Journal of the International Neuropsychological Society (2022), **28**, 473–482

doi:10.1017/S1355617721000643

Performance Validity and Outcome of Cognitive Behavior Therapy in Patients with Chronic Fatigue Syndrome

Jeroen J. Roor[1,2,*], Brechje Dandachi-FitzGerald[3], Maarten J.V. Peters[3], Hans Knoop[4,5] and Rudolf W.H.M. Ponds[6,7]

[1]Department of Medical Psychology, VieCuri Medical Center, Venlo, The Netherlands
[2]School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands
[3]Faculty of Psychology and Neuroscience, Department of Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands
[4]Department of Medical Psychology, Amsterdam University Medical Centers, Amsterdam Public Health Research Institute, University of Amsterdam, Amsterdam, The Netherlands
[5]Department of Medical Psychology, Expert Centre for Chronic Fatigue, Amsterdam University Medical Centers, VU University, Amsterdam, The Netherlands
[6]Department of Medical Psychology, Amsterdam University Medical Centers, Amsterdam, The Netherlands
[7]Department of Psychiatry and Neuropsychology, Limburg Brain Injury Center, Maastricht University, Maastricht, The Netherlands

(Received October 23, 2020; Final revision March 25, 2021; Accepted April 7, 2021; First Published Online June 16, 2021)

Abstract

Objective: There is limited research examining the impact of the validity of cognitive test performance on treatment outcome. All known studies to date have operationalized performance validity dichotomously, leading to the loss of predictive information. Using the range of scores on a performance validity test (PVT), we hypothesized that lower performance at baseline was related to a worse treatment outcome following cognitive behavioral therapy (CBT) in patients with Chronic Fatigue Syndrome (CFS) and to lower adherence to treatment. **Method:** Archival data of 1081 outpatients treated with CBT for CFS were used in this study. At baseline, all patients were assessed with a PVT, the Amsterdam Short-Term Memory test (ASTM). Questionnaires assessing fatigue, physical disabilities, psychological distress, and level of functional impairment were administered before and after CBT. **Results:** Our main hypothesis was not confirmed: the total ASTM score was not significantly associated with outcomes at follow-up. However, patients with a missing follow-up assessment had a lower ASTM performance at baseline, reported higher levels of physical limitations, and completed fewer therapy sessions. **Conclusions:** CFS patients who scored low on the ASTM during baseline assessment are more likely to complete fewer therapy sessions and not to complete follow-up assessment, indicative of limited adherence to treatment. However, if these patients were retained in the intervention, their response to CBT for CFS was comparable with subjects who score high on the ASTM. This finding calls for more research to better understand the impact of performance validity on engagement with treatment and outcomes.

Keywords: Performance validity, Treatment outcome, Chronic fatigue syndrome, Cognitive behavioral therapy, Amsterdam short-term memory test, Effort

INTRODUCTION

The frequency of performance validity test (PVT) failure is substantial in nonforensic clinical settings (Dandachi-FitzGerald, van Twillert, van de Sande, van Os, & Ponds, 2016; Martin & Schroeder, 2020), and its impact on neuropsychological test performance is known to be as large or even greater than various medical and psychiatric conditions (Iverson, 2006; Sollman & Berry, 2011). PVT failure invalidates cognitive test results and hinders the clinician in making adequate diagnoses about cognitive (dis)functioning and, consequently, recommendations for treatment. One might, therefore, anticipate that the impact of performance invalidity is not limited to the diagnostic assessment, but extends to treatment efficacy.

The notion that PVT failure is relevant beyond the diagnostic domain and may also be related to everyday functioning has received some consideration. For example, Lippa et al. (2014) found that PVT failure is related to self-reported community participation in veterans with mild traumatic brain injury. Although research on this topic is limited, performance validity may serve as a behavioral proxy of how a patient copes with everyday life, and as such may convey potentially relevant information for treatment planning, adherence, and outcome.

*Correspondence and reprint requests to: Jeroen J. Roor, School for Mental Health and Neuroscience, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands. Email: jeroen.roor@maastrichtuniversity.nl

To the best of our knowledge, only five studies have examined the relationship between performance validity and treatment. Moore et al. (2013) found that, in comparison with patients who passed a PVT, patients with schizophrenia and schizoaffective disorder ($n = 128$) who failed a PVT had significantly lower therapy attendance—which is known to negatively affect treatment outcome (Long, Dolley, & Hollin, 2012). Psychiatric symptoms or cognitive impairment did not predict group therapy attendance in this study. Horner, VanKirk, Dismuke, Turner, and Muzzy (2014) studied the relationship between invalid performance and healthcare utilization in a heterogeneous outpatient sample of a Veterans Affairs Medical Center ($n = 355$). They found that PVT failure in these patients was associated with increased and longer inpatient hospitalizations and more emergency department visits. A recent study by Jurick and colleagues (2020) of veterans with posttraumatic stress disorder (PTSD) and mild-to-moderate traumatic brain injury ($n = 100$) found that both subjects who passed and those who failed a PVT benefitted from treatment, although valid performers showed the greatest reduction in PTSD symptoms. No significant differences were found in treatment completion between patients who passed and those who failed a PVT (i.e., 57.9% and 46.5%, respectively). Williams and colleagues (2020) found that veterans ($n = 61$) benefitted equally from PTSD treatment, regardless of PVT failure. Goedendorp, van der Werf, Bleijenberg, Tummers, and Knoop (2013) examined whether invalid performance at baseline assessment was related to treatment outcome in patients with chronic fatigue syndrome (CFS; $n = 169$). These authors found a higher loss to follow-up (i.e., missing follow-up assessment) in CFS patients who failed a PVT (i.e., 23%), in comparison with patients who passed the PVT (i.e., 8%). For the patients who failed a PVT, no group differences were found in comparison with the patients who passed the PVT with regard to change in fatigue severity, functional impairments, or physical limitations following cognitive behavioral therapy (CBT) for CFS. Although research on this topic is limited, the studies described suggest that invalid performance is negatively associated with the response or adherence to treatment.

It is important to emphasize that the aforementioned studies all used performance validity dichotomously; subjects who failed a PVT were classified as "invalid performers" and subjects who passed a PVT as "valid performers." Consequently, subjects who scored just on opposite sides of the cutoff were interpreted as being very different, when in fact their PVT scores were close to each other. Subjects who scored below the cutoff were also considered similar, even when their PVT scores varied greatly. However, the positive predictive value (PPP) of invalid performance—i.e., the probability that PVT failure represents true invalid performance—is related to the severity of PVT failure. Consequently, PVT scores well below cutoff are more likely to represent true invalid performance in comparison with PVT scores at cutoff. Consequently, adhering to a dichotomous approach will evidently lead to loss of information and a decrease in the statistical power to detect a relationship between performance validity and treatment outcome (Altman & Royston, 2006).

The current study addresses this limitation. The aim was to replicate the aforementioned study of Goedendorp et al. (2013), but using the total range of scores of a PVT and a considerably larger sample size to examine the impact that performance validity has on the outcome for CFS after CBT. We hypothesized that lower PVT scores (i.e., indicating lower levels of effort to perform to the best of one's abilities) at baseline would be related to worse treatment outcome (i.e., higher levels of self-reported symptoms of CFS after CBT) and lower treatment adherence (i.e., more loss to follow-up and fewer completed treatment sessions) in comparison with patients who produced higher PVT baseline scores.

METHOD

Participants

Archival pre- and posttreatment data on CBT for CFS were used for this study. Patients were consecutively referred to a tertiary treatment facility for chronic fatigue in a university hospital. First, the patients' medical status was assessed by consultants of the Department of Internal Medicine, to rule out other medical explanations for their fatigue, and second to scan for the potential need for additional medical examination. This procedure followed national CFS guidelines (Centraal Begeleidings Orgaan, 2013), which are in accord with the US Centers for Disease Control and Prevention (CDC) guidelines formulated in 2003 (Reeves et al., 2003). If patients met the CDC criteria for CFS, they were referred to the treatment center. All the patients in this sample were seeking treatment for CFS and were seen in the context of routine clinical care. Since it is known that being involved in a legal procedure with respect to disability claims is related to poor treatment outcome of CBT for CFS (Prins, Bazelmans, Van der Werf, Van der Meer, & Bleijenberg, 2001), patients who were engaged in a disability claim were excluded from starting treatment.

Patients were included in this study if they were severely fatigued (i.e., scored 35 or higher on the fatigue severity subscale of the Checklist Individual Strength (CIS) questionnaire (Worm-Smeitink et al., 2017)), and had significant functional impairments in daily life (i.e., had a weighted total score ≥ 700 on the Sickness Impact Profile 8 (SIP8) (Jacobs, Luttik, Touw-Otten, & de Melker, 1990)). Additional inclusion criteria were (1) Dutch language proficiency and (2) being 18 years or older. The data were collected between April 2007 and April 2015 in the context of treatment (i.e., CBT for CFS). The questionnaires and tests used in this study were part of the routine clinical assessment. It was standard practice that all patients completed the Amsterdam Short-Term Memory test (ASTM; Schmand & Lindeboom, 2005) at baseline. The medical ethics committee of Radboud University Medical Centre approved this study. This research was conducted in accordance with the Helsinki declaration.

https://doi.org/10.1017/S1355617721000643 Published online by Cambridge University Press

None of the participants from the Goedendorp et al. (2013) study are represented in the current sample. In the current study, the same clinical procedure, diagnostic criteria for CFS, and inclusion criteria (i.e., score of 35 or higher on the fatigue severity subscale of the Checklist Individual Strength questionnaire and a weighted total score ≥ 700 on the Sickness Impact Profile 8) were used, as in the study of Goedendorp et al. (2013).

Procedures

Before CBT treatment, a neuropsychological assessment was conducted, consisting of a clinical interview by a psychologist, followed by the administration of tests and questionnaires by a test assistant (*see* Instruments). After CBT was completed, a follow-up assessment was conducted, in which only the questionnaires were readministered. Patients were invited for a follow-up assessment with a test assistant, separately from the last treatment session. If they did not respond to the initial invitation, they were contacted multiple times by telephone.

Interventions

Individual and group face-to-face CBT for CFS was provided according to a published treatment protocol (Knoop & Bleijenberg, 2010). The protocol is based on a model of cognitive-behavioral fatigue-perpetuating factors (Knoop, Prins, Moss-Morris, & Bleijenberg, 2010). The aim of CBT is to reduce fatigue and disabilities by changing fatigue-related cognitions and behaviors. CBT for CFS consists of about 12 to 14 sessions during a 6-month period.

Measures

Education was assessed by self-report, classifying formal schooling on an 8-point scale often used in the Netherlands (de Bie, 1987). Based upon Van der Elst, van Boxtel, van Breukelen, and Jolles (2005); three groups of education level were formed: low (those with primary education at most), medium (those with junior vocational training at most), and high (those with senior vocational or academic training).

Performance validity was measured with the Amsterdam Short-Term Memory test (ASTM). The ASTM is a 30-trial forced-choice word recognition procedure. The total calculated score is used as a cutoff for invalid performance. In the original validation studies, a cutoff score lower than 84 was associated with a specificity of 93% and a sensitivity of 84% in discriminating experimental malingerers and a heterogeneous neurological patient group. The internal consistency was found to be excellent (Cronbach's $\alpha = 0.91$) (Schmand & Lindeboom, 2005).

The total score of the revised Dutch-language version of the Symptom Checklist (SCL-90) was used to measure psychological distress, and the Depression subscale (16 items) was used to measure symptoms of depression (Arrindell & Ettema, 2005; Derogatis, 1994). All items are rated on a 5-point Likert scale, ranging from "not at all" (0) to "extremely" (4). Reliability and validity of the revised Dutch-language version of the SCL-90 are qualified as good (Arrindell, et al., 2003).

Fatigue during the past two weeks was assessed with the fatigue severity subscale of the Checklist Individual Strength (CIS) questionnaire (Worm-Smeitink et al., 2017). The CIS fatigue severity subscale contains eight items, with a score range of 8–56. Higher scores indicate higher levels of fatigue. The CIS questionnaire is extensively validated for the assessment of fatigue (Worm-Smeitink et al., 2017).

Physical disabilities were measured with the physical functioning subscale of the Medical Outcomes Survey Short-Form-36 (Ware Jr & Sherbourne, 1992). Scores on this scale range from 0 to 100, with higher scores indicating fewer physical limitations. The SF-36 is a reliable and valid instrument (Scheeres, Wensing, Knoop, & Bleijenberg, 2008; Ware Jr & Sherbourne, 1992).

Functional impairments in daily functioning were assessed using the Sickness Impact Profile 8 (SIP8) (Bergner, Bobbitt, Carter, & Gilson, 1981). The SIP8 total score consists of eight subscales: alertness behavior, sleep/rest, leisure activities, homemaking, work limitations, mobility, social interactions and ambulation. The eight subscales of the SIP are added to a weighted total score, with higher scores indicating more functional impairments [range 0–5799]. The SIP is a reliable instrument (Bergner et al., 1981) and has been validated for the Dutch population (Jacobs et al., 1990).

As with Goedendorp et al. (2013), loss to follow-up was determined by missing follow-up assessment after CBT for CFS. Since treatment dropout was not registered in this study, we examined the number of completed therapy sessions as a proxy of treatment adherence.

Data Analyses

When the amount of missing treatment outcome data is significant, it is likely that complete case analysis (CC) introduces bias and results in estimates with less precision, leading to loss of statistical power. Applying statistical methods that handle missing data appropriately is therefore advocated in reporting observational studies (von Elm et al., 2007). To this end, we used multiple imputation (MI). MI is a commonly used method for handling missing data. It has the potential to counteract the impact that CC has on the results so that bias is reduced and precision is improved. Briefly, in MI, a model is fitted for the missing values of dependent variables. Predictor variables and auxiliary variables (i.e., variables that are not included in the final analyses but are related to variables of interest) are used for this purpose. An estimated (i.e., imputed) value is then calculated for every missing value, ensuring that these scores are near the collected scores of comparable subjects.

Following the suggested guidelines for MI reporting (Sterne et al., 2009), an imputation model with full

https://doi.org/10.1017/S1355617721000643 Published online by Cambridge University Press

conditional specifications was created on the assumption that data were missing at random (MAR). First, binary logistic regression analyses were conducted using the baseline measures (i.e., ASTM, CIS fatigue score, SCL-90 total score, SCL-90 depression score, SIP total score, and SF-36 physical functioning) and demographic information (sex, level of education, and age) to examine which variables predicted missing data at follow-up. We used a less strict significance level ($p < .10$) to include all potential confounding variables. The baseline measures of ASTM and SF-36 physical functioning were negatively associated with missing follow-up data. This suggests an inverse relationship; an increase on the ASTM (i.e., more effort to perform to the best of abilities) or SF36 physical functioning (i.e., reporting fewer physical limitations) was associated with a decrease in missing follow-up data. Therefore, these two variables were used to generate the imputations. Additionally, to preserve the association between outcome measures and predictors, all follow-up variables (i.e., CIS fatigue subscale, SIP total score, SCL-90 total score, and SF36 physical functioning) were retained in the imputation model (Spratt et al., 2010). We used 25 imputations to reduce the impact that random sampling has on pooled data (Spratt et al., 2010). For all subjects ($n = 1081$), missing follow-up values were calculated based upon the multiple imputation procedure outlined. Consequently, all analyses on treatment outcome were performed using the pooled imputed follow-up variables. Complete case (CC) analyses using nonimputed data were used for examining treatment adherence (i.e., missing follow-up assessment and number of completed therapy sessions).

Assumptions concerning linearity were assessed through visual inspection of residuals. To examine the appropriateness of the imputation model, results based on the original (nonimputed) data were compared with those based on the multiple imputations.

Hierarchical linear regression analyses were used to examine the relationship between the continuous ASTM (predictor) score and treatment outcome (criterion), controlling for potential confounding factors. Since we were specifically interested in the impact of performance validity on outcome after CBT for CFS, treatment outcome was defined by follow-up scores on a set of preferred outcome variables used in CFS research (Janse, Wiborg, Bleijenberg, Tummers, & Knoop, 2016): the CIS fatigue subscale, SIP total score, SCL-90 total score, and physical functioning subscale of the SF36, for which Bonferroni correction was applied (alpha = .0125). Older age in combination with depressive symptoms is associated with an increase of false-positive scores on the ASTM (Schmand & Lindeboom, 2005). In addition, low intelligence is known to negatively influence PVT performance (Lippa, 2018). Therefore, predictor variables were entered in two steps. The first step contained level of education (i.e., dummy variables Low and Medium levels of education) as a proxy of intelligence, depression (i.e., SCL-90 depression subscale), and age as predictors for treatment outcome. Step 2 included all of the above predictors and added the total range ASTM score as a predictor. Since SPSS is not able to provide pooled R^2 (change) data for imputed datasets, the mean R^2 (change) values for models 1 and 2 of the 25 imputed datasets were calculated manually. This is the preferred method for combining R^2 (change) across multiple imputed datasets (Van Ginkel, 2019).

Since a significant proportion of patients were lost to follow-up (i.e., did not complete follow-up measurement), we performed a secondary analysis to examine which patient characteristics were related to loss to follow-up, and whether loss to follow-up was related to the number of completed therapy sessions (as a proxy of therapy adherence). Mann-Whitney *U* tests were used to examine differences in test and questionnaire scores administered at baseline. Differences in age, level of education, sex, and number of completed therapy sessions between subjects who completed the follow-up assessment and those who did not were examined using an independent *t*-test or Fisher's exact test as deemed appropriate.

All analyzes were performed using the Statistical Package for the Social Sciences software (SPSS), version 23.0, with $p < .05$ (two-tailed) used as the significance level for baseline analyses and $p < .01$ (Bonferroni correction) for follow-up analyses.

RESULTS

Sample Characteristics

A total of 1382 patients fulfilled the inclusion criteria. Only patients who had started with CBT ($n = 1081$) were included. The variables of age, sex, and level of education were complete. All baseline measures were near to complete (missing < 1%). Data at follow-up were missing for the CIS fatigue subscale ($n = 222$; 20.53%), SF-36 physical functioning ($n = 222$; 20.53%), the SCL-90-R depression subscale and SCL-90-R total score ($n = 273$; 25.25%), and for the SIP total score ($n = 221$; 20.46%).

Table 1 provides an overview of demographics and treatment data at baseline and follow-up using the original (i.e., nonimputed) data of CFS patients provided with CBT. This sample consisted predominantly of women (75.39% female) in their thirties (mean age 36.98 years) with medium to high levels of education.

In addition, the continuous ASTM score was negatively related to all self-reported baseline measures (i.e., CIS fatigue, physical limitations, functional impairment, psychological distress, and depressive symptoms; all p's < .0125).

Hierarchical Linear Regression Analyses on Scores of Fatigue Severity, Physical Limitations, Functional Impairment, and Psychological Distress at Follow-up

The association between the ASTM and all outcome measures (i.e., CIS fatigue, SIP total score, SCL-90 total score,

https://doi.org/10.1017/S1355617721000643 Published online by Cambridge University Press

Table 1. Demographics and treatment data at baseline and follow-up using the original (i.e., nonimputed) data.

	Baseline			Follow-up		
	($n = 1081$)			($n = 859$)		
	M	SD	%	M	SD	%
Patient characteristics						
Age (years)	36.98	(11.74)		–		
Education						
Low			9.53	–		
Medium			59.63	–		
High			30.92	–		
Female			75.39	–		
Treatment data						
Loss to follow-up			–			20.53
ASTM	86.89	(3.78)		–		
[a]ASTM fail (score < 84)			11.38			–
CIS fatigue	50.51	(5.03)		29.83	(14.10)	
SF36 physical functioning	57.25	(20.33)		80.34	(20.81)	
SIP total	1572.31	(551.32)		658.07	(659.25)	
SCL-90 total	164.56	(38.43)		130.71	(36.93)	

[a] cutoff used by Goedendorp et al. (2013). ASTM = Amsterdam short-term memory test; CIS fatigue = checklist for individual strength, fatigue subscale; SF36 physical functioning = medical outcomes survey short-form-36, physical functioning subscale; SIP total = sickness impact profile, total score; SCL-90 total = symptom checklist-90 total score.

and SF36 physical subscale) were linear, based upon visual inspection of their respective residual plots.

The first model accounting for the combined explained variance in age, level of education, and depressive symptoms (i.e., SCL-90 depression subscale) on treatment outcome was significant ($p < .0125$) for all 25 imputation datasets of all criterion variables (i.e., CIS fatigue, SF-36 physical functioning subscale, SIP total score, and SCL-90 total score during follow-up; data not shown). The second model, with the added continuous ASTM score as predictor of treatment outcome, was also significant for all 25 imputed datasets of all criterion variables (data not shown). Importantly, the ASTM score added in Model 2 did not yield a significant improvement in the prediction of treatment outcome for any of 25 imputed datasets of all criterion variables on top of the predictors in Model 1 (i.e., the R^2 change was not significant; data not shown). Importantly, the continuous ASTM score was found not to be significantly associated with any of the follow-up scores (see Table 2). Furthermore, older age was found to be significantly associated with worse outcome on all follow-up scores. Higher levels of depressive symptoms (i.e., a higher SCL-90 depression score) at baseline were significantly associated with worse outcome on the CIS fatigue subscale score, SIP total score, and the SCL-90 total score after treatment. Low and medium levels of education were found to be significantly associated with a worse outcome on SF36 physical functioning. These findings were confirmed using the original (non-imputed) data, where the addition of the continuous ASTM in the regression model did not yield a significant improvement in the prediction of treatment outcome (see Appendix A). In addition, the association of the individual predictors with treatment outcome was comparable based on the original data, with an added significant association between higher levels of self-reported depressive symptoms and worse outcome on SF36 physical functioning (data not shown).

We chose to include depressive symptom reporting in combination with age in the regression models, since older subjects with higher levels of reported depression have a higher chance of producing false-positive ASTM scores—as mentioned in the ASTM manual. Leaving depressive symptom reporting (i.e., the SCL-90 depression subscale) out of the regression models, however, did not alter results: ASTM performance was still not significantly associated with any of the outcome measures at follow-up (i.e., CIS fatigue, the SIP total score, the SCL-90 total score, or SF-36 physical limitations).

Since in practice the ASTM is intended to be used categorically (sufficient versus insufficient performance validity), the mentioned hierarchical linear regression analyses were re-examined using the ASTM cutoff (i.e., score < 84) as a predictor instead of using its continuous score. ASTM failure was not found to be related to any of the outcome variables (data not shown).

When, in line with Goedendorp and colleagues (2013), change scores (baseline minus follow-up scores for CIS fatigue, SF36 physical functioning, the SIP total score, and SCL-90 total score) were used as criterion variables to define treatment outcome, the findings of the hierarchical linear regression analyses were replicated; the continuous ASTM score was found not to be significantly related with

https://doi.org/10.1017/S1355617721000643 Published online by Cambridge University Press

Table 2. Pooled data from a hierarchical linear regression analysis assessing the relationship between level of education, depressive symptoms, age, and the total score range of the Amsterdam Short-Term Memory test at baseline as predictors, and fatigue severity, physical limitations, functional impairment, and psychological distress at follow-up as dependent variables.

Step and predictor variables	$\overline{R}^2$	B	95% CI Lower	Upper	*t*	*p*-value	Effect size Cohen's f^2
		Subscale fatigue severity of the CIS at follow-up					
Step 1	.03						.03
Constant		18.20	13.43	22.97	7.49	< .01	
Low education		1.34	−2.23	4.90	.74	.46	
Medium education		.005	−2.03	2.04	.005	.99	
SCL-90, depression		.19	.08	.291	3.53	< .01*	
Age		.16	.07	.241	3.64	< .01*	
Step 2	.03						< .01
Constant		21.22	−2.51	44.96	1.76	.08	
Low education		1.25	−2.39	4.90	.68	.50	
Medium education		−.02	−2.07	2.03	−.02	.98	
SCL-90, depression		.18	.08	.29	3.49	< .01*	
Age		.15	.07	.24	3.61	< .01*	
ASTM		−.03	−0.29	.22	−.26	.80	
		Physical limitations subscale of the SF-36 at follow-up					
Step 1	.06						.07
Constant		102.38	95.69	109.07	30.02	< .01	
Low education		−8.36	−13.62	−3.09	−3.12	< .01*	
Medium education		−5.11	−8.02	−2.19	−3.44	< .01*	
SCL-90, depression		−.18	−.33	−.03	−.25	.01	
Age		−.35	−.47	−.23	−5.89	< .01*	
Step 2	.06						< .01
Constant		78.19	44.55	111.85	4.56	< .01	
Low education		−7.65	−13.04	−2.31	−2.81	< .01*	
Medium education		−4.89	−7.83	−1.96	−3.27	< .01*	
SCL-90, depression		−.17	−.32	−.02	−2.29	.02	
Age		−.35	−.46	−.23	−5.79	< .01*	
ASTM		.27	−.09	.63	1.45	.15	
		Functional impairment measured with the SIP at follow-up					
Step 1	.06						.06
Constant		−114.64	−326.75	97.48	−1.06	.29	
Low education		−1.27	−166.79	164.25	−.01	.99	
Medium education		28.17	−68.76	125.10	.57	.57	
SCL-90, depression		12.07	7.40	16.73	5.07	< .01*	
Age		10.52	6.88	14.15	5.68	< .01*	
Step 2	.06						< .01
Constant		703.53	−354.50	1761.56	1.30	.19	
Low education		−24.47	−193.41	144.19	−.28	.78	
Medium education		21.06	−76.64	118.76	.42	.67	
SCL-90, depression		11.63	6.94	16.33	4.86	< .01*	
Age		10.33	6.69	13.98	5.57	< .01*	
ASTM		−9.10	−20.57	2.36	−1.56	.12	
		Psychological distress measured with the SCL-90 at follow-up					
Step 1	.01						.11
Constant		76.14	63.86	88.43	12.18	< .01*	
Low education		9.64	.64	18.63	2.10	.04	
Medium education		4.01	−1.35	9.37	1.47	.14	
SCL-90, depression		1.29	1.03	1.55	9.69	< .01*	
Age		.34	.13	.55	3.24	< .01*	
Step 2	.11						.13
Constant		102.26	41.94	162.57	3.33	< .01*	
Low education		8.89	−.28	18.08	1.90	.06	
Medium education		3.78	−1.59	9.15	1.38	.17	
SCL-90, depression		1.27	1.01	1.54	9.57	< .01*	

(*Continued*)

https://doi.org/10.1017/S1355617721000643 Published online by Cambridge University Press

Table 2. (*Continued*)

Step and predictor variables	$\overline{R}^2$	B	95% CI Lower	Upper	*t*	*p*-value	Effect size Cohen's f^2
		Subscale fatigue severity of the CIS at follow-up					
Age		.34	.13	.55	3.17	< .01*	
ASTM		−.29	−.95	.37	−.87	.39	

ASTM = Amsterdam short-term memory test; B = unstandardized B; CI = confidence interval; CIS = checklist individual strength; Effect size: Cohen's $f^2 = \overline{R}^2$ change/(1–R^2 change); SCL-90 = symptom checklist-90 total score; SCL-90 depression = symptom checklist-90 depression subscale; SF36 = medical outcomes survey short-form-36; SIP = sickness impact profile total score; $n = 1081$; *p-value < .0125.

any of the change scores, using both the imputed datasets and the original data (data not shown).

Baseline and Treatment Characteristics of Subjects Without Follow-up Assessment

There were no differences in age ($t[1079] = -1.80$, $p = .07$), level of education ($\chi^2[2] = 3.85$, $p = .15$), or sex (Fisher's exact test, $p = .14$) between patients who did or did not complete follow-up. Subjects lost to follow-up performed significantly lower on the ASTM (Mann-Whitney U test, $p = .02$; $\eta^2 = .005$) and reported higher levels of physical limitations (Mann-Whitney U test, $p < .01$; $\eta^2 = .014$) at baseline. When using the ASTM dichotomously (using the cutoff of < 84), subjects who did not complete follow-up failed the ASTM significantly more often in comparison with the subjects who completed follow-up (resp. 15.8% and 10.1%; Fisher's Exact Test, $p = .023$, $\phi = 0.17$). No group differences were found at baseline for fatigue severity (i.e., CIS fatigue), functional impairment level (i.e., the SIP total score) or psychological distress (i.e., the SCL-90 total score) between subjects with or without follow-up assessment. Moreover, subjects who did not complete the follow-up assessment finished fewer therapy sessions in comparison with subjects who completed follow-up ($t[1079] = 16.40$, $p < .01$; mean scores of 8.67 and 14.42 respectively; Hedge's $g = 1.28$). This suggests that loss to follow-up is closely related to therapy dropout.

DISCUSSION

While considerable attention has been focused on examining the performance validity of diagnostic assessments in various clinical samples including CFS, few studies have examined the impact of performance validity on response or adherence to treatment. Previous studies all took a dichotomous approach to performance validity, leading to loss of information and consequently reducing the statistical power to detect a relationship between performance validity and criterion variables. We chose to use the total PVT score instead, taking into account the limitation of a dichotomous approach to performance validity. To our knowledge, the current study is the first to examine the association between the total score range of a PVT and treatment outcome.

Our main hypothesis that lower ASTM scores are associated with worse treatment outcome (i.e., higher levels of self-reported symptoms or disability following CBT for CFS) was disconfirmed. A continuous ASTM-score yielded no significant associations with any outcome measures (i.e., CIS fatigue, SF-36 physical functioning, SIP total score, SCL-90 total score) in subjects who completed follow-up. However, as hypothesized, loss to follow-up was found to be associated with lower ASTM scores, as well as with higher levels of self-reported physical limitations at baseline and fewer completed therapy sessions. The latter suggests that subjects with missing follow-up assessment were not as engaged in their treatment because they attended significantly fewer therapy sessions in comparison with subjects who completed follow-up. This conclusion is reasonable, since subjects with missing follow-up assessment completed fewer therapy sessions than the 12–14 therapy sessions described in the treatment protocol. To summarize, these results indicated that CFS patients who scored low on the ASTM during baseline assessment were more likely to complete fewer therapy sessions and have missing follow-up data. However, if low-scoring CFS patients are retained in the intervention, their response to CBT for CFS is comparable with that of subjects who scored high on the ASTM (i.e., indicating effort to perform to the best of their abilities).

Our study findings are consistent with those of Goedendorp et al. (2013), who found that ASTM failure was: 1. not associated with outcome after CBT for CFS, and 2. related to loss to follow-up. The replicated findings suggest that low ASTM scores (i.e., indicating lower levels of effort to perform to the best of one's abilities) impact treatment adherence, but are not related to responsiveness to treatment in CFS patients who completed follow-up. Moore and colleagues (2013) directly studied the relation between PVT failure and treatment adherence in a sample of patients with schizophrenia and schizoaffective disorder, who were provided with a skills-training treatment. They found that PVT failure was associated with lower group therapy attendance. This suggests that PVT results are associated with subsequent treatment adherence across existing diagnostic groups.

A multitude of patient characteristics and situational factors may be associated with the relationship between low PVT performance and (study) dropout. For example, financial incentives (e.g., a pending disability claim) are linked to low PVT performance (Bianchini, Curtis, & Greve, 2006;

https://doi.org/10.1017/S1355617721000643 Published online by Cambridge University Press

Sherman, Slick, & Iverson, 2020). Obviously, these incentives may also negatively impact treatment outcome, since improvement in functioning may result in lower disability compensation. However, since participants were excluded when they were engaged in a disability claim, financial incentives are not likely to be present in the current study sample. Besides, external incentives also come in the form of avoiding more basic duties such as work, school, home responsibilities, or any undesirable outcome (Sherman et al., 2020). In most cases, the clinician is unaware of the presence of these incentives, which "may be detrimental to therapeutic success" (Van Egmond, Kummeling, & Balkom, 2005, p. 416). In general, a broader perspective on performance invalidity beyond malingering (i.e., intentionally feigning symptoms for external motives) is desirable. Besides factitious disorder (i.e., intentionally feigning symptoms for internal motives), various psychological constructs are suggested that might result in invalid performance (Silver, 2015). Empirical studies on this topic, however, are limited and focused, for example, on "diagnosis threat" (for a critical review, see Niesten, Merckelbach, Dandachi-FitzGerald, & Jelicic, 2020), perceived injustice (Iverson, Terry, Karr, Panenka, & Silverberg, 2018), and the health locus of control and self-efficacy (Armistead-Jehle, Lippa, & Grills, 2020). Preferably, these constructs are measured independently (of self-reporting) and at least using a check on the validity of self-reported measures. For example, Armistead-Jehle et al. (2020) omitted subjects with noncredible symptom reports, and found no relationship between PVT failure and the self-reported health locus of control and self-efficacy. However, when reanalyzing their data including subjects who failed symptom validity measures, a trend was observed between PVT failure and reporting a lower internal locus of control and higher inefficacy. Taking these caveats into account, it is important to conduct empirical research into possible underlying mechanisms of invalid performance. If one thing is now clear, it is that performance invalidity is not restricted to the realm of malingering and that its relevance extends beyond "noise" during diagnostic decision-making.

This was an observational study using archival treatment data, which prevents causal inferences on the relationship between performance validity and treatment outcome. Furthermore, the current findings using the ASTM cannot readily be generalized to other PVTs, which may have shown different results. Additionally, in the current study, outcome measures relied fully on self-reporting instead of more objective measures. The validity of self-reporting is itself known to be influenced by, for example, intentional symptom exaggeration (Sherman et al., 2020), inattentive responding (Merckelbach, Dandachi-FitzGerald, van Helvoort, Jelicic, & Otgaar, 2019), and the unreliability of memory in general (Loftus, Levidow, & Duensing, 1992). This is not only a limitation of the current study. In general, there is a lack of well-researched methods for evaluating the validity of reported somatic symptoms (e.g., fatigue or pain). Without questioning the clinical value of more general measures of symptom validity (e.g., based upon the validity scales of the MMPI), there is a movement toward assessing the validity of specific symptoms/conditions (Sherman et al., 2020). Promising in this regard is the Self-Report Symptom Inventory (SRSI)—issued after the inclusion period of the current study—containing subscales on pseudo items for fatigue and pain (Merten, Merckelbach, Giger, & Stevens, 2016). However, in the current study, ASTM performance and self-reported symptoms were negatively related at baseline, in accordance with the findings of Goedendorp et al. (2013). Despite this association, baseline ASTM performance did not impact treatment outcome (based upon self-reporting), but did impact loss to follow-up and number of completed therapy sessions. Therefore, since invalid performance and symptom validity can be viewed as "separate but related aspects of the broader construct of symptom exaggeration" (Haggerty, Frazier, Busch, & Naugle, 2007, p. 926), future studies may want to employ both symptom validity and performance validity when examining treatment outcome.

Finally, some may argue that the PVT utilized measured genuine cognitive (dis)functioning in CFS patients instead of performance validity. However, it is important to emphasize that the ASTM—and PVTs in general—are constructed to be relatively insensitive to cognitive dysfunction. By definition, these tests require little cognitive effort. For example, ASTM performance was examined in nonlitigating bonafide neurology patients diagnosed with Parkinson's disease, multiple sclerosis, and cerebrovascular accidents without "obvious clinical cognitive symptoms" (e.g., repeating the same "story," not being able to refer to an earlier subject of conversation). It is highly unlikely that these cognitive symptoms were present in the current sample of relatively young, medium, and highly educated CFS patients. The mean ASTM score in the mentioned sample of neurology patients was 87.3 (SD 2.9), with 92% of these subjects passing the ASTM (Merten Bossink, & Schmand, 2007). On a related note, using known-groups' design, the sensitivity (i.e., detection of insufficient effort to perform to the best of abilities) of the ASTM was found to be excellent in its original validation study (Schagen, Schmand, de Sterke, & Lindeboom, 1997), and comparable with the TOMM Trial 2 and TOMM Retention Trial (Bolan, Foster, Schmand, & Bolan, 2002). Therefore, low scores on the ASTM in the current sample of CFS patients were more likely to be reflective of poor performance validity than of genuine cognitive impairment.

Taken together, our findings have clinical implications. First, that low ASTM performance in CFS patients is not a reason to be excluded from CBT, since these subjects' response to treatment is comparable with subjects who performed to the best of their abilities (i.e., had higher PVT scores) during the baseline assessment. However, low performance on the ASTM was associated with loss to follow-up and fewer completed therapy sessions. Therefore, instead of being an indicator restricted to the assessment of symptom credibility, performance validity may also serve as a behavioral proxy of how patients engage in a behavioral treatment intervention (e.g., some might have reservations about the communicated diagnoses and/or treatment plans). Additional research is necessary to help understand the association between performance

https://doi.org/10.1017/S1355617721000643 Published online by Cambridge University Press

validity, adherence to treatment, and outcomes. Ultimately, a better determination of factors that are known to impact treatment adherence and treatment outcome may sharpen indications for treatment and, consequently, prevent costly specialized tertiary medical care.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S1355617721000643

ACKNOWLEDGMENTS

We thank all patients for participating in this study, and Lianne Vermeeren and Tex de Vroom for help with data management. Part of the results of this study was presented at the 47th Annual Meeting of the International Neuropsychological Society (INS), February 2019, New York City, USA.

FINANCIAL SUPPORT

This study received no financial support.

CONFLICT OF INTEREST

The authors have no competing interests to report.

REFERENCES

Armistead-Jehle, P., Lippa, S.M., & Grills, C.E. (2020). The impact of self-efficacy and health locus of control on performance validity testing [published online ahead of print, 2020 May 7]. *Archives of Clinical Neuropsychology*. doi: 10.1093/arclin/acaa027

Altman, D.G. & Royston, P. (2006). The cost of dichotomizing continuous variables. *The British Medical Journal*, *332*, 1080.

Arrindell, W.A., & Ettema, J.H. (2005). *Symptom Checklist. Handleiding bij een multidimensionale psychopathologie indicator [Symptom Checklist. Manual of a multidimensional psychopathology indicator]*. Amsterdam: Harcourt Test Publishers.

Arrindell, W.A., Ettema, J.H.M., Groenman, N., Brook, F., Janssen, I., Slaets, J., … Dost, S. (2003). De groeiende inbedding van de Nederlandse SCL-90-R [The growing embedding of the Dutch-language version of the SCL-90-R]. *De Psycholoog*, *11*, 576–582.

Bergner, M., Bobbitt, R.A., Carter, W.B., & Gilson, B.S. (1981). The Sickness Impact Profile - development and final revision of a health-status measure. *Medical Care*, *19*, 787–805.

Bianchini, K.J., Curtis, K.L., & Greve, K.W. (2006). Compensation and malingering in traumatic brain injury: A dose-response relationship? *The Clinical Neuropsychologist*, *20*(4), 831–847.

Bolan, B., Foster, J.K., Schmand, B., & Bolan, S. (2002). A comparison of three tests to detect feigned amnesia: The effects of feedback and the measurement of response latency. *Journal of Clinical and Experimental Neuropsychology*, *24*(2), 154–167.

Centraal Begeleidings Orgaan. (2013). Richtlijn Diagnose, behandeling, begeleiding en beoordeling van patiënten met het chronisch vermoeidheidssyndroom (CVS) [Guideline: Diagnosis, treatment, coaching and evaluation of patients suffering chronic fatigue syndrome (CFS)]. Retrieved from https://www.diliguide.nl/document/3435/file/pdf/2013.

Dandachi-FitzGerald, B., van Twillert, B., van de Sande, P., van Os, Y., & Ponds, R.W. (2016). Poor symptom and performance validity in regularly referred hospital outpatients: Link with standard clinical measures, and role of incentives. *Psychiatry Research*, *239*, 47–53.

De Bie, S. (1987). *Standaardvragen 1987: Voorstellen voor uniformering van vraagstellingen naar achtergrondkenmerken en interviews [Standard questions 1987: Proposal for uniformization of questions regarding background variables and interviews]*. Leiden, the Netherlands: Leiden University Press.

Derogatis, L.R. (1994). *SCL-90-R: Administration, scoring and procedures manual* (3rd ed.). Minneapolis, MN: Nation Computer Systems.

Goedendorp, M.M., van der Werf, S.P., Bleijenberg, G., Tummers, M., & Knoop, H. (2013). Does neuropsychological test performance predict outcome of cognitive behavior therapy for chronic fatigue syndrome and what is the role of underperformance? *Journal of Psychosomatic Research*, *75*, 242–248.

Haggerty, K.A., Frazier, Th.W., Busch, R.M., & Naugle, R.I. (2007). Relationships among Victoria Symptom Validity Test indices and Personality Assessment Inventory validity scales in a large clinical sample. *The Clinical Neuropsychologist*, *21*, 917–928.

Horner, M.D., VanKirk, K.K., Dismuke, C.E., Turner, T.H., & Muzzy, W. (2014). Inadequate effort on neuropsychological evaluation is associated with increased healthcare utilization. *The Clinical Neuropsychologist*, *28*(5), 703–713.

Iverson, G.L. (2006). Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Applied Neuropsychology*, *13*(2), 77–90.

Iverson, G.L., Terry, D.G., Karr, J.E., Panenka, W.J., & Silverberg, N.D. (2018). Perceived injustice and its correlates after mild traumatic brain injury. *Journal of Neurotrauma*, *35*, 1156–1166.

Jacobs, H.M., Luttik, A., Touw-Otten, F.W., & de Melker, R.A. (1990). The Sickness Impact Profile; results of an evaluation study of the Dutch version (De sickness impact profile; resultaten van een valideringsonderzoek van de Nederlandse versie). *Nederlands Tijdschrift voor Geneeskunde*, *134*, 1950–1954.

Janse, A., Wiborg, J.F., Bleijenberg, G., Tummers, M., & Knoop, H. (2016). The efficacy of guided self-instruction for patients with idiopathic chronic fatigue: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *84*(5), 377–388.

Jurick, S.M., Crocker, L.D., Merrit, V.C., Hoffman, S.N., Keller, A.V., Eglit, G.M., … Jak, A. M. (2020). Psychological symptoms and rates of performance validity improve following trauma-focused treatment in veterans with PTSD and history of mild-to-moderate TBI. *Journal of the International Neuropsychological Society*, *26*(1), 108–118.

Knoop, H., & Bleijenberg, G. (2010). *Het chronisch vermoeidheidssyndroom. Behandelprotocol cognitieve gedragstherapie voor CVS [Chronic fatigue syndrome. Cognitive behavior therapy for CFS treatment protocol]*. Houten, The Netherlands: Bohn Stafleu Van Loghum.

Knoop, H., Prins, J. B., Moss-Morris, R., & Bleijenberg, G. (2010). The central role of cognitive processes in the perpetuation of chronic fatigue syndrome. *Journal of Psychosomatic Research*, *68*(5), 489–494.

Lippa, S.M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, *32*(3), 391–421.

Lippa, S.M., Pastorek, N.J., Romesser, J., Linck, J., Sim, A.H., Wisdom, N.M., & Miller, B.I. (2014). Ecological validity of performance validity testing. *Archives of Clinical Neuropsychology*, *29*(3), 236–244.

Loftus, E.F., Levidow, B., & Duensing, S. (1992). Who remembers best? Individual differences in memory for events that occurred in a science museum. *Applied Cognitive Psychology*, *6*, 93–107.

Long, C., Dolley, O., & Hollin, C. (2012). Engagement in psychosocial treatment: Its relationship to outcome and care pathway progress for women in medium-secure settings. *Criminal Behaviour and Mental Health*, *22*, 336–349.

Martin, P.K. & Schroeder, R.W. (2020). Base rates of invalid test performance across clinical non-forensic contexts and settings. *Archives of Clinical Neuropsychology*, *35*(6), 717–725.

Merckelbach, H., Dandachi-FitzGerald, B., van Helvoort, D., Jelicic, M., & Otgaar, H. (2019). When patients overreport symptoms: More than just malingering. *Current Directions in Psychological Science*, *28*(3), 321–326.

Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology*, *29*(3), 308–318.

Merten, T., Merckelbach, H., Giger, P., & Stevens, A. (2016). The Self-Report Symptom Inventory (SRSI): A new instrument for the assessment of symptom overreporting. *Psychological Injury and Law*, *9*(2), 102–111.

Moore, R., Davine, T., Harmell, A., Cardenas, V., Palmer, B., & Mausbach, B. (2013). Using the repeatable battery for the assessment of neuropsychological status (RBANS) effort index to predict treatment group attendance in patients with schizophrenia. *Journal of the International Neuropsychological Society*, *19*(2), 198–205.

Niesten, I.J., Merckelbach, H., Dandachi-FitzGerald, B., & Jelicic, M. (2020, April 9). The iatrogenic power of labeling medically unexplained symptoms: A critical review and meta-analysis of "diagnosis threat" in mild head injury. *Psychology of Consciousness: Theory, Research, and Practice*. Advance online publication. http://dx.doi.org/10.1037/cns0000224

Prins, J.B., Bazelmans, E., Van der Werf, S.P., Van de Meer, J., & Bleijenberg, G. (2001). Cognitive-behaviour therapy for chronic fatigue syndrome: predictors of treatment outcome. Paper presented at the psycho-neuro-endocrino- immunology (PNEI): A common language for the whole human body: Proceedings of the 16th World Congress of Psychosomatic Medicine, Göteborg, Sweden.

Reeves, W.C., Lloyd, A., Vernon, S.D., Klimas, N., Jason, L.A., Bleijenberg, G., … Unger, E.R. (2003). Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution. *BMC Health Services Research*, *3*, 25.

Scheeres, K., Wensing, M., Knoop, H., & Bleijenberg, G. (2008). Implementing cognitive behavioral therapy for chronic fatigue syndrome in a mental health center: a benchmarking evaluation. *Journal of Consulting and Clinical Psychology*, *76*(1), 163–171.

Schagen, S., Schmand, B., de Sterke, S., & Lindeboom, J. (1997). Amsterdam Short Term Memory Test: A new procedure for the detection of feigned memory deficits. *Journal of Clinical and Experimental Neuropsychology*, *19*, 43–51.

Schmand, B. & Lindeboom, J. (2005). *Amsterdam Short-Term Memory Test: Manual*. Leiden, The Netherlands: Psychologische Instrumenten, Tests en Services.

Sherman, E.M., Slick, D.J., & Iverson, G.L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology*, *35*(6), 735–764.

Silver, J.M. (2015). Invalid symptom reporting and performance: What are we missing? *NeuroRehabilitation*, *36*, 463–469.

Sollman, M.J. & Berry, D.T. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology*, *26*(8), 774–789.

Spratt, M., Carpenter, J., Sterne, J.A., Carlin, J.B., Heron, J., Henderson, J., & Tilling, K. (2010). Strategies for multiple imputations in longitudinal studies. *American Journal of Epidemiology*, *172*, 478–487.

Sterne, J., White, I.R., Carlin, J.B., Spratt, M.P., Royston, P., Kenward, M.G., … Carpenter, J.R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *The British Medical Journal*, *338*, 157–160.

Van der Elst, W., van Boxtel, M.P., van Breukelen, G.J., & Jolles, J. (2005). Rey's verbal learning test: Normative data for 1855 healthy participants aged 24-81 years and the influence of age, sex, education, and mode of presentation. *Journal of the International Neuropsychological Society*, *11*, 290–302.

Van Egmond, J., Kummeling, I., & Balkom, T. (2005). Secondary gain as hidden motive for getting psychiatric treatment. *European Psychiatry*, *20*(5–6), 416–421.

Von Elm, E., Altman, D.G, Egger, M., Pocock, S.J., Gøtzsche, P.C., Vandenbroucke, J.P, & STROBE initiative (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Lancet*, *370*, 1453–1457.

Van Ginkel, J.R. (2019). Significance tests and estimates for R^2 for multiple regression in multiply imputed datasets: A cautionary note on earlier findings, and alternative solutions. *Multivariate Behavioral Research*, *54*(4), 514–529.

Ware, J.E Jr. & Sherbourne, C.D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection, *Medical Care*, *30*, 473–483.

Williams, M.W., Graham, D., Sciarrino, N.A., Estey, M., McCurry, K.L., Chiu, P., & King-Casas, B. (2020). Does validity measure response affect CPT group outcomes in veterans with PTSD? *Military Medicine*, *185*(3–4), 370–376.

Worm-Smeitink, M., Gielissen, M., Bloot, L., van Laarhoven, H.M., van Engelen, B.M., van Riel, P., … Knoop, H. (2017). The assessment of fatigue: Psychometric qualities and norms for the Checklist Individual Strength. *Journal of Psychosomatic Research*, *98*, 40–46.