 as moderate, and <0.30 as poor.[29] Furthermore, the ability of the EK FCE lifting tests at t_0 (baseline) to predict the outcomes of the IDR at t_1 and t_2 was measured using the area under the receiver operating characteristic curve (AUC).[17 30] Therefore, we used the cut-off point of 38% set by Dutch experts for a high risk for work disability.[22 23] The AUC can be interpreted as follows: $0.7 \leqslant AUC \leqslant 0.8$ as acceptable, $0.8 < AUC \leqslant 0.9$ as excellent, and $AUC > 0.9$ as outstanding.[31]

Predictive validity of the EK FCE lifting tests at t_0 on time to durable RTW was evaluated by conducting Cox proportional hazards regression analysis[30] in order to identify whether the EK FCE lifting tests (independent variables or covariates) were separate predictive factors for time to durable RTW (dependent variable). Interpretation was based on the level of significance (p values) and on the hazard ratio (HR), that can be interpreted in a similar manner to the odds ratio.[32]

RESULTS

Participant's characteristics

Seventy-two construction workers were included in the study, from which 60 (83%) completed the three assessments without any missing information during the 1-year follow-up period. From the 72 subjects included, eight dropped out (11%) for the following assessment 6 months later, an additional one dropped out (1%) between the second and third assessment, and three participants (4%) had missing value(s) on the EK FCE tests during the 1-year follow-up period (fig 1). Compared with the participants who remained in our study, the 12 participants excluded in the analyses because of missing value(s) were slightly younger (mean age of 37 years old; p>0.05) and stayed longer on sick leave (169 days; p>0.05). At t_0, all 60 participants were on sick leave due to MSDs, with the upper extremity MSDs accounting for 17% of the main diagnoses, the lower extremity for 28%, the back for 30%, and a combination of MSDs for the remaining 25%. Participants were assessed on the EK FCE lifting tests in 15 different locations in the Netherlands, depending on their home addresses. Among the participants, carpentry was the most frequent occupation (37%). From the 60 sick listed participants, 47 returned to work 6 months later (t_1; 78%) and 51 returned 1 year later (t_2; 85%). Nine participants were still on sick leave after the 1-year follow-up period. The baseline characteristics of the 60 participants are presented in table 2.

Concurrent validity

Table 2 presents the outcomes at t_0, t_1 and t_2 of the five EK FCE lifting tests and the IDR. The correlations and related PV between the five EK FCE lifting tests scores and the IDR outcomes are presented in table 3. Weak associations were found at t_0 between scores of the five EK FCE lifting tests and the IDR outcomes ($-0.17 \leqslant r \leqslant 0.07$). At t_1 and t_2, the associations between the scores of the two isometric EK FCE lifting tests and the IDR outcomes were also weak. Moderate associations ($p<0.01$) at t_1 and/or t_2 were found between the outcomes of the three dynamic EK FCE lifting tests and the IDR, with an upper value of $r = -0.47$ ($p<0.01$) for the association at t_1 between the carrying lifting strength test and the IDR.

Predictive validity IDR

The correlations between the five EK FCE lifting tests scores at t_0 and the IDR outcomes at t_1 and t_2, the PV and AUC are all presented in table 4. One dynamic EK FCE lifting test, the carrying lifting strength test, had a moderate correlation with the IDR (−0.39 at t_1 and −0.32 at t_2), showing a moderate predictive validity on future work disability risk. In addition, an acceptable predictive ability of the carrying lifting strength test for IDR outcomes was confirmed by an AUC value of 0.72 at t_1. Weak associations ($-0.29 \leqslant r \leqslant -0.04$) were found between the scores on the other four out of the five EK FCE lifting tests and the IDR outcomes.

Table 5 Predictive validity of the Ergo-Kit functional capacity evaluation lifting tests on return to work (number of days on sick leave until return to work): Cox proportional hazards regression analysis (n = 60)

Variables	Estimate of regression coefficient	Hazard ratio (95% CI)
Back-torso lift test	0.008	1.00 (1.00 to 1.02)
Shoulder lift test	0.009	1.00 (1.00 to 1.02)
Carrying lifting strength test	0.030	1.03 (1.00 to 1.05)
Lower lifting strength test	0.045	1.05 (1.02 to 1.07)
Upper lifting strength test	0.027	1.03 (1.00 to 1.06)

Predictive validity durable RTW

Table 5 shows the results of the Cox proportional hazards regression analyses, revealing that two out of the EK FCE lifting tests (carrying and lower lifting strength tests) were significant ($p \leqslant 0.03$) although weak (HR = 1.03; HR = 1.05) predictors of the number of days on sick leave until durable RTW. The HR of the carrying lifting strength test can be interpreted as follows: a change in this test outcome of one or five units (kg) means nearly 3% and 16% (1.03^5), respectively, more chance for the event durable RTW.

DISCUSSION

The aim of this study was to evaluate the criterion-related (concurrent and predictive) validity of two isometric and three dynamic EK FCE lifting tests in construction workers who were on sick leave because of MSDs. Concurrent validity between the two isometric EK FCE lifting tests and the IDR, the reference test for future work disability risk, was found to be poor while concurrent validity between the three dynamic EK FCE lifting tests and the IDR was moderate. One dynamic EK FCE lifting test, the carrying lifting strength test, showed a moderate level of predictive validity on the IDR. The predictive validity of the other four out of the five EK FCE lifting tests on the IDR was poor. Furthermore, the predictive validity of the five EK FCE lifting tests on durable RTW (ie, number of days on sick leave until full durable RTW) could not be established. Overall, the criterion-related validity with future work disability risk was poor for the two isometric EK lifting tests and moderate for the three dynamic lifting tests, especially the carrying lifting strength test. The predictive validity on durable RTW was poor, although weakly significant in two dynamic EK FCE tests.

Conducting a validity study of an instrument inevitably entails some methodological and procedural considerations. First, the study population chosen in any validity study is essential in order to validate correctly the evaluated instrument or test. As FCEs strive to report physical work ability, the selection of construction workers in our validity study seems relevant as construction workers perform jobs particularly exposed to manual material handling, which is strongly related to the occurrence of MSDs and sick leave.[2 3] In addition, among all manual material handling activities performed in the jobs of the construction industry, lifting is definitely a dominant activity.[2 3] With regards to the loss to follow-up in this study, the main reason for dropout was that participants did not find any time or motivation to be assessed again on the EK FCE lifting tests because they already returned to work, or they suffered from an MSD that did not allow them to be assessed again with the EK FCE lifting tests according to our study timetable. One reason for participants remaining in the study could be the financial reward they received: in addition to the travelling expenses, they received €50 per assessment (€150 for the whole study period) and were entered into a lottery for a traveller's cheque with a value of €1000. All in all, the use of construction workers and the few dropouts are strengths of the present study, as it seemed appropriate to select such a population in the validation process of the EK FCE lifting tests.

Second, with regards to the reference test selected, we could put forth reasons to justify the selection of the IDR for our criterion-related validity study. The concept that is measured by the EK FCE lifting tests is physical work ability. As no gold standard is available for physical work ability,[18 19] a well-grounded instrument, accepted and used in practice, measuring physical work ability or an affiliated relevant concept had to be selected. Considering the use of construction workers as participants in our study, especially in the context of the Dutch construction industry, and the need to have a test that was affiliated with the concept of physical work ability, our search for a reference test resulted in the IDR. The IDR is intended to be used in the case of construction workers to assess future work disability risk due to MSDs, which seems an acceptable affiliated concept for (physical) work ability. Furthermore, within the nine questions of the IDR, physical work ability is specifically addressed. It also indirectly assesses the respondent's lifting ability, as this activity is one of the most important for jobs in the construction industry. Thus, as no gold standard is available for physical work ability, the IDR appears as a rational reference test to assess the criterion-related validity of the EK FCE lifting tests in the construction industry.

Finally, to establish relationships between the outcome(s) of evaluated instrument(s) (ie, independent variable(s)) and the outcome(s) of interest (ie, dependent variable(s)) during a follow-up period, an observational prospective longitudinal within-subject design was used to assess criterion-related validity, which seemed the best-suited research design, even if observational studies provide weaker empirical evidence than experimental studies.[33] Also, a strength of our design was the possibility to assess concurrent validity between the EK FCE lifting tests and the IDR at three different moments within 1 year, allowing a comparison over time of the concurrent validity and the evaluation of the durability of validity in a "changing" population, that is, workers recovering from MSDs and sickness absence. In the present study, the concurrent validity level, particularly of the dynamic EK FCE lifting tests, with future work disability risk changed and improved substantially between baseline and either the second or third assessments, which could be explained by the change in the covariance between EK FCE lifting test scores and IDR outcomes.

As FCEs have been recently a topic of interest,[34] our results can be compared with other criterion-related validity studies. As in the present study, some authors tried to assess the concurrent validity of FCE tests with self-reported questionnaires measuring disability-related concepts. Similar to our results, Reneman *et al* and Gross and Battié found low to moderate levels of concurrent validity between the Isernhagen Work Systems (IWS) FCE lifting and carrying tests, and several self-reported disability questionnaires (Roland-Morris Disability questionnaire, Oswestry Back Pain Disability Scale, Quebec Back Pain Disability Scale, Pain Disability Index and pain visual analogue scale).[35–37] From this perspective, it can be suggested that physical work ability, measured through the IWS or EK FCE, and self-reported questionnaires measuring disability-related concepts, can be seen as affiliated or related to each other. However, a study comparing concurrently the IWS and EK FCE lifting tests showed that both the FCEs produced different results, meaning that the IWS and EK cannot be used interchangeably.[38] In our study, only one out of the five EK FCE lifting tests, the carrying lifting strength test, could predict future work disability risk moderately and durable RTW significantly but weakly (HR = 1.03): a change in this test outcome of 1 or 5 km means nearly 3% and 16% (1.03^5), respectively, more chance for the event durable RTW. Gross and Battié also found that a better FCE lifting ability was only weakly related to RTW (either faster or safer); in addition these studies were conducted in a work disability claim context.[39–41] An explanation for our results may be the quick evolution over time of the nature of the participants' MSDs. Another possibility may be that the expectation that FCEs, which measure current physical work ability, have prognostic value on future work-related concepts could be just too ambitious and not realistic.

From the results of this study, the carrying lifting strength test, gives information that is moderately valid for the construction industry. Lasting only a few seconds, the two isometric tests appeared less relevant for the work demands in the construction industry and this may partially explain the results of our study. Compared with the other two dynamic EK FCE lifting tests, the carrying lifting strength test reflects the largest number of activities such as gripping, lifting, bending, carrying and walking. Walking is especially responsible for the longer time needed for the assessment and seems relevant to the physical work demands of construction workers, which could be an explanation for its moderate association with future work disability risk. However, the carrying lifting strength test cannot be used solely for jobs exposed to manual material handling by occupational professionals working in health and safety services as it presents only a moderate evidence of criterion-related validity. In addition, the construct validity of this carrying lifting strength test was not supported.[42]

Thus, it seems necessary to first evaluate whether the information from the carrying lifting strength test, in combination with information provided by anamnesis, clinical examination and self-reported questionnaires, could have an added value for the judgement-making process of occupational professionals in their assessment of physical work ability. If so, and only if so, the assessment of the carrying lifting strength test could provide occupational professionals with useful and valid information on several activities in a rapid and efficient way, and it would also enhance the practicality of using FCEs to some extent. FCE practicality is known to be limited as FCEs are often generic and time-consuming, and has been logically a topic of interest for some authors in order to increase the FCE practicality by selecting functional tests from the full FCE for specific defined jobs.[43–45] However, further research on shorter and more specific FCEs is still needed to support their application in occupational medicine for heavy physical jobs such as construction workers, firefighters or garbage collectors. Furthermore, gathering information from different sources such as self-reported questionnaires, clinical examination and performance-based testing (ie, FCEs), could lead to an optimal assessment of current physical work ability, and should be subject to further research.

CONCLUSION

Criterion-related validity with future work disability risk in sick-listed construction workers with MSDs was poor for the two isometric EK lifting tests and moderate for the three dynamic lifting tests, with the highest value for the carrying lifting strength test. Predictive validity on durable RTW was poor, although weakly significant in two dynamic EK FCE tests, of which one was the carrying lifting strength test.

Acknowledgements: The authors would like to thank Marco van de Velde (''ArboDuo'') for his support in this study. We are also grateful to all certified raters who assessed the EK FCE lifting tests and to all participants.

Funding: This study was financially supported by ''Stichting Arbouw'' and ''Stichting Instituut GAK''.

Competing interests: None declared.

Ethics approval: This study was performed in accordance with the Helsinki Declaration (1964) and received approval from the Medical Ethics Committee of the Academic Medical Center in Amsterdam, the Netherlands.

Patient consent: Obtained.

CvD contributed to the introduction and discussion, but was not involved in the data analyses and results description of this article.

Provenance and peer review: Not commissioned; externally peer reviewed.

REFERENCES

1. **Feldman RP,** Goodrich JT. The Edwin Smith Surgical Papyrus. *Child's Nerv Syst* 1999;**15**:281–4.
2. **Kuiper JI,** Burdorf A, Verbeek JHAM, *et al*. Epidemiologic evidence on manual materials handling as risk factor for back disorders: a systematic review. *Int J Indus Ergon* 1999;**24**:389–404.
3. **van der Molen HF,** Sluiter JK, Hulshof CTJ, *et al*. Effectiveness of measures and implementation strategies in reducing physical work demands due to manual handling at work. *Scand J Work Environ Health* 2005;**31**(Suppl 2):75–87.
4. **Schneider SP.** Musculoskeletal injuries in construction: a review of the literature. *Appl Occup Environ Hyg* 2001;**16**:1056–64.
5. **The Centre for Construction Research and Training**. *The construction chart book - The U.S. Construction Industry and its Workers. Report*. Silver Spring: The Center for Construction Research and Training, 2007.
6. **Wind H,** Gouttebarge V, Kuijer PPFM, *et al*. 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* 2006;**79**:505–13.
7. **King PM,** Tuckwell N, Barrett TE. A critical review of functional capacity evaluations. *Physical Therapy* 1998;**78**:852–66.
8. **Strong S.** Functional capacity evaluation: the good, the bad and the ugly. *OT Now* 2002;**Jan/Feb**:5–9.
9. **Vasudevan SV.** Role of functional capacity assessment in disability evaluation. *J Back Musculoskeletal Rehabil* 1996;**6**:237–48.
10. **Ergo Control**. *Ergo-kit for functional capacity evaluation: User manual*. Enschede, the Netherlands: Ergo Control, 2002. [Ergo-kit Functionele Capaciteit Evaluatie. Handleiding, in Dutch.]
11. **Gouttebarge V,** Wind H, Kuijer PPFM, *et al*. Intra- and interrater reliability of the Ergo-Kit FCE method in adults without musculoskeletal complaints. *Arch Phys Med Rehabil* 2005;**86**:2354–60.
12. **Gouttebarge V,** Wind H, Kuijer PPFM, *et al*. Reliability and agreement of 5 Ergo-Kit Functional Capacity Evaluation lifting tests in subjects with low back pain. *Arch Phys Med Rehabil* 2006;**87**:1365–70.
13. **Matheson LN,** Mooney V, Grant JE, *et al*. Standardized evaluation of work capacity. *J Back Musculoskeletal Rehabil* 1996;**6**:249–64.
14. **Mooney V.** Functional capacity evaluation. *Orthopedics* 2002;**25**:1094–9.
15. **Innes E,** Straker L. Validity of work-related assessments. *Work* 1999;**13**:125–52.
16. **Streiner DL,** Norman GR. *Health measurement scales: a practical guide to their development and use*. New York: Oxford University Press, 2003.
17. **Portney LG,** Watkins MP. *Foundations of clinical research: Applications to practice*. New Jersey: Prentice-Hall, 2000.
18. **Wind H,** Gouttebarge V, Kuijer PPFM, *et al*. Assessment of functional capacity of the musculoskeletal system in the context of work, daily living, and sport: a systematic review. *J Occup Rehabil* 2005;**15**:253–72.
19. **Gouttebarge V,** Wind H, Kuijer PPFM, *et al*. Reliability and validity of Functional Capacity Evaluation methods: a systematic review with reference to Blankenship system, Ergos work simulator, Ergo-Kit and Isernhagen work system. *Int Arch Occup Environ Health* 2004;**77**:527–37.
20. **Baxter P.** Invalid measurement validity. *Dev Med Child Neurol* 2005;**47**:291.
21. **Lambert HC,** Gisel EG, Wood-Dauphinee S. The functional assessment of dysphagia: psychometric standards. *Phys Occup Ther Geriatr* 2002;**19**:1–14.
22. **Burdof A,** Frings-Dresen MHW. Ontwikkeling van de WAO-indicator- Gezondheidskundige identificatie van werknemers met een sterk verhoogd risico op arbeidsongeschiktheid, 2002. [Development of the Instrument for Disability Risk. In Dutch.]
23. **Burdof A,** Frings-Dresen MHW, van Duivenbooden JC, *et al*. Development of a decision model to identify workers at risk of long-term disability in the construction industry. *Scand J Work Environ Health* 2005;**31**(Suppl 2):31–6.
24. **Tuomo K,** Ilmarinen J, Jahkola A, *et al*. Tyokykyindeksi [Work ability index]. Helsinki: Tyoterveyslaitos, 1997.
25. **Wind H,** Gouttebarge V, Kuijer PPFM, *et al*. The complementary value of functional capacity evaluation for physicians in assessing the physical work-ability of workers with musculoskeletal disorders. *Int Arch Occup Environ Health* 2008 Oct 9 [Epub ahead of print].
26. **Wind H,** Gouttebarge V, Kuijer PPFM, *et al*. The effect of functional capacity evaluation assessment on the judgment of insurance physicians of the physical work-ability in the context of disability claims. *Int Arch Occup Environ Health* 2009 May 21 [Epub ahead of print].
27. **Wind H,** Gouttebarge V, Kuijer PPFM, *et al*. 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* 2006;**79**:505–13.
28. **Heymans MW,** de Vet HCW, Knol DL, *et al*. Workers' beliefs and expectations affect return to work over 12 months. *J Occup Rehabil* 2006;**16**:685–95.
29. **Cohen J.** *Statistical power analysis for the behavioral sciences*. Hillsdale, New Jersey: Lawrence Earlbaum Associates, 1988.
30. **Altman DG.** *Practical statistics for medical research*. London: Chapman and Hall, 1991.
31. **Hosmer DW,** Lemeshow S. *Applied logistic regression*. New York: John Wiley & Sons, 2000.
32. **Petrie A,** Sabin S. *Medical statistics at a glance*. Oxford: Blackwell Publishing, 2005.
33. **Concato J.** Observational versus experimental studies: what's the evidence for a hierarchy? *NeuroRx* 2004;**1**:341–7.
34. **Innes Ev**. Reliability and validity of functional capacity evaluations: an update. *Int J Disabil Manag Res* 2006;**1**:135–48.

35. **Reneman MF,** Jorritsma W, Schellekens JMH, *et al*. Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic low back pain. *J Occup Rehabil* 2002;**12**:119–29.
36. **Gross DP,** Battié MC. Construct validity of a kinesiophysical functional capacity evaluation administrated within a worker's compensation environment. *J Occup Rehabil* 2003;**13**:287–95.
37. **Gross DP,** Battié MC. Factors influencing results of functional capacity evaluations in workers' compensation claimant with low back pain. *Phys Ther* 2005;**85**:315–22.
38. **IJmker S,** Gerrits EHJ, Reneman MF. Upper lifting performance of healthy young adults in functional capacity evaluations: a comparison of two protocols. *J Occup Rehabil* 2003;**13**:297–305.
39. **Gross DP,** Battié MC, Cassidy JD. The prognostic value of functional capacity evaluation in patients with chronic low back pain: part 1: timely return to work. *Spine* 2004;**29**:914–19.
40. **Gross DP,** Battié MC, Cassidy JD. The prognostic value of functional capacity evaluation in patients with chronic low back pain: part 2: sustained recovery. *Spine* 2004;**29**:920–4.
41. **Gross DP,** Battié MC. Functional capacity evaluation performance does not predict sustained return to work in claimant with chronic low back pain. *J Occup Rehabil* 2005;**15**:285–94.
42. **Gouttebarge V,** Wind H, Kuijer PPFM, *et al*. Construct validity of functional capacity evaluation lifting tests in construction workers on sick leave due to musculoskeletal disorders. *Arch Phys Med Rehabil* 2009;**90**:302–8
43. **Frings-Dresen MHW,** Sluiter JK. Development of a job-specific FCE protocol: the work demands of hospital nurses as an example. *J Occup Rehabil* 2003;**13**:233–48.
44. **Gross DP,** Battié MC, Asante A. Development and validation of a short-form functional capacity evaluation for use in claimants with low back disorders. *J Occup Rehabil* 2006;**16**:53–62.
45. **Gross DP,** Battié MC, Asante AK. Evaluation of a short-form functional capacity evaluation: less may be best. *J Occup Rehabil* 2007;**17**:422–35.

APPENDIX: INSTRUMENT FOR DISABILITY RISK.[20 21]

1. How would you rate your current work ability compared with the lifetime best, where 0 is ''not able to work'' and 10 is ''best work ability ever''? (0–10 scale.)
2. How would you rate your current work ability with regard to the physical work demands of your job? (5-points Likert scale.)
3. How would you rate your current work ability with regard to the psychological work demands of your job? (5-points Likert scale.)
4. From the following list of 51 diseases, give the number of current diseases you have that were diagnosed by a physician and/or diagnosed by yourself. (Number of diseases.)
5. Give your estimation of work impairment due to diseases. (1–6 scale.)
6. How many days were you on sick leave during the past year? (1–5 scale.)
7. From your own judgement, do you think you will be working in your own job in two years? (3-points Likert scale.)

8a. Lately, do you enjoy your daily life? (5-points Likert scale.)
8b. Lately, have you been active and fit? (5-points Likert scale.)
8c. Lately, have you had trust in the future? (5-points Likert scale.)
9a. Do you have regular neck stiffness or pain? (Binominal.)
9b. Do you have regular stiffness or pain in the upper extremity? (Binominal.)
9c. Do you have regular back stiffness or pain? (Binominal.)
9d. Do you have regular stiffness or pain in the lower extremity? (Binominal.)