
SYMPOSIUM

Systemically Treated Breast Cancer Patients and Controls: An Evaluation of the Presence of Noncredible Performance

Jeffrey S. Wefel,¹ Rozemarijn L. Kornet,² AND Sanne B. Schagen³

¹Department of Neuro-Oncology, M.D. Anderson Cancer Center, Houston, Texas

²Department of Psychology, Vrije Universiteit, Amsterdam, The Netherlands

³Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

(RECEIVED June 27, 2013; FINAL REVISION December 27, 2013; ACCEPTED December 27, 2013; FIRST PUBLISHED ONLINE March 10, 2014)

Abstract

This study sought to define the frequency of noncredible performance in breast cancer patients before, during and after completion of systemic treatment, as well as predictors of noncredible performance. We examined six datasets investigating the cognitive effects of chemotherapy and/or endocrine therapy. Embedded performance validity test (PVT) measures were identified and used to evaluate the datasets. One dataset included a standalone PVT. Possible noncredible performance was operationally defined as performance below criterion on three or more PVT indices. This was undertaken as cancer patients have been observed clinically to fail PVTs both in the context of external gain and independent of such motivators. A total of 534 breast cancer patients and 214 healthy controls were included in the analysis. Percentages of patients performing below cutoff on one or more PVT varied from 0% to 21.2%. Only 1 patient met the criterion of noncredible performance. Calculation of post-test probabilities indicated a more than 90% chance to detect noncredible performance. There is no evidence to suggest noncredible performance in breast cancer patients and healthy controls who choose to participate in research studies examining cognitive function. Thus, the observational data showing that non-central nervous system (CNS) cancer and therapies not targeting the CNS can have untoward effects on cognitive function are unlikely to be due to noncredible performance. (*JINS*, 2014, *20*, 357–369)

INTRODUCTION

Over the past several decades a growing body of research has demonstrated the occurrence of cognitive changes associated with systemic therapies for non-central nervous system (CNS) tumors (Ahles, Root, & Ryan, 2012; O'Farrel, Mackenzie, & Collins, 2013). Many people worldwide are confronted with non-CNS cancer. Age is a risk factor for cancer and given trends toward increased longevity in industrialized countries as well as improvements in survivorship, we are likely to see more and more cancer patients, many of whom may experience treatment associated cognitive dysfunction. Therefore, understanding cognitive changes associated with cancer treatments and their impact on quality of life and daily life functioning is important.

Numerous cross-sectional and longitudinal studies have demonstrated treatment-related cognitive decline in women with breast cancer that receive chemotherapy and/or endocrine therapy (see Wefel and Schagen, 2012, for a review). In this patient group alone, over 60 neuropsychological studies have been published that have investigated whether chemotherapy is associated with cognitive impairment. The vast majority of the neuropsychological studies demonstrate cognitive impairment and/or cognitive decline in a subgroup of breast cancer patients treated with cytotoxic agents compared to breast cancer patients that did not receive chemotherapy or compared to non-cancer controls. Patients show deficits on a wide range of neuropsychological tests but core impairments are related to learning and memory, executive functions and psychomotor speed—indicative of a frontal-subcortical profile (Wefel, Vardy, Ahles, & Schagen, 2011). Cognitive problems may arise during treatment and may persist years after completion of treatment. Endocrine therapy (consisting of selective estrogen receptor modulators

Correspondence and reprint requests to: Sanne B. Schagen, Netherlands Cancer Institute, Division of Psychosocial Research and Epidemiology, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands. E-mail: s.schagen@nki.nl

and/or aromatase inhibitors) is frequently part of the treatment strategy for many women with breast cancer, and may also add to the occurrence of cognitive problems (Phillips et al., 2010; Schilder et al., 2010). Preclinical studies as well as imaging studies have provided insights on the neurobiological changes associated with many chemotherapeutic agents that result in the cognitive changes that have been observed (Seigers, Schagen, van Tellingen, & Dietrich, 2013; Seigers, & Fardell, 2011). Specifically, imaging studies have shown gray and white matter volume loss, altered white matter integrity, and abnormalities in functional activation as well as resting state network connectivity (Pomykala, de Ruiter, Deprez, McDonald, & Silverman, 2013; Scherling, & Smith, 2013). Preclinical animal studies have shown increased apoptosis in healthy proliferating cells in the central nervous system as well as damage to neural precursor cells.

Despite this strong, multidimensional body of preclinical, human, and neuroimaging research supporting a neurobiological relationship between chemotherapy and cognitive decline, critics have asserted that poor neuropsychological test performance may be influenced by factors other than neurological disease, including conditions such as anxiety and depression, or inattentiveness secondary to various causes including limited motivation. Studies on the cognitive effects of cancer treatments have shown that anxiety and depression do not seem to drive the relationship between cognitive problems as assessed by neuropsychological tests and systemic therapies (Wefel & Schagen, 2012).

In neuropsychological practice, performance validity testing (PVT) (Larrabee, 2012) has been advanced as an important element in the evaluation of both litigating and non-litigating patient populations. In fact, many neuropsychologists recommend PVT to be included in standard practice (Boone, 2009; Dandachi-Fitzgerald, 2013). However, PVT failure has not been systematically explored in research on the cognitive effects of systemic therapies in non-CNS cancer patients. Various terms are used in the literature to describe both intentional and unintentional underperformance that may undermine the validity of a neuropsychological examination, and at present no clear consensus exists on the preferred descriptor. We choose to use the term “noncredible performance” derived from the consensus conference statement on neuropsychological assessment of effort, response bias and malingering (Heilbronner et al., 2009).

It is plausible to imagine situations in which noncredible performance on cognitive testing is likely in this patient population and thus could invalidate the results of these studies. In clinical practice, one must always be aware of the potential of a patient to intentionally underperform for purposes of external gain. Although it may seem unlikely to practitioners not routinely caring for patients in an oncologic setting, we have each experienced patients failing PVTs both in the context of overt external gain scenarios and outside of this context. Furthermore, many practicing in this area have experienced occasions when the medical team asserts the observed cognitive deficits are due to patient attempting to appear impaired rather than treatment-related neurotoxicity.

Unlike in traumatic brain injury, to date there are no studies in cancer populations reporting on noncredible performance to assist practicing clinicians in discussing the relative frequency or infrequency of noncredible performance in this population. If noncredible performance in this patient population is frequent, reconsideration of the interpretation that cognitive decline can occur with chemotherapy would be essential.

It is also imperative to evaluate the possibility of noncredible performance in non-cancer control groups as many published studies on cognitive performance of cancer patients compare patients to either published normative data of healthy individuals or to a control group consisting of women without a history of cancer. Non-patient control populations may also produce noncredible performance on formal cognitive tests when participating in a study on the cognitive effects of a disease/treatment not applicable to themselves.

There is a large body of research on the use and the limitations of PVTs (Strauss, Sherman, & Green, 2006). Over the years, several methods have been developed to detect noncredible performance. Examples of PVTs include the Test of Memory Malingering (TOMM) (Tombaugh, 1996), the Amsterdam Short-Term Memory Test (ASTM) (Schmand & Lindeboom, 2003), the Hiscock Forced-Choice Test (Hiscock & Hiscock, 1989) and the Word Memory Test (WMT) (Green, 2003); the latter being one of the most popular and best investigated free standing PVTs currently available.

Another method to detect noncredible performance is to derive effort indices from conventional neuropsychological tests, so-called embedded PVT measures. Examples of such measures are the WAIS-III Digit Span subtest, several indices of the California Verbal Learning Test, the Rey Complex Figure Test, and Warrington's Recognition Memory test (Strauss et al., 2006).

PVT cutoffs are available for both stand-alone tests and embedded indices. The sensitivity and specificity of these measures at specific cutoffs have been studied in naïve malingerers, in subjects instructed to simulate deficits, and in clinical populations (generally mild TBI populations).

The current study sought to define the frequency of noncredible performance in breast cancer patients before, during and/or after completion of systemic treatment, as well as potential predictors of noncredible performance in this patient population. We examined six datasets from studies (Van Dam, et al., 1998; Schagen, van Dam, Muller, Boogerd, & Lindeboom, 1999; Schagen, Muller, Boogerd, Mellenbergh, & van Dam, 2006; Schilder et al., 2010; Wefel, Lenzi, Theriault, Davis, & Meyers, 2004; Wefel, Saleeba, Buzdar, & Meyers, 2010; Wefel, 2002) on the effects of either chemotherapy and/or endocrine therapy on the cognitive functioning of breast cancer patients. Occurrence of possible noncredible performance and differences in rates of possible noncredible performance between treatments and assessment times were examined. Also, differences in possible noncredible performance between patients undergoing systemic therapies and disease specific and healthy controls were investigated. Ruling out this potential explanatory factor for

the observed deficits in cognitive function would strengthen the conclusion that therapies not targeting the CNS can have untoward effects on cognitive function.

METHODS

Datasets

The characteristics of the different datasets are listed in Table 1 and Table 2. Datasets 1–3 were obtained at the Netherlands Cancer Institute (NKI). Datasets 4–6 were obtained at M.D. Anderson Cancer Center (MDACC). None of the datasets contain overlapping data.

1. *Dataset 1.* The first dataset originated from a prospective study performed by Schagen et al. (2006). In this study, changes in cognitive performance were examined in three different breast cancer groups and a healthy control group. The breast cancer patient groups consisted of (1) patients who adjuvantly received four cycles of standard-dose chemotherapy (FEC) followed by one cycle of high-dose chemotherapy (CTC) (labeled “CTC”, $n = 28$); (2) patients who received five cycles of standard-dose FEC chemotherapy (labeled “FEC”, $n = 39$); and (3) a breast cancer control group (labeled “Control”, $n = 57$) comprised by stage-1 breast cancer patients for whom systemic therapy was not part of the treatment strategy. The breast cancer patients receiving chemotherapy were randomized between either high-dose or standard dose-treatment (Rodenhuis et al., 2003). Patients who received chemotherapy also received endocrine therapy (20 mg a day for a period of 2 years). All patients underwent neuropsychological testing before chemotherapy and 6 months after treatment over a 12-month interval. The healthy control group underwent repeated testing over a 6-month interval.
2. *Dataset 2.* This is the dataset of a prospective study in which the influence of endocrine therapy on cognitive functioning was examined in postmenopausal patients with breast cancer (Schilder et al., 2010). Patients were randomized to either tamoxifen ($n = 80$) or exemestane ($n = 99$) and underwent neuropsychological testing after breast surgery and before the start of endocrine treatment. Follow-up assessment was conducted after 1 year of endocrine treatment. A healthy control group ($n = 120$) underwent the same assessments with a similar time interval.
3. *Dataset 3.* This dataset describes four patient groups that were examined on average 2 years after completion of treatment (Van Dam et al., 1998; Schagen et al., 1999). In this study, the cognitive status of breast cancer patients treated with various cytotoxic regimens as part of an adjuvant therapy strategy was examined. A group of patients treated with high-dose adjuvant chemotherapy (CTC group; $n = 34$), two groups treated with conventional dose chemotherapy (FEC group; $n = 36$, CMF group; $n = 39$). CTC and FEC patients and a select number of randomized CMF patients ($n = 20$) also received endocrine treatment with Tamoxifen for 2 years. A group of breast cancer patients without any systemic treatment was also included ($n = 34$).
4. *Dataset 4.* The fourth dataset stems from a prospective longitudinal study performed by Wefel et al. (2010). In this study, the incidence, nature, and chronicity of

Table 1. Characteristics of patients in NKI datasets

	<i>n</i>	Age Mean (SD)	Education ¹ Mean (SD)	Time(months) between assessments Mean (SD)	Fatigue ² (baseline) Mean (SD)	Fatigue ² (follow-up) Mean (SD)	Anxiety ³ (baseline) Mean (SD)	Anxiety ³ (follow-up) Mean (SD)	Depression ³ (baseline) Mean (SD)	Depression ³ (follow-up) Mean (SD)
Dataset 1										
CTC	28	45.5 (6.6)	4.2 (2.0)	11.9 (1.2)	34.1 (16.8)	32.9 (18.5)	17.5 (11.6)	12.7 (8.4)	13.3 (9.8)	12.9 (9.0)
FEC	39	45.5 (6.6)	4.4 (2.0)	11.9 (1.2)	34.8 (24.6)	25.4 (20.4)	14.3 (10.0)	12.1 (11.7)	13.2 (10.7)	12.9 (11.1)
No CT	57	50.7 (7.7)	4.1 (2.0)	12.6 (1.2)	31.4 (22.2)	31.4 (25.0)	17.5 (12.5)	16.5 (14.8)	16.0 (12.2)	15.1 (15.4)
Control	60	48.8 (6.0)	4.3 (1.8)	6.6 (1.2)	17.4 (19.3)	16.6 (15.1)	8.7 (7.9)	9.2 (8.2)	9.6 (9.2)	9.5 (11.9)
Dataset 2										
TAM	80	68.7 (7.6)	4.1 (1.4)	12.4 (1.2)	33.6 (21.5)	28.8 (23.2)	14.1 (12.4)	15.5 (14.9)	13.8 (11.9)	18.3 (15.8)
EXE	99	68.3 (6.8)	4.3 (1.3)	12.2 (1.4)	32.4 (22.3)	26.6 (21.3)	11.4 (11.0)	11.8 (10.1)	11.9 (11.1)	17.6 (16.3)
Control	120	66.2 (7.9)	4.7 (1.2)	12.3 (0.6)	18.3 (15.7)	21.5 (18.4)	9.5 (9.3)	11.2 (9.7)	10.9 (10.0)	13.5 (10.4)
Dataset 3										
CTC	34	45.1 (6.1)	4.9 (1.3)	x	35.3 (26.2)	x	19.4 (15.4)	x	18.8 (22.0)	x
FEC	36	48.1 (6.8)	5.2 (1.0)	x	25.3 (27.7)	x	14.1 (13.5)	x	13.1 (11.0)	x
CMF	39	47.1 (6.5)	5.2 (1.1)	x	25.6 (21.7)	x	15.5 (12.4)	x	16.9 (13.7)	x
Control	34	46.1 (5.2)	4.8 (0.7)	x	18.6 (20.4)	x	14.3 (15.9)	x	9.3 (10.1)	x

¹Education is based on Verhage education scores 1–7 (Verhage, 1964), corresponding with the following US years of education; 1: 1–5 years; 2: 6 years; 3: 7–8 years; 4: 7–9 years; 5: 7–10 years; 6: 7–16 years; 7: 17–20 years.

²Fatigue based on EORTC QLQ-C30, EORTC = European Organization for Research and Treatment of Cancer: EORTC QLQ-C30 is a health-related quality of life questionnaire (Aaronson et al., 1993). Score on the Fatigue scale ranges from 0 to 100, with a higher score meaning more bothered by complaint.

³Anxiety and Depression based HSCL-25, HSCL-25 = the Hopkins symptom checklist-25 (Hesbacher et al., 1978). Scores range from 0 to 100; a higher score indicates more complaints.

CTC = cyclophosphamide, thiotepa, carboplatin; FEC = 5-fluorouracil, epirubicin, cyclophosphamide; CMF = cyclophosphamide, methotrexate, 5-fluorouracil; CT = chemotherapy; TAM = tamoxifen; EXE = exemestane.

Table 2. Characteristics of patients in MDACC datasets

Dataset	Treatment	n	Age Mean (SD)	Years of education Mean (SD)	Time (months) from baseline Mean (SD)	Fatigue ¹ Mean (SD)	Anxiety ² Mean (SD)	Depression ³ Mean (SD)
Dataset 4								
Baseline	FAC +/- paclitaxel	42	48.8 (8.1)	13.0 (2.5)		x	-0.94 (1.4)	8.5 (6.9)
Eval.2		29	47.9 (9.0)	12.9 (2.3)	2.9 (0.6)	x	-0.52 (1.7)	8.7 (8.3)
Eval.3		33	47.6 (7.9)	13.2 (2.4)	7.0 (1.4)	x	0.24 (1.2)	5.8 (5.0)
Eval.4		30	47.7 (8.0)	13.1 (2.6)	13.1 (2.8)	x	0.07 (1.0)	5.8 (5.0)
Dataset 5								
Baseline	FAC +/- Methotrexate/vinblastine	18	45.4 (6.7)	14.1 (2.6)		x	56.7 (9.0)	57.9 (9.4)
Eval.2		18	45.4 (6.7)	14.1 (2.6)	7.5 (2.4)	x	55.4 (5.6)	63.4 (10.0)
Eval.3		15	45.3 (7.0)	14.3 (2.7)	19.5 (6.2)	x	52.7 (6.7)	55.6 (8.8)
Dataset 6								
Baseline	TAM	62	52.8 (9.2)	15.4 (2.1)		2.7 (2.1)	5.4 (5.6)	6.2 (6.7)
Eval.2		49	53.0 (9.5)	15.4 (2.1)	4.4 (2.2)	2.3 (1.5)	8.0 (8.3)	6.3 (6.1)
Eval.3		40	52.8 (9.1)	15.5 (2.0)	8.8 (3.2)	2.5 (2.0)	4.7 (4.6)	4.1 (4.3)

¹ Fatigue level on a 0-10 scale with 10 being the worst.

² Anxiety based on State-Trait Anxiety Inventory (z-scale) (Spielberger et al., 1983), Minnesota Multiphasic Personality Inventory (MMPI) (Greene, 1991) scale 7 (T-scores) and, Beck Anxiety Inventory (Beck, 1990). A higher score indicated a higher level of anxiety.

³ Depression based on the Beck Depression Inventory (Beck, 1996) and, MMPI (Greene, 1991) scale 2 (T-scores).

FAC = 5-fluorouracil, doxorubicin, cyclophosphamide; TAM = tamoxifen.

cognitive dysfunction in 42 breast cancer patients undergoing adjuvant chemotherapy according to the FAC regimen with or without paclitaxel was examined. Patients were tested before, during, and at two time points after chemotherapy.

- Dataset 5.** This set contains data from a prospective longitudinal study performed by Wefel et al. (2004). In this study, cognitive changes of breast cancer patients who received chemotherapy with FAC with or without methotrexate and vinblastine was examined over time ($n = 18$). Patients underwent neuropsychological testing before and at three time points after completion of chemotherapy.
- Dataset 6.** The last dataset originates from a prospective longitudinal study performed by Wefel (2002) on cognitive changes in breast cancer patients treated with endocrine therapy ($n = 62$). Patients underwent neuropsychological testing after their primary medical treatment (e.g., surgery, radiation, chemotherapy) and before starting Tamoxifen. Follow-up assessment was performed twice after starting endocrine treatment.

We searched Pubmed, EMBASE, and PsychInfo (January 1980–January 2013) and reference lists of articles in the English language for embedded PVTs that corresponded with the neuropsychological tests used in the six studies. Search terms were: (noncredible performance) AND (embedded symptom or performance validity) AND (malingering) AND (embedded effort) in combination with the neuropsychological instruments used in our datasets.

The following tests were determined to contain embedded PVT indices: California Verbal Learning Test (CVLT), Controlled Oral Word Association (COWA), WAIS-III subtest Digit Span, Rey-Osterrieth Complex Figure Test (ROCF), Rey Auditory Verbal Learning Test (RAVLT), Stroop test, and Trail Making Test (TMT). We also examined the Fake Bad Scale (FBS) from the Minnesota Multiphasic Personality Inventory (MMPI).

Table 3 summarizes the identified performance validity indices, the operational definitions of noncredible performance for each index, and their sensitivity and specificity. For each test, the embedded PVT index was evaluated based on cutoffs previously described in published literature. Additionally, measures of fatigue, depression, and anxiety were obtained in most of the study datasets and were examined in relation to PVT performance.

Measures

- California Verbal Learning Test.** The Dutch version of the CVLT (Delis, Kramer, Kaplan, & Ober, 1987) was used in dataset 1. The indices calculated in the Dutch version are different compared to the indices in the original CVLT (Mulder, Dekker, & Dekker, 1996). However, the administration and the data obtained are identical. Millis and Putman (1995) identified the following five indices suitable for examining noncredible

Table 3. Characteristics of the studies from which cutoffs were determined

Test	Cutoff	Specificity	Sensitivity	N	Population	External incentive for poor effort ¹	Age Mean (SD)	Years of education Mean (SD)	Reference
CVLT									
Total (total correct)	≤ 34	91%	74%	23	Mild head injury patients	Yes	38.7 (12.9)	11.6 (1.5)	Millis and Putman, 1995
Recognition (total correct)	≤ 6	87%	96%	23	Severe head injury patient	No	38.6 (13.2)	11.5 (1.4)	
LTCR (total correct)	≤ 10	91%	83%						
Discrimination (percentage)	≤ 80	91%	96%						
Formula	> 0	91%	91%						
COWA									
Total Correct T-score	< 32	89%	36%	108	Mild TBI patients	Mixed	40.9 (11.8)	12.5 (3.3)	Curtis et al., 2008
				96	Moderate/severe TBI patients	Mixed	38.2 (15.6)	12.1 (3.1)	
				488	General clinical patients	No	62.3 (15.6)	12.6 (3.2)	
Digit Span									
Reliable Digit Span	≤ 6	97%	38%	24	Mild TBI patients probable malingering	Yes	38.9 (12.3)	12.5 (2.8)	Mathias et al, 2002
				30	Mild TBI patients control	No	33.6 (13.5)	12.5 (2.4)	
ROCFT									
Copy (total correct)	≤ 26	89%	52%	146	Credible psychiatric and TBI patients	Yes	42.7 (13.7)	13.3 (2.8)	Reedy et al., 2013
Delay (total correct)	≤ 10	83.3%	45%	157	Noncredible psychiatric and TBI patients	No	44.2 (10.8)	12.5 (2.9)	
RAVLT									
Delayed Recognition (total correct)	≤ 9	90%	67%	61	Suspect effort patients	Yes	42.3 (13.1)	13.6 (2.5)	Boone et al., 2005
				88	Clinical patients	No	39.8 (13.5)	13.1 (2.7)	
				25	Controls	No	44.1 (19.0)	13.4 (1.8)	
Stroop									
Word-naming card (in seconds)	≥ 66	90.56%	53.49%	233	Credible psychiatric and TBI patients	No	43.3 (14.6)	13.3 (2.7)	Arentsen et al., 2013
Color-naming card (in seconds)	≥ 93	90.13%	48.82%	129	Noncredible psychiatric and TBI patients	Yes	43.6 (11.1)	12.7 (4.4)	
TMT									
Part A (in seconds)	≥ 63	98.8%	14.5%	328	Uncomplicated Mild brain injury patients	Mixed	31.6 (12.2)	12.3 (2.3)	Iverson et al., 2010
Part B (in seconds)	≥ 200	97.5%	11.6%	86	Mild brain injury with skull fracture patients	Mixed	30.2 (11.5)	11.7 (1.8)	
				117	Complicated Mild brain injury patients	Mixed	34.8 (14.4)	12.2 (2.0)	
				40	Moderate/severe brain injury patients	Mixed	34.7 (15.0)	11.8 (1.8)	
MMPI									
Fake Bad Scale	≥ 26	89.7%	NA	53	Non compensation seeking epilepsy patients	No	33.7 (10.3)	14.1 (2.7)	Nelson et al., 2006

¹Study participants include participants instructed to simulate, patients with and without credible performance, and patients evaluated in context of litigation. LTCR = Long term cued recall; NA = not available

- performance: CVLT Total (sum of the scores on List A trial 1 to 5); Long Delay Cued Recall (number of correct responses on List A LDCR), Discriminability (score on the following function $[1 - (\text{false positives} + \text{misses})/44] \times 100$); Recognition Hits (number of correct answers on List A Long Term Recognition, using the number of false positives and misses on List A Long Term Recognition) and Discriminant function score $((\text{CVLT Total raw score}) * (-0.00406) + (\text{LDCR raw score}) * (0.21099) + (\text{Discriminability}) * (0.04988) - 5.06399)$.
2. *Fluency*. Various versions of a letter fluency test were used among the different datasets. In dataset 2, the Dutch letter combination DAT was used (Schmand, Groenink, & van den Dungen, 2008) that corresponds with the US FAS-test. In dataset 4 to 6, the letter combinations of the Multilingual Aphasia Examination were used (Benton, Hamsher, & Sivan, 1994). Curtis, Thompson, Greve, and Bianchini (2008) defined a cutoff of a T-score below 32 as an indicator of possible noncredible performance.
 3. *Digit Span*. Using the WAIS-III (Wechsler, 1997) or WAIS-R (Wechsler, 1981) Digit Span subtest in dataset 3 to 6, the Reliable Digit Span (RDS) was calculated. The RDS is the sum of the longest span forward and the longest span backward. An RDS score below 7 has been suggested as a cutoff for possible noncredible performance (Mathias, Greve, Bianchini, Houston, & Crouch, 2002).
 4. *Rey-Osterrieth Complex Figure Test*. The ROCFT (Rey, 1941) was partly administrated in datasets 3 and 6. Dataset 3 contained information about the copy score and the 3-min delay score of the ROCFT. Dataset 6 contained information only about the copy score. Reedy et al. (2002) identified a copy score of 26 or lower and/or a 3-min delay score of 10 or lower as an indication of possible noncredible performance.
 5. *Rey Auditory Verbal Learning Test*. Datasets 2 and 3 provided information about the recognition score of the RAVLT (van den Burg, Saan, & Deelman, 1985). A recognition score of 9 or lower has been recommended as an indicator of possible noncredible performance (Boone, Lu, & Wen, 2005).
 6. *Stroop test*. The Dutch datasets 1 to 3 contained information on the Stroop color-word test (Hammes, 1978). Based on a study by Arentsen et al. (2013) in which the superiority was demonstrated of Word Naming trials and Color Naming trials over an inverted Stroop effect as an effective PV measure, we choose the cutoff of ≥ 66 s for Word Naming and ≥ 93 s for Color Naming as indicators of possible noncredible performance, maintaining adequate specificity and sensitivity. In the Arentsen paper, the Comalli version of the Stroop Color Word Test was used. This version is highly comparable to the version used in the Dutch datasets (i.e., identical number of trials and items). The Dutch version included 4 response options compared to 3 in the Comalli version.
 7. *Trail Making Test (TMT)*. All datasets contained information on Part A and Part B of the TMT (Reitan, 1958). Iverson, Lange, Green, and Franzen (2002) defined a completion time of 63 s or higher on Part A and/or a completion time of 200 s or higher on Part B as an indicator of possible noncredible performance.
 8. *Hiscock and Hiscock*. Hiscock and Hiscock (1989) developed a stand-alone PVT using the forced-choice method. One dataset contained information from a modified Hiscock & Hiscock test where 20 cards were shown for 5 s with a 5-s delay between presentation and recognition conditions. A score below chance is regarded as an indicator of possible noncredible performance.
 9. *Minnesota Multiphasic Personality Inventory (MMPI)*. The FBS developed by Lees-Haley, English, and Glenn (1991) and derived from items on the MMPI (Greene, 1991) was intended to detect self-reported symptom magnification. Larrabee's (2003) regression formula was used to estimate the FBS score from the 370-item protocol for all patients in Dataset 5. A FBS score of ≥ 26 was used to identify possible symptom exaggeration (Nelson, Parsons, Grote, Smith, & Sisung, 2006).
 10. *Years of education*. In the Dutch datasets 1 to 3, Verhage education scores (Verhage, 1964) were used, according to Dutch standards (Bouma, Mulder, Lindeboom, & Schmand, 2012). The US datasets 4 to 6 describe years of education.
 11. *Fatigue*. In datasets 1 to 3, the fatigue scale of the EORTC-QLQ-30 (Aaronson, Ahmedzi, Bergman, Bullinger, & Cull, 1993) was administrated. The fatigue scale consists of three items. Scores range from 0 to 100, with higher scores indicating more symptoms. In dataset 6 patients were asked to rate their fatigue level on a 0–10 scale with 10 being the worst.
 12. *Anxiety and Depression*. In datasets 1 to 3, the Hopkins Symptom Checklist-25 (HSCL-25) (Hesbacher, Rickels, Downing, & Stepansky, 1978) was used to determine anxiety and depression. The HSCL-25 is a self-report measure of psychological distress. It includes an anxiety and depression subscale. The instructions inquire about the intensity of symptoms in the previous week. Answers are scored on a scale from 1 (never) to 4 (always). A higher score indicates greater distress.
- In dataset 4, the State-Trait Anxiety Inventory (STAI) (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) was used to determine anxiety. The STAI is a 20-item self-report measure to assess levels of situation-related anxiety. Items are rated on a 4-point Likert-type scale. The score is then converted into a Z-scale distribution based on comparisons with a healthy non-anxious sample. To assess depression, the Beck Depression Inventory (BDI) (Beck, 1996) was used. The BDI is a 21-item self-report questionnaire measuring symptoms of depression. Items are rated on a 4-point scale. A higher score indicates greater depressive symptoms.

In dataset 5, the MMPI was used to determine anxiety and depression. MMPI scale-2 measures depression and MMPI scale-7 measures anxiety. Each scale is converted into a standardized T-score. A higher score indicates more symptoms.

In dataset 6, the BDI (Beck, 1996) was used to determine depression. To determine anxiety the Beck Anxiety Inventory (BAI) (Beck, 1990) was used. The BAI is a 21-item self-report measure to assess the cognitive aspects of anxiety. Items are rated on 4-point scale. A higher score indicates higher anxiety levels.

Statistical Analysis

The Statistical Package for Social Sciences (SPSS) version 20.0 was used for all statistical analyses. For each PVT index, a dichotomous variable was created indicating performance above or below the cutoff. The percentage of patients and controls with performances below the cutoff for each test and between group differences in these percentages were analyzed using χ^2 tests. Possible noncredible performance at the subject level was operationally defined as performance below the cutoff on three or more PVT indices. As discussed in Larrabee (2007), combining PVTs improves specificity with evidence that failure on three indicators essentially ensures no false positive classifications. We applied this strict criterion to the measures as we primarily had these available in the datasets, the cutoffs for the measures were not derived in our study population, the subjects in our studies did not have incentive for external gain, and we wished to minimize risk of falsely declaring noncredible performance (as opposed to cognitive dysfunction) in a population where there is evidence of brain dysfunction associated with treatment. Only one embedded PVT index was counted for tests that yielded multiple embedded PVT measures (Larrabee, 2008) with preference given to including PVT indices below the cutoff. Logistic regression analyses were performed to examine group differences. Additional factors including age, education, fatigue, anxiety and depression were also examined. Assuming a low but non-negligible base rate of noncredible performance of 10%, likelihood ratios, pre- and post-test odds and posttest probabilities were calculated using the sensitivity and specificity of the applied cutoffs as depicted in the literature, according to the method described by Larrabee (2008). Using chained likelihood ratios, comparisons were made between posttest probabilities of the different combinations of PVTs in each dataset.

RESULTS

In total, 534 breast cancer patients and 214 healthy controls were included in the current analysis. The percentage of patients that performed below criterion on each PVT index and the percentage of subjects classified as exhibiting possible noncredible performance are listed in Table 4 (datasets 1 to 3) and Table 5 (datasets 4 to 6). Overall, only 1 patient was classified as exhibiting possible noncredible performance. None of the control subjects met the criteria for possible noncredible performance.

In dataset 1, 0–7.1% of participants performed below criterion on any PVT. PVTs with at least one patient meeting criterion included: CVLT Total index, CVLT Recognition index, CVLT Recognition, CVLT Discriminability, CVLT Long Term Cued Recall index, Stroop A and Stroop B, and TMT Part A and B. At both assessment times, no group differences were found on any PVTs or on the frequency of possible noncredible performance.

In dataset 2, 0–21.2% of participants performed below criterion on a PVT. PVTs with at least one patient meeting the criterion included: COWA, Stroop A and B and TMT Part A and B. At the first assessment, significant differences between groups were found on COWA ($\chi^2 = 8.1$; $p = .017$; Tamoxifen > Control, $p = .032$), TMT Part A ($\chi^2 = 10.1$; $p = .006$; Exemestane > Control, $p = .003$), and TMT Part B ($\chi^2 = 6.2$; $p = .044$; Exemestane > Control, $p = .025$). Logistic regression analysis indicated no associations between age, level of education, fatigue, anxiety or depression, and PVT classification. At the second assessment, no significant group differences were found. For this particular set of older subjects we conducted an exploratory analysis of the TMT using an age-adjusted cutoff of 3 standard deviations below the normative mean. This cutoff corresponded to a TMT Part A performance > 87 s, and to a TMT Part B performance of > 260 s. Table 4 shows the rates of subjects failing this age adjusted PVT classification.

In dataset 3, 0% to 7.7% of participants performed below criterion on a PVT. PVTs with at least one patient meeting the criterion included: ROCFT Delayed Recall, Stroop A and B and the TMT Part A. No significant group differences on any PVTs or on frequency of possible noncredible performance were found.

In dataset 4, 0% to 4.8% of patients performed below criterion on a PVT. At the first assessment, PVTs with at least one patient performing below criterion included: COWA, Digit Span, and TMT Part A and B. One patient met the definition of noncredible performance. At the second, third, and fourth assessment, patients performed below the PVT criterion on the TMT Part A, the Digit Span and on the COWA, respectively.

In dataset 5, no patients showed performance below criterion on a PVT or the FBS scale at the first and third assessment time. At the second assessment, patients performed below criterion on Digit Span. Additionally, one patient at this time point scored above the cutoff for the FBS. However, examination of the cognitive test results found her to be stable or improved on 21 of 22 cognitive tests compared to the first assessment.

In dataset 6, 0% to 12.9% of patients performed below criterion on a PVT. PVTs with at least one patient below criterion included: COWA, ROCFT Copy, and TMT Part A and B.

Table 6 shows the diagnostic statistics and probabilities of noncredible performance for failure on three or more PVT measures. The post-probability of noncredible performance is on average well above .95, ranging from .918 to .999. Inspection of the data suggests that this conclusion remains

Table 4. Percentage of patients in NKI datasets performing below PVT criterion

		CVLT Total	CVLT LTCR	CVLT Recognition	CVLT Discriminability	CVLT Function	COWAT	Digit Span	ROCFT copy	ROCFT delay	RAVLT	Stroop A	Stroop B	TMTA [#]	TMTB [#]	Noncredible performance
Dataset 1																
M1	CTC (n = 28)	3.6	0	0	0	0	x	x	x	x	x	7.1	0	0	0	0
	FEC (n = 39)	0	0	2.6 ¹	0 ¹	0 ¹	x	x	x	x	x	0	2.6	0	0	0
	No CT (n = 57)	0	0 ¹	3.6 ¹	1.8 ¹	0 ¹	x	x	x	x	x	1.8	0	0	1.8	0
	Control (n = 60)	1.7 ¹	0	0	1.7	0 ¹	x	x	x	x	x	0	0	1.7	0	0
M2	CTC (n = 28)	0	3.6	0	0	0	x	x	x	x	x	0	0	0 ¹	0 ¹	0 ¹
	FEC (n = 39)	0	0	0	0	0	x	x	x	x	x	2.6	2.6	0	0	0
	No CT (n = 57)	0	1.8	0	0	0	x	x	x	x	x	1.8	1.8	1.8 ¹	1.8 ¹	0 ¹
	Control (n = 60)	0	0	0 ¹	0 ¹	0 ¹	x	x	x	x	x	0	0	0	0	0 ¹
Dataset 2																
M1	TAM (n = 80)	x	x	x	x	x	8.9 ^{1*}	x	x	x	0	5	1.2	12.5* (1.2)	7.5* (2.5)	0 ¹
	EME (n = 99)	x	x	x	x	x	2*	x	x	x	0	4	2	21.2* (3.0)	9.1* (3.0)	0 ¹
	Control (n = 120)	x	x	x	x	x	1.7*	x	x	x	0 ²	3.3	0.8	6.7* (1.7)	1.7* (0.8)	0 ¹
M2	TAM (n = 80)	x	x	x	x	x	7.6 ¹	x	x	x	0	6.2	3.8	16.2 (2.5)	8.8 (1.2)	0 ¹
	EME (n = 99)	x	x	x	x	x	2	x	x	x	0 ²	3	2	11.1 (3.0)	5.1 (3.0)	0 ²
	Control (n = 120)	x	x	x	x	x	1.7	x	x	x	0 ¹	1.7	0	6.7 (0)	1.7 (0.8)	0 ¹
Dataset 3																
	CTC (n = 34)	x	x	x	x	x	x	0	0	5.9	0	5.9	0	5.9	0	0
	FEC (n = 36)	x	x	x	x	x	x	0	0	5.6	0	5.6	2.8	0	0	0
	CMF (n = 39)	x	x	x	x	x	x	0	0	7.7	0	0	0	2.6	0	0
	Control (n = 34)	x	x	x	x	x	x	0	0	0	0	2.9	0	2.9	0	0

¹Missing data on 1 patient.²Missing data on 2 patients.*Significant differences between groups ($p < .05$).

% of subjects who fail TMT PV measures when cutoffs are defined as 3 SD below age corrected normative mean (i.e., TMT A >87 s; TMT B > 260 s).

LTCR = Long term cued recall; Stroop a = word-naming; Stroop b = color-naming. CTC = cyclophosphamide, thiotepa, carboplatin; FEC = 5-FU, epirubicin, cyclophosphamide; CMF = cyclophosphamide, MTX, 5-FU; TAM = tamoxifen; EXE = exemestane.

Table 5. Percentage of patients in MDACC datasets performing below PVT criterion

	n	COWA	Digit Span	ROCFT Copy	TMTA	TMTB	Hiscock & Hiscock	Noncredible performance
Dataset 4								
Baseline	42	2.4	2.4	x	4.8	4.8	x	2.4
Eval 2	29	0	0	x	3.4	0	x	0 ¹
Eval 3	33	0	3.1 ¹	x	0	0	x	0 ²
Eval 4	30	3.3	0 ¹	x	0	0	x	0 ¹
Dataset 5								
Baseline	18	0	0	x	0	0	x	0
Eval 2	18	0	5.6	x	0	0	x	0
Eval 3	15	0	0 ¹	x	0	0	x	0 ¹
Dataset 6								
Baseline	62	8.1	0	12.9	3.2	1.6	0	0
Eval 2	49	2.1 ¹	0	10.2	0	0	0	0
Eval 3	40	0 ¹	0	5.0	0	2.5	0	0 ¹

¹Missing data on 1 patient.

²Missing data on 2 patients.

FAC = 5-fluorouracil, doxorubicin, cyclophosphamide TAM = tamoxifen.

true when taking expected intercorrelations between PVTs into account; that is, post-test probabilities are still well above .95 for combinations of PVTs that are derived from different performance domains.

DISCUSSION

Cognitive changes in patients with cancer outside the CNS (e.g., breast cancer) and who receive therapy that is not directed at the CNS (e.g., chemotherapy) have been reported to demonstrate cognitive dysfunction on neuropsychological tests. However, there are no reports to date examining the potential contribution of noncredible performance. Ruling out this potential explanatory factor for the observed deficits in cognitive function before treatment and the decline in cognitive function after treatment would strengthen the conclusion that non-CNS cancer and therapies not targeting the CNS can have untoward effects on cognitive function.

The current study examined 534 breast cancer patients and 214 healthy controls from six datasets of studies on the effects of chemotherapy and/or endocrine therapy on cognitive functioning to determine if there was evidence for noncredible performance. Out of 917 assessments, only one patient was classified as exhibiting suspected noncredible performance at the pre-chemotherapy time point. No evidence of noncredible performance was found for any other time points in any other patient. No healthy controls were classified as exhibiting suspected noncredible performance at any time point.

Patients' report of mood disturbance and affective distress was, as expected, quite varied with reports ranging from no distress to severe symptoms of depression/anxiety. The occurrence of noncredible performance was near absent despite the presence of an array of psychological and clinical symptoms that could influence effort/motivation, that is, an unbiased response. Similarly, there was no evidence that the current sample of healthy controls, who may not have any personal investment in the research study at hand, was prone to noncredible performance. These findings should not be interpreted as indicating that psychological and clinical symptoms have no impact on test performance, as this was not addressed by these analyses. However, in prior publications based on these subjects (Schagen et al., 2006; Schilder et al., 2010; Schagen et al., 1999; Van Dam et al., 1998; Wefel et al., 2004; Wefel et al., 2010) as well as other studies in this area (O'Farrell et al., 2013), the relationship between distress, fatigue, and cognition is minimal.

There are several limitations to the current study. This was a retrospective analysis; the original studies were not designed to specifically assess noncredible performance; the majority of PVT measures were "embedded" rather than stand alone; and, for some embedded effort measures the preferred test condition was not administered limiting the sensitivity and specificity of those measures.

Furthermore, many of the PVT cutoffs were established in much younger populations with known or suspected mild traumatic brain injury and it is unlikely that the performance

Table 6. Diagnostic statistics and probabilities of noncredible performance for combinations of PVTs

Dataset	Test combination	Pre-test odds	Likelihood ratio	Post-test odds	Post-test probability
1	tot/tma/sta	11.04	5.67	62.55	0.98
	tot/tma/stb	11.04	4.95	54.60	0.98
	tot/tmb/sta	4.24	5.67	24.02	0.96
	tot/tmb/stb	4.24	4.95	20.97	0.95
	rec/tma/sta	27.86	5.67	157.86	0.99
	rec/tma/stb	27.86	4.95	137.80	0.99
	rec/tmb/sta	10.70	5.67	60.62	0.98
	rec/tmb/stb	10.70	4.95	52.91	0.98
	ltc/tma/sta	12.38	5.67	70.16	0.99
	ltc/tma/stb	12.38	4.95	61.24	0.98
	ltc/tmb/sta	4.75	5.67	26.94	0.96
	ltc/tmb/stb	4.75	4.95	23.52	0.96
	dis/tma/sta	14.32	5.67	81.15	0.99
	dis/tma/stb	14.32	4.95	70.84	0.99
	dis/tmb/sta	5.50	5.67	31.16	0.97
	dis/tmb/stb	5.50	4.95	27.20	0.96
	for/tma/sta	13.58	5.67	76.92	0.99
	for/tma/stb	13.58	4.95	67.15	0.99
	for/tmb/sta	5.21	5.67	29.54	0.97
	for/tmb/stb	5.21	4.95	25.78	0.96
2	cow/rav/tma	2.44	12.08	29.44	0.97
	cow/rav/tmb	2.44	4.64	11.30	0.92
	cow/rav/tma/sta	29.44	5.67	166.81	0.99
	cow/rav/tma/stb	29.44	4.95	145.62	0.99
	cow/rav/tmb/sta	11.30	5.67	64.06	0.98
	cow/rav/tmb/stb	11.30	4.95	55.92	0.98
	cow/rav/tma	2.44	12.08	29.44	0.97
	cow/rav/tmb	2.44	4.64	11.30	0.92
	cow/rav/tma/sta	29.44	5.67	166.81	0.99
	cow/rav/tma/stb	29.44	4.95	145.62	0.99
	cow/rav/tmb/sta	11.30	5.67	64.06	0.98
	cow/rav/tmb/stb	11.30	4.95	55.92	0.98
3	rds/rav/cop	9.43	4.73	44.58	0.98
	rds/rav/del	9.43	2.69	25.41	0.96
	rds/rav/cop/tma	44.58	12.08	538.63	1.00
	rds/rav/cop/tmb	44.58	4.64	206.83	1.00
	rds/rav/del/tma	25.41	12.08	307.03	1.00
	rds/rav/del/tmb	25.41	4.64	117.90	0.99
	rds/rav/cop/tma/sta	538.63	5.67	3052.06	1.00
	rds/rav/cop/tma/stb	538.63	4.95	2664.24	1.00
	rds/rav/cop/tmb/sta	206.83	5.67	1171.99	1.00
	rds/rav/cop/tmb/stb	206.83	4.95	1023.07	1.00
	rds/rav/del/tma/sta	307.03	5.67	1739.71	1.00
	rds/rav/del/tma/stb	307.03	4.95	1518.65	1.00
	rds/rav/del/tmb/sta	117.90	5.67	668.05	1.00
	rds/rav/del/tmb/stb	117.90	4.95	583.16	1.00
4	cow/rds/tma	4.61	12.08	55.66	0.98
	cow/rds/tmb	4.61	4.64	21.37	0.96
5	cow/rds/tma/cop	55.66	4.73	263.10	1.00
	cow/rds/tmb/cop	21.37	4.73	101.03	0.99
6	cow/rds/tma/cop	55.66	4.73	263.10	1.00
	cow/rds/tmb/cop	21.37	4.73	101.03	0.99

Tot = California Verbal Learning Test Total; rec = CVLT recognition; ltc = CVLT Long Term Cued Recall; dis = CVLT discriminability. for = CVLT Formula; tma = Trail Making Test part A; tmb = Trail Making Test part B; sta = Stroop Word-naming card; stb = Stroop Color-naming card; cow = COWA T; rav = Rey Auditory Verbal Learning Test; rds = Reliable Digit Span; cop = Rey-Osterreith Complex Figure Test Copy; del = ROCFT 3-min delay.

characteristics of these cut points behave as well in our older medically ill population (Heilbrunner et al., 2009). This is especially evident in Dataset 2 where the subjects on average

were in their mid to late-60s and performance below the criterion for the Trail Making Tests were much more frequent. The criterion for the Trail Making Tests was derived

from a much younger population that were on average in their early 30 s. Measures dependent on processing speed (such as the Trail Making Tests) are known to be more vulnerable to age-related cognitive decline (Robins Wahlin, Bäckman, Wahlin, Winblad, 1996; Rodriguez-Aranda & Martinussen, 2010). Thus, using the Trail Making Tests as an embedded effort measure with conventional cutoff criterion such as described in this study leads to an overestimation of noncredible performance. Furthermore, several studies that examined the TMT as an embedded PVT have reported unsatisfactory diagnostic accuracy and have cautioned against its use (Busse & Whiteside, 2012; Iverson et al., 2002; Powell, Locke, Smigielski, & McCrea, 2011). Below criterion performance on the ROCFT was the next most common pattern. However, when compared to the results of Reddy et al. (2013) from which the cutoffs were derived, there was no difference between rates of possible noncredible performance (12.9% in the current study compared to 11% in Reddy et al.).

An additional potential limitation is that the PVT cutoffs we used were based on studies conducted on English speaking participants in the United States. However, half the datasets were from studies conducted in the Netherlands that used different versions or translated versions of tests. One particular notable example of this difficulty is for the measures of lexical fluency. The cutoff for the lexical fluency measures was based on the English version of the FAS lexical fluency test (Curtis et al., 2008). The US datasets contained the Multilingual Aphasia Examination Controlled Oral Word Association test, while the Dutch datasets contained the letter combination DAT, which corresponds with the US FAS-lexical fluency measure. In addition to possible cross-cultural differences in lexical fluency performances on these tests, the English FAS and COWA have been demonstrated not to be equivalent (Barry, Bates, & Labouvie, 2008). Although the cutoff was based on the standardized score rather than the raw score, it is unclear if the diagnostic accuracy of this cutoff remains the same on different fluency measures and across cultures. That being said, the frequency of below criterion performance on the fluency tests was not dissimilar across most datasets.

Slick, Sherman, and Iverson (1999) proposed two or more types of evidence of noncredible performance from neuropsychological testing in combination with presence of a substantial external incentive as criteria of probable malingering neurocognitive dysfunction. However, the breast cancer participants and healthy controls in the datasets we investigated did not have any known external incentives that may have contributed to noncredible performance.

In his influential study of 2008, Larrabee demonstrates that ranges of post-test probability of noncredible performance vary greatly for one failed PVT, using chaining of likelihood ratios. This variation lessens for two failed PVT and is quite restricted when three PVT are failed. He applied this method to a sample of litigating or compensation-seeking subjects and found posttest probabilities of malingering ranging from .989 to .995 at various base rates. This suggests that even at

low base rates of non-credible performance, failure on three or more PVT provides a high probability that the diagnosis of noncredible performance is correct. In the current study, we applied this method to our data, where the base rate for noncredible performance is low, and observed similar high chances of detecting noncredible performance, with post-test probabilities approaching nearly 1.00.

Based on the preponderance of the evidence, there is no suggestion that breast cancer patients who chose to participate in research studies examining cognitive function exhibit noncredible performance. Similarly, healthy controls who chose to participate in research studies examining cognitive function also did not exhibit test performances indicative of noncredible performance. It does not appear to be essential for research studies within this population to use PVTs on a routine basis.

ACKNOWLEDGMENT

The information in this manuscript has not been previously published either electronically or in print. The authors have no financial conflicts of interest. We received no financial support for this project.

REFERENCES

- Aaronson, N.K., Ahmedzi, S., Bergman, B., Bullinger, M., & Cull, A. (1993). The European Organization for Research and Treatment of Cancer QLQ-30; A quality of life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute*, *85*, 365–376.
- Ahles, T.A., Root, J.C., & Ryan, E.L. (2012). Cancer- and cancer treatment-associated cognitive change: An update on the state of the science. *Journal of Clinical Oncology*, *30*(30), 3675–3686.
- Arentsen, T.J., Boone, K.B., Lo, T.Y., Goldberg, H.E., Cottingham, M.E., Victor, T.L., ... Zeller, M.A. (2013). Effectiveness of the Comalli Stroop Test as a measure of negative response bias. *The Clinical Neuropsychologist*, *27*(6), 1060–1076.
- Barry, D., Bates, M.E., & Labouvie, E. (2008). FAS and CFL forms of verbal fluency differ in difficulty: A meta-analytic study. *Applied Neuropsychology*, *15*(2), 97–106.
- Beck, A.T. (1990). *Beck Anxiety Inventory*. San Antonio, TX: The Psychological Corporation.
- Beck, A.T. (1996). *Beck Depression Inventory-II*. San Antonio, TX: The Psychological Corporation.
- Benton, A.L., Hamsher, K., & Sivan, A.B. (1994). *Multilingual Aphasia Examination* (3rd ed.). Iowa City, IA: AJA Associates.
- Boone, K.B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, *23*, 729–741.
- Boone, K.B., Lu, P., & Wen, J. (2005). Comparison of various RAVLT scores in the detection of noncredible memory performance. *Archives of Clinical Neuropsychology*, *20*, 301–319.
- Bouma, A., Mulder, J., Lindeboom, J., & Schmand, B. (2012). *Handleiding neuropsychologische diagnostiek*. Amsterdam, the Netherlands: Pearson Assessment and Information.
- Busse, M., & Whiteside, D. (2012). Detecting suboptimal cognitive effort: Classification accuracy of the Connor's Continuous Performance Test-II, Brief Test of Attention, and Trail Making Test. *The Clinical Neuropsychologist*, *26*(4), 675–687.

- Curtis, K.L., Thompson, L.K., Greve, K.W., & Bianchini, K.J. (2008). Verbal fluency indicators of malingering in traumatic brain injury. *The Clinical Neuropsychologist*, 22(5), 930–945.
- Dandachi-Fitzgerald, B. (2013). *Symptom validity tests in neuropsychology: No excuse for not using them*. Paper presented at the ESN/GNP Meeting, Berlin, Germany, 2013. Abstract 197.
- Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B.A. (1987). *The California Verbal Learning Test*. San Antonio, TX: Psychological Corporation.
- Green, P. (2003). *Green's Word Memory Test for Microsoft Windows*. Edmonton, Alberta: Green's Publishing Inc.
- Greene, R.L. (1991). *The MMPI-2/MMPI: An interpretive manual*. Boston: Allyn and Bacon.
- Hammes, J.G. (1978). *De Stroop kleur-woord test. Handleiding, tweede gewijzigde druk*. Lisse, the Netherlands: Swets & Zeitlinger.
- Heilbronner, R.L., Sweet, J.J., Morgan, J.E., Larrabee, G.J., Millis, S.R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23, 1093–1129.
- Hesbacher, P.T., Rickels, K., Downing, R.W., & Stepansky, P. (1978). Assessment of psychiatric illness severity by family physicians. *Social Science & Medicine*, 12, 45–47.
- Hiscock, M., & Hiscock, C.K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology*, 11(6), 967–974.
- Iverson, G.L., Lange, R.T., Green, P., & Franzen, M.D. (2002). Detecting exaggeration and malingering with the Trail Making Test. *The Clinical Neuropsychologist*, 16(3), 398–406.
- Larrabee, G.J. (2007). *Assessment of malingered neuropsychological deficits*. New York, NY: Oxford University Press.
- Larrabee, G.J. (2003). Detection of symptom exaggeration with the MMPI-2 in litigants with malingered neurocognitive dysfunction. *The Clinical Neuropsychologist*, 17, 54–68.
- Larrabee, G.J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, 22(4), 666–679.
- Larrabee, G.J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychology Society*, 18(4), 625–630.
- Lees-Haley, P.R., English, L.T., & Glenn, W.J. (1991). A fake bad scale on the MMPI-2 for personal injury claimants. *Psychological Reports*, 68, 203–210.
- Mathias, C.W., Greve, K.W., Bianchini, K.J., Houston, R.J., & Crouch, J.A. (2002). Detecting malingered neurocognitive dysfunction using the reliable digit span in traumatic brain injury. *Assessment*, 9(3), 301–308.
- Millis, S.R., & Putman, S.H. (1995). The California Verbal Learning Test in the detection of incomplete effort in neuropsychological evaluation. *Psychological Assessment*, 7(4), 463–471.
- Mulder, J.L., Dekker, R., & Dekker, P.H. (1996). *Verbale Leer en Geheugen Test. Handleiding*. Lisse, the Netherlands: Swets & Zeitlinger.
- Nelson, N.W., Parsons, T.D., Grote, C.L., Smith, C.A., & Sisung, J.R., II (2006). The MMPI-2 Fake Bad Scale: Concordance and specificity of true and estimated scores. *Journal of Clinical and Experimental Neuropsychology*, 28, 1–12.
- O'Farrell, E., Mackenzie, J., & Collins, B. (2013). Clearing the air: A review of our current understanding of "Chemo Fog". *Current Oncology Reports*, 15(3), 260–269.
- Phillips, K.A., Ribbi, K., Sun, Z., Stephens, A., Thompson, A., Harvey, V., ... Bernhard, J. (2010). Cognitive function in postmenopausal women receiving adjuvant letrozole or tamoxifen for breast cancer in the BIG 1-98 randomized trial. *Breast*, 19(5), 388–395.
- Pomykala, K.L., de Ruiter, M.B., Deprez, S., McDonald, B.C., & Silverman, D.H. (2013). Integrating imaging findings in evaluating the post-chemotherapy brain. *Brain Imaging and Behavior*, 7, 436–452.
- Powell, M.R., Locke, D.E.C., Smigielski, J.S., & McCrea, M. (2011). Estimating the diagnostic value of the Trail Making Test for suboptimal effort in acquired brain injury rehabilitation patients. *The Clinical Neuropsychologist*, 25(1), 108–118.
- Reedy, S.D., Boone, K.B., Cottingham, M.E., Glaser, D.F., Lu, P.H., Victor, T.L., ... Wright, M.J. (2013). Cross validation of the Lu and colleagues (2003) Rey-Osterrieth Complex Figure Test effort equation in a large known-group sample. *Archives of Clinical Neuropsychology*, 28(1), 30–37.
- Reitan, R.M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, 8, 271–276.
- Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives de Psychologie*, 28, 286–340.
- Robins Wahlin, T.B., Bäckman, L., Wahlin, A., & Winblad, B. (1996). Trail Making Test performance in a community-based sample of healthy very old adults: Effects of age on completion time, but not on accuracy. *Archives of Gerontology and Geriatrics*, 22(1), 87–102.
- Rodenhuis, S., Bontenbal, M., Beex, L.V., Wagstaff, J., Richel, D.J., Nooij, M.A., ... de Vries, E.G. (2003). High-dose chemotherapy with hematopoietic stem-cell rescue for high-risk breast cancer. *The New England Journal of Medicine*, 349, 7–16.
- Rodriguez-Aranda, C., & Martinussen, M. (2010). Age-related differences in performance of phonemic verbal fluency measured by the Controlled Oral Word Association Task (COWAT): A meta-analytic study. *Developmental Neuropsychology*, 30(2), 697–717.
- Schagen, S.B., Muller, M.J., Boogerd, W., Mellenbergh, G.J., & van Dam, F.S. (2006). Change in cognitive function after chemotherapy: A prospective longitudinal study in breast cancer patients. *Journal of the National Cancer Institute*, 98(23), 1742–1745.
- Schagen, S.B., van Dam, F.S., Muller, M.J., Boogerd, W., & Lindeboom, J. (1999). Cognitive deficits after postoperative adjuvant chemotherapy for breast carcinoma. *Cancer*, 85(5), 640–650.
- Scherling, C.S., & Smith, A. (2013). Opening up the window into "Chemobrain": A neuroimaging review. *Sensors*, 13, 3169–3203.
- Schilder, C.M., Seynaeve, C., Beex, L.V., Boogerd, W., Linn, S.C., Gundy, C.M., ... Schagen, S.B. (2010). Effects of tamoxifen and exemestane on cognitive functioning of postmenopausal patients with breast cancer: Results from the neuropsychological side study of the tamoxifen and exemestane adjuvant multinational trial. *Journal of Clinical Oncology*, 28(8), 1294–1300.
- Schmand, B., Groenink, S.C., & van den Dungen, M. (2008). Letterfluency: Psychometric properties and Dutch normative data. *Tijdschrift voor Gerontologie en Geriatrie*, 39(2), 64–76.
- Schmand, B., & Lindeboom, J. (2003). *Amsterdam Short-Term Memory Test (ASTM)*. PITS: Psychologische Instrumenten, Tests en Services.
- Seigers, R., & Fardell, J.E. (2011). Neurobiological basis of chemotherapy induced cognitive impairment: A review of

- rodent research. *Neuroscience and Biobehavioral Reviews*, 35, 729–741.
- Seigers, R., Schagen, S.B., Van Tellingen, O., & Dietrich, J. (2013). Chemotherapy-related cognitive dysfunction: Current animal studies and future directions. *Brain Imaging Research*, 7, 453–459.
- Slick, D.J., Sherman, E.M.S., & Iverson, G.L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13(4), 545–561.
- Spielberger, C.D., Gorsuch, R.L., Lushene, R., Vagg, P.R., & Jacobs, G.A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Strauss, E., Sherman, E.M.S., & Green, O. (2006). Assessment of response bias and suboptimal performance. In *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed., pp. 1145–1189). New York, NY: Oxford University Press.
- Tombaugh, T. (1996). *The test of memory malingering (TOMM)*. North Tonawanda, NY: Multi Health Systems.
- Van Dam, F.S., Schagen, S.B., Muller, M.J., Boogerd, W., vd Wall, E., Droogleever Fortuyn, M.E., & Rodenhuis, S. (1998). Impairment of cognitive function in women receiving adjuvant treatment for high-risk breast cancer: High-dose versus standard-dose chemotherapy. *Journal of the National Cancer Institute*, 90(3), 210–218.
- Van den Burg, W., Saan, R.J., & Deelman, B.G. (1985). *15-woordentest. Provisional Manual*. Groningen, the Netherlands: University Hospital, Department of Neuropsychology.
- Verhage, F. (1964). *Intelligentie en leeftijd. Onderzoek bij Nederlanders van twaalf tot zeventenzeventig jaar*. Assen, the Netherlands: Van Gorcum.
- Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale* (3rd ed). San Antonio, TX: Psychological Corporation.
- Wefel, J. S. (2002, August). Psychoneuroendocrinology in breast cancer: The effect of adjuvant tamoxifen therapy on cognitive and emotional functioning in women with early breast cancer. A dissertation presented at the faculty of the department of psychology, University of Houston, TX, United States.
- Wefel, J.S., Lenzi, R., Theriault, R.L., Davis, R.N., & Meyers, C.A. (2004). The cognitive sequelae of standard-dose adjuvant chemotherapy in women with breast carcinoma: Results of a prospective, randomized, longitudinal trial. *Cancer*, 100(11), 2292–2299.
- Wefel, J.S., Saleeba, A.K., Buzdar, A.U., & Meyers, C.A. (2010). Acute and late onset cognitive dysfunction associated with chemotherapy in women with breast cancer. *Cancer*, 116(14), 3348–3356.
- Wefel, J.S., & Schagen, S.B. (2012). Chemotherapy-related cognitive dysfunction. *Current Neurology and Neuroscience Reports*, 12(3), 267–275.
- Wefel, J.S., Vardy, J., Ahles, T., & Schagen, S.B. (2011). International Cognition and Cancer Task Force recommendations to harmonise studies of cognitive function in patients with cancer. *The Lancet Oncology*, 12(7), 703–708.