**ORIGINAL ARTICLE**

# Do patient characteristics affect the predictive validity of Functional Capacity Evaluations?

David Bühne[1] · Torsten Alles[1] · Christian Hetzel[1] · Marco Streibelt[2] · Ingo Froböse[3]

## Abstract

**Purpose** The aim of this study was to identify patient-related characteristics that affect the predictive validity of the FCE assessment ELA.

**Methods** A prospective multicenter study was conducted on 303 patients with musculoskeletal disorders (MSD) recruited from eleven rehabilitation centers. The ELA-based estimation of the participants' ability to cope with physical work demands was considered valid if RTW was paired with a positive ELA outcome ($\geq$ moderate) as well as if non-RTW was accompanied by a negative ELA outcome (rather or very poor). In the remaining cases, the ELA result was judged as non-valid. To reduce the risk of false conclusions, the rating was performed inversely in participants that (1) reported severe limitations regarding their productivity at work, (2) attributed RTW to a change in job resp. a reduction of their physical work demands and in those that (3) attributed non-RTW to non-physical reintegration barriers only. Using questionnaires, 28 patient-related characteristics were collected. Logistic regression models were calculated to identify characteristics that affected the predictive validity of ELA.

**Results** ELA was considered valid in 208 of 303 (69%) participants. A moderate and strong pain-related disability at work were associated with a 0.15-fold (95% confidence interval (95% CI) 0.05–0.46), respectively, 0.19-fold (95% CI 0.05–0.72) chance for a valid outcome. In addition, a negative influence was found in participants that reported psychosocial distress (odds ratio (OR) 0.35; 95% CI 0.15–0.82), a native language different from the national language (OR 0.16; 95% CI 0.05–0.56) as well as in those that expected to return to work, but not within one month (OR 0.17; 95% CI 0.06–0.46). Further variables—including age, employment status, fear-avoidance beliefs and the level of physical work demands—did not affect the predictive validity of ELA.

**Conclusions** The results suggest that the predictive validity of ELA is primarily limited by patients that report a moderate or strong pain-related disability at work, psychosocial distress as well as the expectation to return to work, but not within one month. Furthermore, a negative influence can be assumed for language barriers.

✉ David Bühne
buehne@iqpr.de

1 Institute for Quality Assurance in Prevention and Rehabilitation at the German Sport University Cologne (IQPR), Eupener Str. 70, 50933 Cologne, Germany

2 Department of Rehabilitation, German Federal Pension Insurance, Ruhrstr. 2, 10704 Berlin, Germany

3 Institute of Movement Therapy and Movement-Oriented Prevention and Rehabilitation at the German Sport University Cologne, Am Sportpark Müngersdorf 6, 50933 Cologne, Germany

## Background

Due to a large number of days on sick leave, rehabilitation measures and pension applications that are related to musculoskeletal disorders (MSD) (Bevan 2015; Murray et al. 2015; Deutsche Rentenversicherung Bund (DRV) 2019a), the question of whether an individual's physical capacity meets the physical demands of work frequently arises in the context of occupational and rehabilitation medicine. To promote a valid assessment of the ability to safely perform activities required at work, Functional Capacity Evaluation (FCE) procedures became part of common clinical practice (Ansuategui Echeita et al. 2019). These assessments

are defined as "an evaluation of capacity of activities that is used to make recommendations for participation in work while considering the person's body functions and structures, environmental factors, personal factors and health status" (Soer et al. 2008).

Various studies have been conducted to examine the measurement properties of FCE procedures. The most recently published reviews (Baets et al. 2018; Kuijer et al. 2012) indicate that there is some evidence for the predictive validity of FCE information with regard to work participation. Overall, the predictive validity seems to be modest. However, the comparability of these results is limited due to differences in participants (e.g. chronic low back pain (Gross and Battié 2004) vs. upper extremity disorders (Gross and Battié 2006)), the FCE information that was used to predict work participation (e.g. number of failed tests (Branton et al. 2010) vs. carrying and lifting tests (Gouttebarge et al. 2009)) as well as the operationalization of work participation (e.g. time until the suspension of time-loss benefits (Gross and Battié 2006) vs. return to work (RTW) (Streibelt et al. 2009)). In addition, several attempts were made to evaluate criteria in terms of their impact on FCE outcomes like the maximal lifting capacity. For example, Gross and Battié (2005) found that "younger, male subjects who reported lower levels of perceived disability and pain intensity lifted at higher levels". The last three criteria were confirmed in a multicountry study (Ansuategui Echeita et al. 2019).

These results are consistent with the assumption that female sex, pain as well as the perceived disability are negatively related to the ability to handle heavy loads. Therefore, they do not allow conclusions to be drawn regarding whether sex, pain intensity and the perceived disability affect the validity of FCE outcomes. In fact, little is known about criteria that are associated with lower resp. increased risk of assessing the ability to perform work-related activities or the ability to meet physical demands of work incorrectly (Bühne et al. 2020). Such findings would be of great benefit for both FCE developers and health care professionals. According to the first-mentioned, they might provide the opportunity to estimate the utility or necessity of modifications and further developments. Health care providers as well as other professionals that use FCE information for therapy planning or e.g. RTW decisions could derive information about the robustness of FCE results. To our knowledge, only two studies addressed this issue so far. The data were collected in patients with nonspecific low back pain (Cheng and Cheng 2010) and distal radius fracture, respectively (Cheng and Cheng 2011). Within the first study, a negative impact was found for compensable injury and in the case of high physical work demands. In addition, "a 1-day increase on the days from injury to FCE would reduce it [the predictive validity] by 51.7%" (Cheng and Cheng 2010). However, due to methodological uncertainties, the results cannot be transferred.

Thus, the aim of the study was to identify patient-related criteria that affect the predictive validity of FCE-based estimations of the ability to cope with physical demands of work.

## Methods

### Study design and setting

A multicenter cohort study was conducted in eleven German rehabilitation centers offering work-related medical rehabilitation (WMR) programs (Bethge et al. 2019) for patients with various chronic or injury-related MSD. These WMR interventions last three to four weeks and are performed both in- and outpatient. In addition to ordinary rehabilitation, work-related components with a total extent of at least eleven hours are provided. In patients with MSD, FCE procedures constitute a key component of job-related diagnostics. The additional therapeutic components include work-related psychosocial groups, intensified social counselling and work-related functional capacity training (Bethge et al. 2019).

Data were collected by questionnaires at admission, discharge and the 3-month follow-up between August 2018 and August 2020. The study was reviewed by the Ethics Committee of the German Sport University Cologne (118/2017) and funded by the German Pension Insurance.

## Participants

Eligible patients with MSD and high risk of non-RTW, e.g. due to a low self-efficacy with regard to work demands or numerous days of sick leave, were referred to rehabilitation by the German Pension Insurance (ten centers), respectively, the German Social Accident Insurance (one center). The specific criteria were dependent on the screening methods used. In the case of the very common procedure SIMBO (Screening Instrument for Identification of a Demand for Medical-Vocational Oriented Rehabilitation), for example, individuals are considered to have a high risk for non-RTW if they expect not to be able to work in their profession in the near future and if the number of sick days within the last 12 months exceeds 26 weeks (Streibelt et al. 2007). The inclusion criteria were the presence of musculoskeletal disorders, absence of contraindications such as acute injuries or serious comorbidities (e.g. cardiac insufficiency, angina pectoris) and signed consent to participate. Patients were excluded if they reported significant changes in their physical work demands or their physical capacity, that occurred after discharge and were unrelated to the individual physical work ability (e.g. loss of physical capacity due to an injury

during leisure time, job change that was not initiated as a consequence of excessive physical demands).

## Measures

### RTW outcome

A successful return to work was considered if participants were working full- ($\geq 35$ h/week) or part-time (15-34 h/week) at the 3-month follow-up and did not reach or exceed a total duration of 1.5 weeks of sick leave due to MSD following discharge from the WMR program (Streibelt et al. 2009; Bühne et al. 2020). In both cases, the information was taken from the written post-survey questionnaire.

### FCE-based estimation of the ability to cope with the physical demands of work

The ELA method (German: Einschätzung körperlicher Leistungsfähigkeiten bei arbeitsbezogenen Aktivitäten) was used at discharge to estimate the ability to cope with the physical demands of work. ELA is a customized short-form FCE containing 24 work-related activity tests such as kneeling, torso rotation and lifting (Bühne et al. 2020). Prior to the test, an interview was carried out between a health care professional and the patient to identify work demands that might exceed the current physical capacity. Based on this, individually relevant activity tests were selected. In ELA, these tests are to be terminated by the assessor in case of a clear insecurity in the execution, dizziness or comparable observations that indicate a potential health risk. Further termination criteria are a heart rate increase above 85% of the estimated maximum (220 minus age) and achieving a weight in lifting and carrying that exceeds the work demands as well as 60% of the patient's body weight. In addition, subjects are instructed to abort the test if they perceive to have reached their current maximum performance level.

After performing the activity tests, the FCE-certified assessors compared the demonstrated functional capacity to the individual job demands for each activity. These extrapolations to the entire working day were then summarized in an estimation of the patient's ability to cope with physical work demands using a 5-point scale (very poor via moderate to very good). The characteristics of the procedure are described in detail elsewhere (Bühne et al. 2020). The assessments were guided by 26 physical therapists, eight occupational therapists and ten sports scientists with an average professional experience of 11.3 year (0–27 years). The average FCE experience was 2.0 years (0–15 years).

The predictive validity of ELA was demonstrated in workers with MSD. A positive ELA outcome (moderate to very good) was associated with a sixfold higher RTW chance after adjusting for the current pain intensity, the Work Ability Score and further patient-reported data (Bühne et al. 2020). According to our knowledge, reliability has not yet been evaluated.

### Predictive validity of the ELA outcome

Based on the results of a previous study (Bühne et al. 2020), the ELA-based estimation of the patient's ability to cope with physical work demands was considered valid if RTW was paired with a positive FCE outcome (moderate to very good) as well as if non-RTW was accompanied by a negative FCE outcome (rather or very bad). In the remaining cases, the ELA result was judged as non-valid.

To avoid false conclusions regarding the predictive validity of the ELA result, adjustments were made in three conditions. The rating was inversely converted in participants that reported a change in job or a reduction of their physical work demands which enabled them to return to work successfully. The rating was also converted when participants reported "having great difficulty" or "being unable" to perform the usual tasks at their workplace without restrictions (Müller et al. 2014). In these cases, despite RTW, participants were considered not to have the required functional capacity to cope with the demand of work defined at discharge. Moreover, positive ratings of the predictive validity of ELA were converted into negative and vice versa, when N-RTW was attributed to non-physical reintegration barriers only (e.g. unsuccessful job search).

### Patient characteristics

Using questionnaires, the characteristics shown in Table 1 were collected. These procedures resp. variables were chosen as they are recommended for use in WMR or have been shown to be meaningful with regard to FCE resp. RTW outcomes.

In addition, FCE assessors recorded the level of physical work demands according to the REFA (Association for Work Design, Business Organization and Business Development) classification (DRV 2019b) as well as the injury location (upper extremities, lower extremities, trunk or multiple regions). ICD-10 diagnoses were extracted from the medical report. Regional unemployment rates were assigned based on the counties of residence.

### Statistical methods

Descriptive statistics were used to report sample characteristics. Differences between dropouts and the final sample were analyzed using the Chi-square test, the Mann and Whitney's $U$ test and the $T$ test. Logistic regression models were calculated to identify patient characteristics that affect the

**Table 1** Patient-reported characteristics collected to identify aspects that affect the predictive validity of FCE information

| Sociodemographic | Primarily health-related patient-reports | Primarily work-related patient-reports |
|---|---|---|
| Admission<br>  Age, gender, marital status<br>  Migration background<br>  Native language (German vs. others)<br>  Highest professional qualification ($\leq$ apprenticeship vs. higher)<br>  Months of employment within the last year<br>  Employment status (unemployed/ < 5 years/ $\geq$ 5 years)<br>  Type of work requirements (primarily physically demanding vs. others) | Admission<br>  Psychosocial distress[#] (0–3 vs. $\geq$ 4 points) (Schmidt et al. 2019)<br>  Duration of illness in months (0–6/7–12/13–24/ > 24)<br>  Degree of disablement<br>  Duration of sickness absence within the previous 12 months ($\leq$ 26 weeks vs. > 26 weeks)<br>Discharge<br>  Current pain intensity (NRS; 0–3/4–7/8–10)<br>  General health perception (Ware and Sherbourne 1992) ($\leq$ rather bad/good/ $\geq$ very good)<br>  Tampa Scale for Kinesiophobia (Rusu et al. 2014) ($\leq$ 29/30–35/ $\geq$ 36) (Lakke et al. 2015) | Admission<br>  Work motivation ($\leq$ low/moderate/ $\geq$ very strong)<br>  Presence of workloads (0 to max. 5)[##]<br>Discharge<br>  Health-related disability at work (Streibelt et al. 2009)<br>  Expected duration until RTW (Löffler et al. 2009)<br>  Pension request<br>  Work Ability Score (Ilmarinen et al. 1997) (NRS; 0–3/4–7/8–10)<br>  Pain-related disability at work (NRS; 0–3/4–7/8–10) (Tait et al. 1990)<br>  Attitudes towards work (Ranft et al. 2009) ( < 14/14–22/23–30) |

Cut-off values used in the present analysis are shown in parentheses

[#]Ultra-Short Screening Questionnaire

[##] Workloads: monotonous work, shift work, time pressure resp. pressure to perform, environmental influences and control by superiors

predictive validity of ELA. Continuous health- and work-related characteristics were thereby split according to established cut-offs, respectively, transformed into three-level variables to simulate u-shaped associations (Table 1). These $U$-shaped associations were to be expected, since valid estimations of the ability to cope with physical demands of work are more likely in cases in which patients report very favorable (e.g. high level of perceived work ability, no pain) or very unfavorable RTW-conditions (e.g. low level of perceived work ability, severe pain). Initially, adjusted only for the ELA result, all 28 characteristics were examined separately. Those characteristics that did not exceed the cut-off ($p < 0.1$) were then included into further models collectively, while differentiating (1) sociodemographic characteristics, (2) primarily health-related and (3) primarily work-related patient reports as shown in Table 1. The final model was calculated based on those characteristics that still remained below the cut-off ($p < 0.1$) within these three multiple models. All models were adjusted for the ELA result. Given a small number of participants whose ability to cope with physical work demands was classified as very good resp. very bad, the ELA outcome was converted into a three-point scale ($\geq$ rather good vs. moderate vs. $\leq$ rather poor). Nagelkerke´s $R^2$ and the area under the receiver operating characteristic curve (AUC) were calculated to assess the model fit. The discrimination was considered acceptable if AUC $\geq 0.7$ and excellent if AUC $\geq 0.8$ (Hosmer et al. 2013). Odds ratios (OR) with 95% confidence intervals (CIs) were calculated as effect estimates. Statistical tests were regarded as significant if the $p$ value was less than 0.05. Before calculating multiple models, the presence of multicollinearity ($r \geq 0.7$; variance inflation factor $\geq 10$) was evaluated. A complete case analysis was performed. All calculations were carried out with SPSS 27. Expecting a drop-out rate of 35%, a proportion of valid ELA outcomes of 75% and a multiple model with about ten dependent variables, the recruitment of 650 participants was targeted a priori to avoid overfitting.

# Results

## Flow of participants and sample characteristics

The recruitment was stopped prematurely in April 2020 due to the COVID-19 pandemic. By this time, 550 participants had already been recruited. The following analyses are based on 303 participants (the flow of participants is illustrated in Fig. 1). Overall, the proportion of missing values was 1.3%.

Compared to the participants who did not participate in the follow-up survey, the final sample was older (51.0 vs. 44.5 years; $p < 0.001$), rather female (48.2% vs. 36.1%; $p = 0.011$) and employed at admission (84.8% vs. 76.3%; $p = 0.022$) as well as more likely to be in a stable relationship (73.6% vs. 62.4%; $p = 0.010$).

The number of cases per clinic ranged from 3 to 65 (mean = 27.5). The characteristics of the resulting sample are presented in Table 2 and the appendix. The most prevalent ICD-10 diagnoses were M54.4 (Lumbago with sciatica; $n = 40$), M53.1 (Cervicobrachial syndrome; $n = 28$) and M51.2 (Other specified intervertebral disc displacement; $n = 18$).
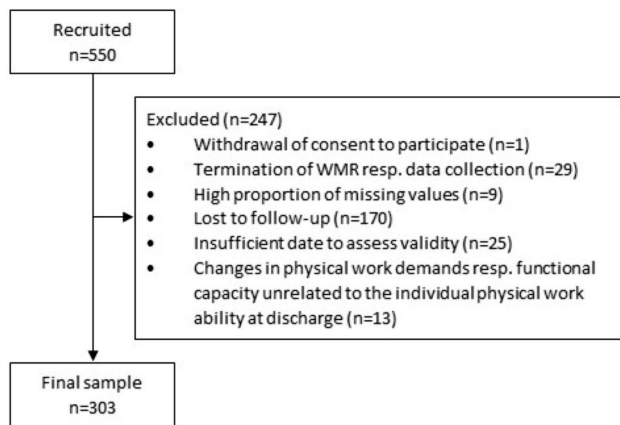
**Fig. 1** Flow of participants

**Table 2** Patients characteristics (*N* = 303)

| | |
|---|---|
| Age in years, mean (SD) | 51.0 (8.7) |
| Female, *n* (%) | 146 (48.2) |
| Married/stable relationship, *n* (%) | 223 (73.6) |
| Work status, *n* (%) employed | 257 (84.8) |
| Duration of sickness absence > 26 weeks (in the year before rehabilitation), *n* (%) | 114 (37.6) |
| Level of physical work demands, *n* (%) | |
| Light, *n* (%) | 23 (7.6) |
| Light to medium, *n* (%) | 97 (32.0) |
| Medium, *n* (%) | 123 (40.6) |
| Heavy, *n* (%) | 50 (16.5) |
| Main diagnosis according to the ICD 10th revision, *n* (%) | |
| M00–M25 (Arthropathies), *n* (%) | 41 (13.5) |
| M40–M54 (Dorsopathies), *n* (%) | 184 (60.7) |
| M60–M79 (Soft tissue disorders), *n* (%) | 35 (11.6) |
| S40–S49 (Injuries to the shoulder and upper arm), *n* (%) | 11 (3.6) |
| Current pain intensity (NRS, 0–10), mean (SD)[#] | 4.6 (2.3) |
| Work Ability Score (NRS, 0–10), mean (SD)[#] | 4.8 (2.5) |
| Pain-related disability at work (NRS, 0–10), mean (SD)[#] | 5.6 (2.6) |
| Expected RTW-duration[#] | |
| ≤ 1 month | 139 (45.9) |
| > more than 1 month | 90 (29.7) |
| No RTW | 63 (20.8) |

*SD* standard deviation; *NRS* numeric rating scale; *ICD-10* 10th revision of the international statistical classification of diseases and related health problems

[#]Patient-reports at discharge

## RTW and the predictive validity of ELA

The ability to cope with the physical demands of work was estimated as moderate or better in 180 participants (59.4%) and rather or very poor in 123 participants (40.6%). These estimations were based on an average of 3.6 activity tests (SD = 1.5). At 3-month follow-up, 117 participants (38.6%) had successfully returned to work. Concordance between the FCE and RTW outcome was found in 210 participants (69.3%; Table 3). Eight participants reported that RTW resulted from a change in job or a reduction in work demands and ten stated limitations in productivity. N-RTW was attributed to non-physical reintegration barriers only by 12 participants. After considering these aspects, the ELA result was rated as valid in 208 (68.6%) cases.

## Characteristics influencing the predictive validity of ELA

Regarding the current pain intensity, a strong correlation with the pain-related disability at work ($r = 0.742$; $p < 0.001$) was observed. Due to multicollinearity, the current pain intensity was not considered in the analyses. Among sociodemographic characteristics, the following variables exceeded the critical $p$ value ($< 0.1$) and were therefore excluded from further analyses: gender, age, marital status, highest professional qualification, migration background, unemployment rate and level of physical work demands (see appendix). Based on the multiple model calculated for the remaining variables ("model 1"), employment status and native language were also included into the final model (Table 4). As a result of model 2, general health perception, duration of sickness absence within the previous 12 months as well as psychosocial distress were selected for further analysis among the primarily health-related characteristics. With regard to the third group of primarily work-related patient reports, two characteristics did not exceed the critical p-value in model 3: pain-related disability at work and expected RTW-duration. In the fourth and final model, which achieved a high goodness of fit (AUC = 0.877), a native language other than German (OR 0.16; 95%-CI 0.05–0.56), psychosocial distress (OR 0.35; 95%-CI 0.15–0.82), a moderate (OR 0.15; 95%-CI 0.05–0.46) and strong pain-related disability at work (OR 0.19; 95%-CI 0.05–0.72) as well as the expectation to return to work after one month (OR

**Table 3** RTW depending on the ELA-based estimation of the ability to cope with the physical demands of work

| | Very good | Rather good | Moderate | Rather bad | Very bad | Total |
|---|---|---|---|---|---|---|
| n | 20 | 76 | 84 | 89 | 34 | 303 |
| RTW | 18 (90.0%) | 53 (69.7%) | 31 (36.9%) | 13 (14.6%) | 2 (5.9%) | 117 |
| N-RTW | 2 (10.0%) | 23 (30.3%) | 53 (63.1%) | 76 (85.4%) | 32 (94.1%) | 186 |

**Table 4** Associations between patient characteristics and the predictive validity of ELA: logistic regression models

| | Model 1 Sociodemographic | | Model 2 Primary health-related patient-reports | | Model 3 Primary health-related patient-reports | | Model 4 Final model | |
|---|---|---|---|---|---|---|---|---|
| | OR (95% CI) | $p$ | OR (95% CI) | $p$ | OR (95% CI) | $p$ | OR (95% CI) | $p$ |
| Native language ($R$ = German) | **0.31 (0.11–0.92)** | **0.035** | | | | | **0.16 (0.05–0.56)** | **0.004** |
| Months of employment within the last year (0–12) | 1.06 (0.99–1.14) | 0.117 | | | | | | |
| Type of work requirements ($R$ = not primarily physically demanding) | 0.57 (0.28–1.16) | 0.120 | | | | | | |
| Employment status ($R$ = employed ≥ 5 years) | | | | | | | | |
|   Employed < 5 years | **0.40 (0.17–0.92)** | **0.032** | | | | | 0.45 (0.17–1.18) | 0.104 |
|   Unemployed | **0.37 (0.14–0.95)** | **0.038** | | | | | 0.96 (0.33–2.77) | 0.938 |
| General health perception ($R$ = ≥ very good) | | | | | | | | |
|   Good | | | 0.47 (0.12–1.88) | 0.289 | | | 1.26 (0.25–6.33) | 0.783 |
|   ≤ Rather bad | | | **0.20 (0.05–0.84)** | **0.028** | | | 0.87 (0.15–4.87) | 0.869 |
|   Duration of sickness absence within the previous 12 months ($R$ = ≤ 26 weeks) | | | **0.35 (0.18–0.70)** | **0.003** | | | 0.81 (0.36–1.85) | 0.620 |
| Tampa Scale for Kinesiophobia ($R$ = < 30) | | | | | | | | |
|   30–35 | | | 0.91 (0.47–1.78) | 0.787 | | | | |
|   ≥ 36 | | | 1.07 (0.34–3.40) | 0.908 | | | | |
|   Psychosocial distress ($R$ = < 4 points) | | | 0.51 (0.25–1.04) | 0.064 | | | **0.35 (0.15–0.82)** | **0.015** |
| Work Ability Score ($R$ = 0–3) | | | | | | | | |
|   4–6 | | | | | 0.61 (0.22–1.66) | 0.330 | | |
|   7–10 | | | | | 1.42 (0.43–4.64) | 0.562 | | |
| Health-related disability at work ($R$ = not limited) | | | | | | | | |
|   Low | | | | | 0.42 (0.08–2.15) | 0.295 | | |
|   Severe | | | | | 0.36 (0.06–2.30) | 0.282 | | |
| Pain-related disability at work ($R$ = 0–3) | | | | | | | | |
|   4–6 | | | | | **0.26 (0.10–0.72)** | **0.009** | **0.15 (0.05–0.46)** | **0.001** |
|   7–10 | | | | | 0.30 (0.09–1.01) | 0.052 | **0.19 (0.05–0.72)** | **0.014** |
| Expected RTW-duration ($R$ = ≤ 1 month) | | | | | | | | |
|   > 1 month | | | | | **0.13 (0.05–0.35)** | **< 0.001** | **0.17 (0.06–0.46)** | **0.001** |
|   No RTW | | | | | **0.25 (0.07–0.91)** | **0.036** | 0.35 (0.09–1.28) | 0.111 |
| $N$ | 277 | | 286 | | 289 | | 278 | |
| Nagelkerke $R^2$ | 0.370 | | 0.389 | | 0.492 | | 0.535 | |
| AUC | 0.820 | | 0.808 | | 0.863 | | 0.877 | |

Bold values indicate statistically significance $p < 0.05$

Valid ELA results were 1-coded. All models were adjusted for the result of the FCE-based estimation of the ability to cope with the physical demands of work (≤ rather poor vs. moderate vs. ≥ rather good)

*OR* Odds ratio; *CI* Confidence Interval; *p* Wald statistic; *AUC* Area Under the ROC Curve; *R* Reference category

0.17; 95%-CI 0.06–0.46) were associated with a lower chance for a valid ELA outcome. Compared with 0-coded

characteristics only, for example, the likelihood of valid ELA results decreased from 96.9 to 61.7% when psychosocial stress and a moderate pain-related disability at work were reported.

## Discussion

The aim of the present study was to systematically identify patient-related characteristics that affect the predictive validity of FCE information. Data were collected from patients with MSD recruited in eleven rehabilitation centers offering WMR. As previous studies did not allow a pre-selection of relevant characteristics, the study had an exploratory character.

The ELA-based estimation of the participants' ability to cope with physical work demands was found to be valid in 208 of 303 (68.6%) cases. Among the numerous variables considered, four proved to be significant in predicting the validity of ELA: native language, psychosocial distress, the expected duration until RTW and the pain-related disability at work.

An increased chance for non-valid ELA outcomes was found in participants who reported an expected RTW duration of more than one month (OR 0.17). This result can be explained by the nonlinear relationship between RTW beliefs and the predictive validity of FCE-based estimations of the ability to cope with physical demands of work: The greater the (actual and perceived) discrepancy between abilities and requirements, whether positive or negative, the easier and clearer the interpretation of test results. For individuals who neither believe that they will return to work within a short period of time nor that they will not succeed in doing so at all, the profile comparison is, in general, less evident for FCE assessors and the possibility to return to work thus more difficult to estimate. This assumption is consistent with the finding that the ability to cope with the physical demands of work was predominantly estimated as moderate based on the ELA test in patients with an expected RTW duration of more than one month in case of non-valid ELA-results (27 of 44). Furthermore, it is to be considered that the RTW status might have been caused by a self-fulfilling prophecy in some cases. This may not only have influenced the proportion of valid ELA results but may also have led to an overestimation of the importance of the expected RTW duration.

Compared to those participants who reported a low pain-related disability at work or less, a moderate (OR 0.15) and strong disability (OR 0.19) were associated with an increased chance for non-valid ELA results. This finding can also be attributed to the u-shaped association described above. Furthermore, an increase in the pain-related disability at work seemed to have an independent negative impact on the predictive validity of ELA. This finding was expectable, as health care professionals frequently describe the difficulty of differentiating between pain and the (work-related) physical capacity as a key challenge in FCE. Under what conditions is it acceptable to confront patients with work demands while they are in pain? And under what conditions can patients be considered to be able to cope with the physical demands of work despite a pain reaction during FCE? The comparable influence of a moderate and strong pain-related disability at work may therefore be attributed to an increased importance of these questions with greater perceived limitations.

Considering the non-significant influence of the migration background (and thus possible cultural differences), the negative impact of a native language other than German is probably related to a greater difficulty in describing the physical work demands specifically. In contrast, the negative influence of psychosocial distress (OR 0.35) can be attributed to non-physical reintegration barriers.

Non-valid ELA results were primarily related to individuals that did not achieve RTW despite an ability to cope with the physical demands of work that was estimated as moderate. It is therefore questionable whether this estimation should be interpreted positively regarding the ability to cope with the physical demands of work like in the previous study (Bühne et al. 2020). The present findings rather indicate that this category should be interpreted neither positively nor negatively. However, the influence of the four characteristics changed only slightly when restricting the analysis to those individuals whose FCE result could be clearly assigned as positive ($\geq$ rather good) or negative ($\leq$ rather bad). Hence, we consider a substantial bias (resulting from the interpretation of the category "moderate") to be unlikely.

In contrast, no significant associations were found among the numerous other characteristics. Consequently, the predictive validity of ELA was not affected by age, gender, ICD-10 diagnoses, the employment status or fear-avoidance beliefs as well as those aspects, that Cheng and Cheng (Cheng and Cheng 2010, 2011) associated with lower predictive validity, the level of physical work demands and days from injury to FCE.

It is questionable to what extent the present findings can be applied to further FCE procedures. With regard to future studies, it would therefore be beneficial to conduct a comparable analysis using, for example, outcomes of widely used assessments like the WorkWell Systems FCE (Bieniek and Bethge 2014) or Ergo-Kit (Caron et al. 2015).

## Limitations

Despite the strengths of the study—multicentric design, holistic consideration of participant characteristics as well as the inclusion of productivity, adjustments in work

demands post discharge and nonphysical RTW-barriers in estimating the predictive validity of ELA—we have to consider the following limitations: (1) With regard to the transformations of continuous variables, which were carried out to address non-linear associations, it must be pointed out that these changes are accompanied by a loss of information. With respect to the work-related characteristics, however, the u-shaped association was confirmed when using the non-modified continuous work-related variables including their second and third power. (2) Studies indicate that the performance of the FCE is associated with an improvement of the patient-reported functional ability (Schindl et al. 2019; Bühne et al. 2017). Since participants were not blinded for the result of the ELA test, this may have affected the RTW outcome. The same applies to a possible positive reinforcement by the FCE assessors. However, in the context of WMR, there does not appear to be an overall positive impact of FCE on perceived work ability (Bühne et al. 2017). In addition, individuals who achieved RTW despite a negative RTW expectation (duration of more than one month or no RTW) constituted a clear minority in this sample. Therefore, and due to the consideration of productivity as well as the days of the sick leave post discharge in estimating the ability to cope with the physical demands of work, a significant bias is considered unlikely. Assuming that the impact on self-perception is part of the essence of FCE, it should also be considered that blinding may affect the practical relevance. (3) It cannot be ruled out that the predictive validity of ELA was also affected by insufficient information about the physical demands of work—related information from the employer is rarely available in WMR—or an inadequate selection of tests. Such a selection has to be carried out in all facilities of WMR, whereby the methods have not been specified by the German Pension Insurance. Neither does ELA provide a standardized approach. (4) The researchers were also not blinded for the ELA outcome. (5) The multidimensionality of RTW was not sufficiently considered in the present analysis. For example, days of sick leave might have been caused by social conflicts at work in some cases. Others, however, may have returned to work despite low functional capacity as they received support from their employer and co-workers. (6) 12 participants were excluded as FCE assessors claimed not to be able to assess the ability to cope with the physical demands of work. The underlying reasons cannot be clearly derived from the available data. It may have been due to a lack of knowledge regarding the individual work demands—six patients were unemployed at his time. (7) No protocol was developed a priori. Although this was due to the exploratory nature of the analysis, it constitutes another limitation. In this respect, further studies are necessary to confirm or reject the present methods and findings. This applies in particular to the observed u-shaped associations, the procedure of estimating of the predictive validity of the ELA outcome as well as the generalizability of the present findings. Despite the limitations mentioned above, we consider a substantial bias to be unlikely.

## Conclusion

Among the 28 patient-related characteristics collected, an increased likelihood for non-valid ELA results was present in participants who reported a moderate or strong pain-related disability at work, a native language other than German, psychosocial distress as well as in those who expected to return to work after more than one month.

## Declarations

**Conflict of interest** Some authors (DB, TA, CH) are employees of the Institute for Quality Assurance in Prevention and Rehabilitation at the German Sport University Cologne, which is responsible for the development and training of ELA. MS is working at the German Pension Insurance (funding institution). IF declares that he has no conflict of interest.

**Ethical approval** All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as modified in 2000.

**Informed consent** Informed consent was obtained from all patients for being included in the study.

**Consent for publication** Not applicable.

## References

Ansuategui Echeita J, Bethge M, van Holland BJ, Gross DP, Kool J, Oesch P et al (2019) Functional capacity evaluation in different societal contexts: results of a multicountry study. J Occup Rehabil 29(1):222–236. https://doi.org/10.1007/s10926-018-9782-x

Bethge M, Markus M, Streibelt M, Gerlich C, Schuler M (2019) Effects of nationwide implementation of work-related medical rehabilitation in Germany: propensity score matched analysis. Occup Environ Med 76(12):913–919. https://doi.org/10.1136/oemed-2019-106065

Bevan S (2015) Economic impact of musculoskeletal disorders (MSDs) on work in Europe. Best Pract Res Clin Rheumatol 29(3):356–373. https://doi.org/10.1016/j.berh.2015.08.002

Bieniek S, Bethge M (2014) The reliability of WorkWell systems functional capacity evaluation: a systematic review.

BMC Musculoskelet Disord 15:106. https://doi.org/10.1186/1471-2474-15-106

Branton EN, Arnold KM, Appelt SR, Hodges MM, Battié MC, Gross DP (2010) A short-form functional capacity evaluation predicts time to recovery but not sustained return-to-work. J Occup Rehabil 20(3):387–393. https://doi.org/10.1007/s10926-010-9233-9

Bühne D, Alles T, Froböse I (2017) Der Einfluss des FCE-Verfahrens ELA auf die Selbsteinschätzung des Patienten in der MBOR [Impact of the FCE procedure "ELA" on patients' self-estimation in work-related medical rehabilitation (WMR) programs ("MBOR")]. DRV-Schriften 111:195–197

Bühne D, Alles T, Hetzel C, Streibelt M, Froböse I, Bethge M (2020) Predictive validity of a customized functional capacity evaluation in patients with musculoskeletal disorders. Int Arch Occup Environ Health 93(5):635–643. https://doi.org/10.1007/s00420-020-01518-5

Caron J, Ronzi Y, Bodin J, Richard I, Bontoux L, Roquelaure Y, Petit A (2015) Interest of the Ergo-Kit for the clinical practice of the occupational physician. A study of 149 patients recruited in a rehabilitation program. Ann Phys Rehabil Med 58(5):289–297. https://doi.org/10.1016/j.rehab.2015.08.002

Cheng AS-K, Cheng SW-C (2010) The predictive validity of job-specific functional capacity evaluation on the employment status of patients with nonspecific low back pain. J Occup Environ Med 52(7):719–724. https://doi.org/10.1097/JOM.0b013e3181e48d47

Cheng ASK, Cheng SWC (2011) Use of job-specific functional capacity evaluation to predict the return to work of patients with a distal radius fracture. Am J Occup Ther 65(4):445–452. https://doi.org/10.5014/ajot.2011.001057

de Baets S, Calders P, Schalley N, Vermeulen K, Vertriest S, van Peteghem L et al (2018) Updating the evidence on functional capacity evaluation methods: a systematic review. J Occup Rehabil 28(3):418–428. https://doi.org/10.1007/s10926-017-9734-x

Deutsche Rentenversicherung Bund (DRV) (2019a) Rehab report 2019 [German]. Deutsche Rentenversicherung Bund, Berlin

Deutsche Rentenversicherung Bund (DRV) (2019b) Guideline for work-related medical rehabilitation of the German Pension Insurance [German]. Deutsche Rentenversicherung Bund, Berlin

Gouttebarge V, Kuijer PPFM, Wind H, van Duivenbooden C, Sluiter JK, Frings-Dresen MHW (2009) Criterion-related validity of functional capacity evaluation lifting tests on future work disability risk and return to work in the construction industry. Occup Environ Med 66(10):657–663. https://doi.org/10.1136/oem.2008.042903

Gross DP, Battié MC (2004) The prognostic value of functional capacity evaluation in patients with chronic low back pain: part 2: sustained recovery. Spine (phila Pa 1976) 29(8):920–924. https://doi.org/10.1097/00007632-200404150-00020

Gross DP, Battié MC (2005) Factors influencing results of functional capacity evaluations in workers' compensation claimants with low back pain. Phys Ther 85(4):315–322

Gross DP, Battié MC (2006) Does functional capacity evaluation predict recovery in workers' compensation claimants with upper extremity disorders? Occup Environ Med 63(6):404–410. https://doi.org/10.1136/oem.2005.020446

Hosmer DW, Lemeshow S, Sturdivant RX (2013) Applied logistic regression, 3rd edn. Wiley, Hoboken, NJ

Ilmarinen J, Tuomi K, Klockars M (1997) Changes in the work ability of active employees over an 11-year period. Scand J Work Environ Health 23(Suppl 1):49–57

Kuijer PPFM, Gouttebarge V, Brouwer S, Reneman MF, Frings-Dresen MHW (2012) Are performance-based measures predictive of work participation in patients with musculoskeletal disorders? A systematic review. Int Arch Occup Environ Health 85(2):109–123. https://doi.org/10.1007/s00420-011-0659-y

Lakke SE, Soer R, Krijnen WP, van der Schans CP, Reneman MF, Geertzen JHB (2015) Influence of physical therapists' kinesiophobic beliefs on lifting capacity in healthy adults. Phys Ther 95(9):1224–1233. https://doi.org/10.2522/ptj.20130194

Löffler S, Wolf H-D, Neuderth S, Vogel H (2009) Screening-Verfahren in der medizinischen Rehabilitation. In: Hillert A, Albrecht DG (eds) Medizinisch-beruflich orientierte Rehabilitation: Grundlagen und klinische Praxis; mit 61 Tabellen. Dt Ärzte-Verl, Köln, pp 133–140

Müller E, Bengel J, Wirtz M (2014) Validity and responsiveness of the scales for the assessment of functioning at work [German]. Rehabilitation 53:176–183. https://doi.org/10.6099/1000070

Murray CJL, Barber RM, Foreman KJ, Ozgoren AA, Abd-Allah F, Abera SF et al (2015) Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: quantifying the epidemiological transition. Lancet 386(10009):2145–2191. https://doi.org/10.1016/S0140-6736(15)61340-X

Ranft A, Fiedler R, Greitemann B, Heuft G (2009) Diagnostics of work motivation (DIAMO): optimization and construct validity [German]. Psychother Psychosom Med Psychol 59(1):21–30. https://doi.org/10.1055/s-2007-986296

Rusu AC, Kreddig N, Hallner D, Hülsebusch J, Hasenbring MI (2014) Fear of movement/(Re)injury in low back pain: confirmatory validation of a German version of the Tampa Scale for Kinesiophobia. BMC Musculoskelet Disord 15:280. https://doi.org/10.1186/1471-2474-15-280

Schindl M, Wassipaul S, Wagner T, Gstaltner K, Bethge M (2019) Impact of functional capacity evaluation on patient-reported functional ability: an exploratory diagnostic before-after study. J Occup Rehabil 29(4):711–717. https://doi.org/10.1007/s10926-019-09829-2

Schmidt S, Samos F-AZ, Klement A, Krüger J-M, Mau W (2019) Implementation and acceptance of a psychosocial screening in primary und orthopedic care to manage rehabilitation access in musculoskeletal diseases [German]. Phys Med Rehab Kuror 29(4):206–214. https://doi.org/10.1055/a-0842-6901

Soer R, van der Schans CP, Groothoff JW, Geertzen JHB, Reneman MF (2008) Towards consensus in operational definitions in functional capacity evaluation: a Delphi Survey. J Occup Rehabil 18(4):389–400. https://doi.org/10.1007/s10926-008-9155-y

Streibelt M, Gerwinn H, Hansmeier T, Thren K, Müller-Fahrnow W (2007) SIMBO: a screening instrument for identification of work-related disabilities—analyses of construct and prognostic validity [German]. Rehabilitation 46(5):266–275. https://doi.org/10.1055/s-2007-970583

Streibelt M, Blume C, Thren K, Reneman MF, Mueller-Fahrnow W (2009) Value of functional capacity evaluation information in a clinical setting for predicting return to work. Arch Phys Med Rehabil 90(3):429–434. https://doi.org/10.1016/j.apmr.2008.08.218

Tait RC, Chibnall JT, Krause S (1990) The Pain Disability Index: psychometric properties. Pain 40(2):171–182. https://doi.org/10.1016/0304-3959(90)90068-O

Ware JE, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 30(6):473–483