HHS Public Access

Author manuscript

Psychooncology. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

Psychooncology. 2016 January ; 25(1): 43–50. doi:10.1002/pon.3799.

Reliable change in neuropsychological assessment of breast cancer survivors

Charissa Andreotti[1,*], **James C. Root**[1], **Sanne B. Schagen**[2], **Brenna C. McDonald**[3], **Andrew J. Saykin**[3], **Thomas M. Atkinson**[1], **Yuelin Li**[1], and **Tim A. Ahles**[1]

[1]Department of Psychiatry and Behavioral Sciences, Memorial Sloan-Kettering Cancer Center, New York, NY, USA [2]Department of Psychosocial Research and Epidemiology, Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands [3]Indiana University Melvin and Bren Simon Cancer Center, Indiana University School of Medicine, Indianapolis, IN, USA

Abstract

Background—The purpose of this study was to enhance the current understanding and interpretation of longitudinal change on tests of neurocognitive function in individuals with cancer. Scores on standard neuropsychological instruments may be impacted by practice effects and other random forms of error.

Methods—The current study assessed the test–retest reliability of several tests and overarching cognitive domains comprising a neurocognitive battery typical of those used for research and clinical evaluation using relevant time frames. Practice effect-adjusted reliable change confidence intervals for test–retest difference scores based on a sample of patient-matched healthy controls are provided.

Results—By applying reliable change confidence intervals to scores from two samples of breast cancer patients at post-treatment follow-up assessment, meaningful levels of detectable change in cognitive functioning in breast cancer survivors were ascertained and indicate that standardized neuropsychological instruments may be subject to limitations in detection of subtle cognitive dysfunction over clinically relevant intervals, especially in patient samples with average to above average range baseline functioning.

Conclusions—These results are discussed in relation to reported prevalence of cognitive change in breast cancer patients along with recommendations for study designs that enhance detection of treatment effects.

[*]Correspondence to: Department of Psychiatry and Behavioral Sciences, Memorial Sloan-Kettering Cancer Center, 641 Lexington Avenue, 7th Floor, New York, NY 10022, USA. ahlest@mskcc.org.

Conflict of interest

The authors have declared no conflicts of interest.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.

Introduction

A growing body of research has provided evidence for cognitive change associated with adjuvant treatment for breast cancer [1]. However, inconsistencies remain, including widely varying prevalence rates (i.e., 0–77% post-treatment impairment), higher prevalence rates based on patient self-report compared with objective neuropsychological assessments, and discrepancies in the level of severity of cognitive change (i.e., survivor reports of inability to return to work or school versus relatively subtle or absent changes based on neuropsychological test performance). Variability in prevalence has been attributed to differences in study design, test batteries used, variation in treatment regimens, and differences in sample characteristics. Self-report of cognitive function has also been questioned due to the influence of psychological factors, such as depression and anxiety. However, another source of variation may relate to basic psychometric properties of neuropsychological tests and their sensitivity to detecting relatively subtle change, particularly within the normal range of cognitive function.

The standardized neuropsychological instruments commonly used to measure cognitive change in individuals with cancer are often those developed originally to determine lesion location and impairment in patients with overt neurological injuries and illnesses, such as traumatic brain injury or degenerative dementing conditions. The degree of impairment accompanying these conditions is often severe [2,3], particularly compared with neurocognitive effects expected following cancer treatment. Further, test–retest reliability data for many of these measures are only available for shorter durations (e.g., 1–3 weeks), and only limited data exist over more extended time frames of greater clinical or research relevance to cancer patients (e.g., 6 or more months). However, the potential implications of measurement-related error for the use of these measures in research and clinical evaluation of cancer-related cognitive decline have been largely unexplored in the cancer context. That is, the use of these same neuropsychological instruments in cancer-treated samples may be limited because of the test–retest reliability as well as ceiling effects, restricted range of test scores, and low sensitivity in samples with average range (or above) premorbid cognitive abilities and potentially subtle cognitive changes [4].

The purpose of this study was therefore two-fold. Because scores on standard neuropsychological instruments may be impacted by several factors, including true changes in performance, practice effects, regression towards the mean, and random measurement error, interpretation of change involves acknowledgment of the full range of measurement error for each test–retest difference interval. We first sought to assess the test–retest reliability of several tests comprising a neurocognitive battery typical of those used for research and clinical evaluation using time frames typical of longitudinal research studies and to provide reliable change confidence intervals for test–retest difference scores based on a sample of patient-matched healthy controls. Second, we sought to enhance the current understanding and interpretation of longitudinal change on tests of neurocognitive function in individuals with breast cancer. In order to examine effects of site, treatment type, and test–retest interval, we present analyses for two samples of patients and matched controls, one collected as part of a US study involving patients receiving chemotherapy and another collected as part of a study of endocrine therapy at a site in the Netherlands. By applying

reliable change confidence intervals to scores from these two samples of breast cancer patients at post-treatment follow-up assessment, meaningful levels of detectable change in cognitive functioning in breast cancer survivors were ascertained and are discussed in relation to reported prevalence of cognitive change in breast cancer patients.

Methods

Patients

Sample 1—Eligible patients were newly diagnosed patients with breast cancer recruited from the Breast Cancer Service of the Dartmouth-Hitchcock Norris Cotton Cancer Center as part of a longitudinal study of cognitive change in breast cancer survivors exposed to chemotherapy. Extended data on inclusion/exclusion criteria for study participation as well as sample characteristics have been described elsewhere [5]. Briefly, patients (*n*=60) were eligible for participation if they were diagnosed with noninvasive (stage 0) or invasive (stage 1, 2, or 3A) breast cancer, undergoing first treatment with systemic chemotherapy, between 18 and 70 years of age at time of diagnosis, and fluent in English and able to read English. Patients were excluded on the basis of the following criteria: central nervous system (CNS) disease; previous history of cancer (except basal cell carcinoma) or treatment with chemotherapy, CNS radiation, or intrathecal therapy; neurobehavioral risk factors, including history of neurologic disorder (e.g., Parkinson's disease, seizure disorder, and dementia), alcohol/substance abuse, or moderate to severe head trauma (loss of consciousness >60 min or structural brain changes on imaging); or Axis I psychiatric disorder (according to the Diagnostic and Statistical Manual of Mental Disorders, ed 4 [DSM-IV]; (e.g., schizophrenia, bipolar disorder, and depression).

Female healthy controls (*n*=45) who met the same inclusion (except for cancer diagnosis) and exclusion criteria were recruited through community advertisements. Healthy controls were frequency matched to patients on age and education. All methods and procedures were approved by the institutional review board of Dartmouth Medical School, and all participants provided written informed consent.

For patients, the pretreatment assessment occurred after surgery but before initiation of adjuvant therapy. Follow-up assessment for patients treated with chemotherapy was conducted 6 months after the baseline assessment, corresponding to approximately 1-month post-treatment completion. Because the length of chemotherapy varied, the test–retest interval for the follow-up assessment for healthy control participants was frequency matched to the interval for the chemotherapy patients. Analysis of the intervals between neuropsychological assessments by group revealed no differences. See Table 1 in the supporting information for full list of tests in the neuropsychological battery for Sample 1 [12–17].

Sample 2—Eligible patients were Dutch postmenopausal women participating in the tamoxifen exemestane adjuvant multinational (TEAM) trial; an international, open label, randomized study comparing the efficacy and safety of 5 years of adjuvant exemestane (25 mg/d; *n*=99) with 2.5 to 3 years of tamoxifen (20 mg/d; *n*=80) followed by 2 to 2.5 years of exemestane.

Additional information on inclusion/exclusion criteria of the TEAM trial as well as sample characteristics have been described elsewhere [17]. In short, patients had histologically confirmed adenocarcinoma of the breast, positive estrogen and/or progesterone receptor status, and had undergone surgery with a curative intent. For this neuropsychological side-study, additional exclusion criteria included the following: adjuvant chemotherapy, not being fluent in the Dutch language, and CNS disease or signs of dementia according to a dementia screening tool [18]. In order to take into account the test–retest effects of neuropsychological tests, a control group was included that consisted of healthy female friends or relatives age-matched to TEAM patients (n=120). Inclusion criteria for controls were postmenopausal status, no history of CNS or malignant disease, fluent in the Dutch language, and no signs of dementia according to the dementia screening tool. The study was approved by the central review board (Erasmus MC, Rotterdam, The Netherlands) and the local medical ethics committees of all participating hospitals. All participants provided written informed consent.

Initial neuropsychological assessments (T1) were performed after definite breast surgery, and immediately before the start of adjuvant endocrine treatment. This point in time was chosen in order to minimize potential effects of other treatments on cognition in the interval between T1 and T2. Follow-up assessments were conducted 1 year after the baseline assessment (T2). Healthy control participants underwent the same assessments with a similar time interval of 1 year. See Table 1 in the supporting information for full list of tests in neuropsychological battery for Sample 2 [15,18–25].

Statistical analysis

Descriptive statistics were calculated for the healthy control group for each test in the neuropsychological battery at baseline and follow-up time points and are presented in Tables 1 and 2. The descriptive statistics for the healthy control samples were used to calculate reliable change confidence intervals based on the procedure described by Jacobson and Truax [6]. According to this procedure, the standard error of measurement (SEM) from each of the baseline (SEM_1) and follow-up (SEM_2) testing sessions and the standard error of the difference (SE_{diff}) were used to compute the reliable change confidence intervals based on the following equation:

$$\mathrm{CI}=\mathrm{SE}_{\mathrm{diff}} \times 1.28(\text{z score for } 80\%\mathrm{CI});$$
$$\mathrm{CI}=\mathrm{SE}_{\mathrm{diff}} \times 1.96(\text{z score for } 95\%\mathrm{CI}).$$

Paired sample t-tests were then used to calculate repeat testing effects in each group, accounting for score improvement because of practice and procedural learning. For tests exhibiting significant repeat testing effects (p=< 0.05), mean improvements in the healthy control group were added to the confidence intervals.

These reliable change intervals were then applied to the patient and healthy control samples to determine the percentage of patients and controls that declined at both 80% and 95% confidence intervals. This procedure has been used in previous studies assessing sensitivity

of cognitive measures in novel populations (e.g., concussion [7], dementia [8], and cognitive status in the elderly [9]).

Results

Test–retest reliability

Pearson correlations indicating test–retest reliability between baseline and follow-up assessments for the healthy control groups of Samples 1 and 2 are shown in Tables 1 and 2 at the individual test and domain levels. For Sample 1, test–retest reliability at the level of individual measures ranged from 0.23 to 0.90 (mean=0.67) and from 0.64 to 0.89 (mean=0.74) at the domain level. For Sample 2, test–retest reliability at the level of individual measures ranged from 0.57 to 0.88 (mean=0.74) and from 0.60 to 0.89 (mean=0.80) at the domain level.

Longitudinal change in performance

As shown in Table 1, the healthy control group in Sample 1 exhibited significant improvement on several measures (i.e., Digit Symbol-Coding, Logical Memory I & II, Faces I & II, CVLT Total Trials 1–5, CVLT Long Delay Recall, CPT Vigilance Total Correct, Trail Making 3, Trail Making 4, and Trail Making 5) between the baseline and follow-up assessments ($p<0.05$). As shown in Table 2, the healthy control group in Sample 2 exhibited significant improvement on selected tests (i.e., RAVLT Delayed Recall, Visual Association Test) between the baseline and follow-up assessments ($p<0.05$).

Reliable change

Reliable change confidence intervals are presented in Tables 1 and 2 and include calculated practice effects for tests, which showed a significant performance improvement from baseline to follow-up assessments. Effect sizes (i.e., Cohen's Delta) are also presented in Tables 1 and 2 as a standardized metric signifying the magnitude of change for each interval. The standardized effect sizes corresponding to the 80% reliable change confidence intervals ranged from approximately 0.60 to 1.73 for Sample 1 and from 0.63 to 1.20 for Sample 2. The standard effect sizes corresponding to the 95% reliable change confidence intervals ranged from approximately 0.91 to 2.40 for Sample 1 and from 0.96 to 1.82 for Sample 2. These are considered 'medium' to 'very large' changes (0.2, small; 0.5, medium; 0.8, large; 1, very large) [10].

For Sample 1, when the 80% reliable change confidence interval was applied to each measure, the percentage of patients indicated as declined ranged from approximately 0% to 31%, and when the 95% reliable change confidence interval was applied to each measure, the percentage of patients indicated as declined ranged from approximately 0% to 22% (Table 3). For Sample 2, when the 80% reliable change confidence interval was applied to each measure, the percentage of patients indicated as declined ranged from approximately 4% to 19% for the TMX group and 3% to 15% for the EXE group, and when the 95% reliable change confidence interval was applied to each measure, the percentage of patients indicated as declined ranged from approximately 0% to 9% for the TMX group and 1% to 10% for the EXE group (Table 4).

Discussion

The present study sought to enhance the current understanding and interpretation of longitudinal change on tests of neurocognitive function in individuals with cancer. Using numerous tests comprising two comprehensive neurocognitive batteries typical of those used for research and clinical evaluation of cancer-related cognitive decline over clinically relevant (i.e., 6 months and 1 year) time frames, we calculated reliable change indices based on 80% and 95% confidence intervals, taking into account any significant practice effects for each individual test.

We believe the results of this analysis have implications for the design and analysis of future studies of cognitive function in cancer survivors. First, results indicated attenuated test–retest reliability at longer intervals (i.e., 6 months and 1 year) compared with published reliability values during standardization that are derived from shorter intervals (i.e., 1–3 weeks). Acceptable reliability values for standard neuropsychological measures are generally considered at $r>/=0.8$. In contrast, our analyses of two healthy control samples at extended, but perhaps more clinically- or research-relevant intervals, generally fell below this value with a subset of measures exhibiting reliability values as low as $r= 0.23$ to 0.35. This finding will have particular importance in detecting subtle cognitive dysfunction typical of cancer survivors. In order to detect meaningful cognitive change (i.e., the signal), differences in test scores will need to exceed the random measurement error inherent in each test (i.e., the noise). The range of random variation between time 1 and time 2 in our healthy control samples during which no change should be evident (i.e., the effect size from time 1 to time 2) represents medium to large effects, and treatment-related changes in cancer survivors post-treatment is generally expected to be much smaller.

Our results would therefore suggest that these measures could only reliably detect moderate to large changes in a given cognitive ability using a sizeable sample over a 6-month or 1-year time frame, and more subtle changes in ability may thus be lost in the ‘noise’ of a measure’s random sources of error. Further, when reliable change intervals were applied to patient samples, the percentage of patients exhibiting significant decline was generally lower than that of typically self-reported by breast cancer patients. It is of note that these findings were observed in two datasets comprised of patients from different assessment sites/countries (i.e., USA and the Netherlands), receiving different cancer treatments (i.e., chemotherapy and endocrine therapy), and tested across different intervals (i.e., 6 months and 1 year).

Second, single-arm study designs that rely on published test–retest reliability values for calculation of a reliable change index may overestimate decline in patient groups. Because published test–retest reliability at shorter time points is higher, confidence intervals for reliable change that are calculated from published reliability data will be reduced. As a result, change in performance over longer time periods that is due to random measurement error may be misidentified as true change in performance when relying on published reliability values. To address this, we recommend continued accrual of true test–retest reliability data at intervals similar to research study time points. More accurate reliable change indices can then be calculated for use in studies that collect only patient group

cognitive data. Even with this adjustment, however, true change in performance may be undetected because of large confidence intervals, and thus collection of a sizeable healthy control group may be preferable.

As this may prove burdensome, an alternative approach to overcome the measurement challenges we present is the aggregation of tests into cognitive domains. Here, we show that relying on confirmatory factor analytic approaches to aggregate individual measures into cognitive factor domains by summing standard scores (e.g., z-scores) of individual tests by domain can provide greater test–retest reliability and may thus reduce error in measurement and provide more accurate indications of decline in this population.

Third, the relation between test reliability and sensitivity may be linked to the specific pattern of cognitive dysfunction observed in previous studies of cancer survivors. As discussed, a majority of longitudinal studies indicate some degree of post-treatment decline, but results suggest that these subtle changes are limited to select cognitive domains. In the past, as in the current study, timed measures of psychomotor speed, specifically, have been found to be associated with treatment-related effects on cognition [5]. However, our results raise the possibility that such findings may be less related to specific patterns of cognitive function affected by treatment and, instead, potentially related to increased sensitivity resulting from enhanced reliability of psychomotor speed measures. That is, measures of psychomotor speed are more reliable and less subject to random ‘noise,’ as exhibited by the somewhat smaller effect sizes corresponding to the reliable change confidence interval that must be overcome to detect ‘meaningful change’ with a measure.

Lastly, a recent collection of studies suggests more substantial effects in specific high-risk subgroups that may currently be moderated by performance improvements because of positive practice effects on most standardized instruments in the majority of patients as seen in this study. For example, older patients with limited cognitive reserve exposed to chemotherapy as well as individuals carrying adverse genetic alleles (e.g., APOE ε4) are shown to be at significantly increased risk for post-treatment cognitive decline [5,11]. As such, future analyses taking into account sample characteristics are needed to ascertain which measures may be most sensitive to detection of treatment effects in vulnerable subgroups. Other primary confounds of cognitive function (e.g., sleep and mood) may also be assessed at each time point and used as covariates when examining cognitive trajectories in survivors.

In summary, neuropsychological measures remain the gold standard in assessing treatment-related cognitive changes and dysfunction in cancer survivors. Several observations from our analysis strongly support changes in study design and methods to improve the sensitivity of these measures to the subtle cognitive changes seen in treatment-related dysfunction. Chief among these are the importance of establishing reliability values and reliable change indices of cognitive measures at clinically meaningful intervals, assessment of practice effects at longer intervals to more realistically anticipate changes in performance, collection of a control group particularly when this information is not already available, and use of aggregate, domain-level performance scores to improve test–retest stability over time.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The data used in this study was collected as part of studies supported by Grants No. R01 CA87845, R01 CA101318, R01 CA129769, CA172119, U54CA137788 from the National Cancer Institute (Sample 1), and an independent research grant from Pfizer (Sample 2). The current study was supported by a T32 NCI Institutional Training Grant (CA009461) in Psycho-oncology.

References

1. Ahles TA, Root JC, Ryan EL. Cancer- and cancer treatment-associated cognitive change: an update on the state of the science. J Clin Oncol. 2012; 30(30):3675–3686. [PubMed: 23008308]
2. Hutchinson AD, Mathias JL. Neuropsychological deficits in frontotemporal dementia and Alzheimer's disease: a meta-analytic review. J Neurol Neurosurg Psychiatry. 2007; 78(9):917–928. [PubMed: 17371908]
3. Mathias JL, Wheaton P. Changes in attention and information-processing speed following severe traumatic brain injury: a meta-analytic review. Neuropsychology. 2007; 21(2):212–223. [PubMed: 17402821]
4. Shilling V, Jenkins V, Trapala IS. The (mis) classification of chemo-fog–methodological inconsistencies in the investigation of cognitive impairment after chemotherapy. Breast Cancer Res Treat. 2006; 95(2):125–129. [PubMed: 16261398]
5. Ahles TA, Saykin AJ, McDonald BC, et al. Longitudinal assessment of cognitive changes associated with adjuvant treatment for breast cancer: impact of age and cognitive reserve. J Clin Oncol. 2010; 28(29):4434–4440. [PubMed: 20837957]
6. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991; 59(1):12–19. [PubMed: 2002127]
7. Register-Mihalik JK, Guskiewicz KM, Mihalik JP, et al. Reliable change, sensitivity, and specificity of a multidimensional concussion assessment battery: implications for caution in clinical practice. J Head Trauma Rehabil. 2013; 28(4):274–283. [PubMed: 22691965]
8. Pedraza O, Smith GE, Ivnik RJ, et al. Reliable change on the dementia rating scale. J Int Neuropsychol Soc. 2007; 13(4):716–720. [PubMed: 17521486]
9. Hensel A, Angermeyer MC, Riedel-Heller SG. Measuring cognitive change in older adults: reliable change indices for the mini-mental state examination. J Neurol Neurosurg Psychiatry. 2007; 78(12): 1298–1303. [PubMed: 17442763]
10. Cohen, J. Statistical Power Analyses for the Behavioral Sciences. 2. Lawrence Erlbaum; New Jersey: 1988.
11. Ahles TA, Saykin AJ, Noll WW, et al. The relationship of APOE genotype to neuropsychological performance in long-term cancer survivors treated with standard dose chemotherapy. Psycho-Oncology. 2003; 12(6):612–619. [PubMed: 12923801]
12. Delis, JH.; Kramer, JH.; Kaplan, E.; Ober, BA. California Verbal Learning Test – (2nd edn). Adult Version Manual. Psychological Corporation; San Antonio, TX: 2000.
13. Gordon, MFD.; Aylward, GP. Gordon Diagnostic System: Instruction Manual and Interpretive Guide. Gordon Systems, Inc.; DeWitt, NY: 1986.
14. Delis, DC.; Kaplan, E.; Kramer, JH. Delis-Kaplan Executive Function System. The Psychological Corporation; San Antonio, TX: 2001.
15. Reitan, R.; Wolfson, D. The Halstead-Reitan Neuropsychological Test Battery. Neuropsychological Press; Tucson, AZ: 1985.
16. Wechsler, D. WAIS—III administration and scoring manual. 3. The Psychological Corporation; San Antonio, TX: 1997.
17. Wechsler, D. Weschsler Memory Scale. 4. Pearson Education; San Antonio, TX: 2008.

18. Alpherts, W.; Aldenkamp, A. FePsy: The Iron Psyche. Instituut voor Epilepsiebestrijding; Hemmstede, the Netherlands: 1994.
19. van den Burg, W.; Saan, RJ.; Dellman, BG. 15-woordentest: Provisional Manual. University Hospital, Department of Neuropsychology; Groningen, the Netherlands: 1985.
20. Hammes, J. De Stroop kleur-woord test: Handleiding. 2. Swets & Zeitlinger; Lisse, the Netherlands: 1978.
21. Reitan R. Validity of the trail making test as an indicator of organic brain damage. Perceptual Motor Skills. 1958; 8:271–276.
22. Van der Elst W, Dekker S, Hurks P, Jolles J. Normative data for the Animal, Profession and Letter M Naming verbal fluency tests for Dutch speaking participants and the effects of age, education, and sex. J Int Neuropsychol Soc. 2006; 12(1):80–89. [PubMed: 16433947]
23. Lindeboom J, Schmand B, Tulner L, Walstra G, Jonker C. Visual association test to detect early dementia of the Alzheimer type. J Neurol Neurosurg Psychiatry. 2002; 73(2):126–133. [PubMed: 12122168]
24. Wechsler, D. WAIS-III Nederlandstalige Bewerking. Swets & Zeitlinger; Lisse, the Netherlands: 2000.
25. Wechsler, D. Wechsler Memory Scale: Revised. Psychological Corporation; New York, NY: 1987.

Table 1

Descriptive statistics for Sample 1 healthy controls, test/domain reliability, and reliable change confidence intervals

		Baseline (n = 45)			Follow-up (n = 38)									
Domain	**Test**	**Mean**	**SD**	**SEM**	**Mean**	**SD**	**SEM**	***r***	**SE_{diff}**	***t*-test**	**80% CI (Raw)**	**95% CI (Raw)**	**80% CI ES**	**95% CI ES**
Verbal memory								0.77						
	CVLT-II total trials 1–5	55.84	8.03	4.09	57.79	8.49	4.3	0.74	5.93	−2.41*	9.55	13.58	1.16	1.65
	CVLT-II short delay	12.67	2.43	1.42	13.21	2.76	1.62	0.66	2.15	−1.63	2.76	4.22	1.07	1.63
	CVLT-II long delay	12.62	2.56	1.50	13.55	2.35	1.37	0.66	2.03	−2.72*	3.53	4.90	1.43	1.99
	CVLT-II recognition	15.27	1.07	0.95	15.58	0.72	0.64	0.23	1.14	−1.13	1.46	2.23	1.57	2.40
	WMS-III logical memory I	47.27	7.55	4.57	51.26	7.45	4.52	0.63	6.43	−3 93***	12.23	16.59	1.63	2.21
	WMS-III logical memory II	29.78	6.06	3.38	34.13	5.96	3.32	0.69	4.73	−5.64***	10.42	13.63	1.73	2.27
	WMS-III logical	27.80	1.58	1.01	27.74	1.52	0.98	0.59	1.41	−0.34	1.81	2.76	1.17	1.78
	Memory recognition													
Visual memory								0.75						
	WMS-III faces I	38.44	4.64	2.58	41.87	4.92	2.73	0.69	3.75	−6.16***	8.24	10.79	1.73	2.26
	WMS-III faces II	38.51	4.10	2.21	41.38	4.41	2.38	0.71	3.25	−5.80***	7.03	9.23	1.66	2.18
Processing speed								0.89						
	WAIS-III digit symbol-coding	81.73	15.94	4.94	85.37	17.88	5.54	0.90	7.42	−408***	13.16	18.19	0.78	1.08
	D-KEFS trails 1-visual scanning	20.47	6.07	2.83	20.22	5.57	2.60	0.78	3.84	0.50	4.92	7.53	0.84	1.29
	D-KEFS trails 2-number Sequencing	29.80	8.94	5.59	26.92	8.09	5.06	0.61	7.54	2.01	9.67	14.78	1.13	1.72
	D-KEFS trails 3-letter sequencing	29.13	9.71	6.53	25.86	10.47	7.04	0.55	9.60	2.05*	15.58	22.09	1.55	2.20
	D-KEFS trails 4-number–letter switching	64.84	26.40	13.82	58.95	22.57	11.82	0.73	18.18	2.26*	29.21	41.54	1.18	1.68
	D-KEFS trails 5-motor speed	23.80	10.76	4.55	21.35	9.75	4.12	0.82	6.14	3.06**	10.32	14.49	1.00	1.40
	D-KEFS color naming	27.24	5.03	2.38	26.61	5.03	2.38	0.78	3.36	1.23	4.31	6.59	0.86	1.31
	D-KEFS word reading	20.42	3.61	1.40	21.03	3.98	1.54	0.85	2.08	−1.53	2.67	4.08	0.71	1.08
	D-KEFS inhibition	50.89	10.93	3.61	49.74	10.52	3.47	0.89	5.01	1.44	6.42	9.82	0.60	0.91
	D-KEFS inhibition/switching	59.44	13.20	7.45	57.58	13.36	7.55	0.68	10.61	1.56	13.60	20.79	1.02	1.57
	Grooved pegboard (R)	70.78	16.92	6.15	70.27	19.95	7.25	0.87	9.50	0.23	12.18	18.63	0.66	1.02

		Baseline (n = 45)			Follow-up (n = 38)									
Domain	**Test**	**Mean**	**SD**	**SEM**	**Mean**	**SD**	**SEM**	***r***	**SE_{diff}**	***t*-test**	**80% CI (Raw)**	**95% CI (Raw)**	**80% CI ES**	**95% CI ES**
	Grooved pegboard (L)	75.98	20.41	9.58	74.59	21.29	9.98	0.78	13.83	−1.80	17.73	27.11	0.85	1.30
Simple vigilance								0.76						
	CPT vigilance total correct	39.59	0.87	0.69	29.83	0.51	0.40	0.37	0.80	−2.06*	10.79	11.33	1.47	1.55
	CPT vigilance reaction time	41.45	8.21	3.79	40.83	8.27	3.82	0.79	5.38	0.90	6.89	10.54	0.84	1.28
Distractibility								0.64						
	CPT distractibility total correct	26.42	5.28	3.29	27.76	4.02	2.51	0.61	4.14	−1.82	5.31	8.11	1.11	1.70
	CPT distractibility reaction time	42.37	7.69	3.60	41.85	7.44	3.48	0.78	5.01	0.24	6.42	9.82	0.85	1.29
Executive function														
	D-KEFS sorting-free sorting score	44.60	6.35	4.30	45.16	6.72	4.56	0.54	6.27	−0.53	8.03	12.28	1.23	1.88
Verbal Ability								0.74						
	WAIS-III vocabulary	68.51	5.87	3.23	68.87	5.85	3.22	0.70	4.56	−0.57	5.85	8.94	1.00	1.53
	D-KEFS animals	21.33	3.93	3.17	22.68	3.78	3.06	0.35	4.41	−1.64	5.65	8.64	1.46	2.24
	D-KEFS boys' names	23.27	3.41	2.18	22.47	4.49	2.88	0.59	3.61	1.09	4.63	7.08	1.18	1.80

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

CI, confidence interval; ES, effect size; SEM, standard error of measurement; SD, standard deviation. Significant differences between time 1 and time 2 performances on individual measures indicate a practice effect.

Table 2

Descriptive statistics for Sample 2 healthy controls, test/domain reliability, and reliable change confidence intervals

		Baseline (n = 120)			Follow-up (n = 120)									
Domain	**Test**	**Mean**	**SD**	**SEM**	**Mean**	**SD**	**SEM**	***r***	**SE_{diff}**	***t*-test**	**80% CI**	**95% CI**	**80% CI ES**	**95% CI ES**
Verbal memory								0.77						
	RAVLT total trials 1–5	22.20	5.70	3.27	22.90	5.40	3.10	0.67	4.51	−1.42	5.77	8.84	1.04	1.59
	RAVLT delayed recall	6.40	3.00	1.64	7.00	2.80	1.53	0.70	2.25	−2.35*	3.48	5.01	1.20	1.73
	Visual association test	20.90	2.80	1.53	21.40	2.60	1.42	0.70	2.09	−2.11*	3.18	4.10	1.18	1.52
Visual memory								0.83						
	WMS-R visual reproduction immediate	31.40	5.50	2.96	31.20	5.40	2.91	0.71	4.15	0.41	5.31	8.64	0.97	1.58
	WMS-R visual reproduction delayed	28.20	6.90	3.90	28.50	7.00	3.96	0.68	5.56	−0.47	7.12	10.90	1.02	1.57
Processing speed								0.89						
	Stroop word reading	46.60	7.40	3.39	47.50	7.10	3.25	0.79	4.70	−1.39	6.02	9.21	0.83	1.27
	Stroop color naming	61.00	10.00	3.46	60.30	9.60	3.33	0.88	4.80	0.80	6.15	9.41	0.63	0.96
	Trail making test A	42.20	14.00	6.57	40.50	14.10	6.61	0.78	9.32	1.32	11.93	18.27	0.85	1.30
Executive function								0.88						
	Stroop interference	107.50	32.50	13.00	104.70	33.80	13.52	0.84	18.76	0.91	24.01	36.76	0.72	1.11
	Trail making test B	97.00	40.60	21.10	93.00	36.60	19.02	0.73	28.40	1.20	36.36	55.67	0.94	1.44
Manual motor speed								0.88						
	Finger tapping (dominant)	52.40	10.00	3.74	53.10	8.80	3.29	0.86	4.98	−0.87	6.38	9.77	0.68	1.04
	Finger tapping (non-dominant)	48.10	8.60	3.10	49.00	7.90	2.85	0.87	4.21	−1.25	5.39	8.25	0.65	1.00
Verbal fluency								0.89						
	Category – animals	23.30	5.90	3.28	23.30	6.10	3.40	0.69	4.73	0.00	6.05	9.26	1.01	1.54
	Category – professions	19.40	5.20	2.85	19.10	5.60	3.07	0.70	4.19	0.59	5.36	8.20	0.99	1.52
	Letters (D, A, T)	39.30	11.70	4.53	40.60	12.00	4.65	0.85	6.49	−1.19	8.31	12.72	0.70	1.07
Reaction time								0.64						
	Reaction speed (dominant)	309.00	58.00	38.03	307.00	54.00	35.41	0.57	51.97	0.41	66.52	101.85	1.19	1.82
	Reaction speed (non-dominant)	309.00	57.00	34.20	313.00	59.00	35.40	0.64	49.22	−0.74	63.00	96.48	1.09	1.66
Working memory								0.60						
	WAIS-III letter-number sequencing	9.50	2.60	1.66	9.70	2.30	1.47	0.59	2.22	−0.95	2.85	4.36	1.16	1.78

* $p < 0.05$.

CI, confidence interval; ES, effect size; SEM, standard error of measurement; SD, standard deviation. Significant differences between time 1 and time 2 performances on individual measures indicate a practice effect.

Table 3

Percentage of Sample 1 (patient n = 60) that declined with 80% and 95% reliable change confidence interval on individual measures and mean percentage per domain

		80% CI		95% CI	
Domain	**Test**	**Patients**	**Controls**	**Patients**	**Controls**
Verbal memory	Mean percentage declined	11.7	2.2	5.9	0.0
	CVLT-II total trials 1–5	16.5	2.6	11.6	0.0
	CVLT-II short delay	14.9	7.9	11.6	0.0
	CVLT-II long delay	13.2	0.0	8.3	0.0
	CVLT-II recognition	15.7	2.6	0.0	0.0
	WMS-III logical memory I	8.9	0.0	4.9	0.0
	WMS-III logical memory II	5.7	0.0	2.4	0.0
	WMS-III logical memory recognition	6.6	2.6	2.5	0.0
Visual memory	Mean percentage declined	4.1	0.0	1.6	0.00
	WMS-III faces I	4.1	0.0	1.6	0.0
	WMS-III faces II	3.3	0.0	1.6	0.0
Processing speed	Mean percentage declined	16.2	4.9	9.5	2.44
	WAIS-III Digit Symbol-Coding	0.0	0.0	0.0	0.0
	D-KEFS Trails 1-visual scanning	16.3	5.4	7.3	2.7
	D-KEFS Trails 2-number sequencing	15.5	5.4	8.1	5.4
	D-KEFS Trails 3-letter sequencing	5.7	2.7	1.6	2.7
	D-KEFS Trails 4-number-letter switching	13.0	0.0	7.3	0.0
	D-KEFS trails 5-motor speed	13.0	2.7	9.8	0.0
	D-KEFS color naming	20.3	7.9	10.6	2.6
	D-KEFS word reading	20.3	7.9	14.6	5.3
	D-KEFS inhibition	31.2	13.2	22.1	2.6
	D-KEFS inhibition/switching	17.2	7.9	10.7	2.6
	Grooved pegboard (R)	21.3	5.4	10.7	5.4
	Grooved pegboard (L)	20.7	0.0	10.7	0.0
Simple vigilance	Mean percentage declined	8.4	2.8	6.3	0.0
	CPT vigilance total correct	0.0	0.0	0.0	0.0

		80% CI		95% CI	
Domain	**Test**	**Patients**	**Controls**	**Patients**	**Controls**
	CPT vigilance reaction time	16.8	5.6	12.6	0.0
Distractibility	Mean percentage declined	13.6	4.4	7.78	0.0
	CPT distractibility total correct	10.0	0.0	4.6	0.0
	CPT distractibility reaction time	17.1	8.8	10.8	0.0
Executive function	D-KEFS sorting-free sorting description score	18.7	7.9	8.1	2.6
Verbal ability	Mean percentage declined	18.6	6.1	9.56	1.73
	WAIS-III vocabulary	25.6	2.6	14.1	2.6
	D-KEFS animals	14.6	2.6	6.5	0.0
	D-KEFS boys' names	15.5	13.2	8.1	2.6

CI, confidence interval.

Table 4

Percentage of Sample 2 (TMX n=99; EXE n=80) that declined with 80% and 95% reliable change confidence interval

		80% CI			95% CI		
Domain	**Test**	**TMX**	**EXE**	**Controls**	**TMX**	**EXE**	**Controls**
Verbal memory	Mean percentage declined	9.10	8.45	5.85	2.48	2.28	1.65
	RAVLT total trials 1–5	13.8	8.1	10.0	2.5	2.0	2.5
	RAVLT delayed recall	3.8	3.0	2.5	1.2	1.0	0.8
	RAVLT recognition	8.8	14.6	7.6	1.2	1.0	0.8
	Visual association test	10.0	8.1	3.3	5.0	5.1	2.5
Visual memory	Mean percentage declined	8.10	5.6	9.55	3.2	2.0	2.9
	WMS-III visual reproduction immediate	6.2	7.1	5.8	2.5	1.0	3.3
	WMS-III visual reproduction delayed	10.0	4.0	13.3	3.8	3.0	2.5
Processing speed	Mean percentage declined	14.6	9.4	7.8	7.1	5.4	2.8
	Stroop word reading	13.8	14.1	12.5	6.2	7.1	4.2
	Stroop color naming	16.2	9.1	5.0	7.5	6.1	1.7
	Trail making test A	13.8	5.1	5.8	7.5	3.0	2.5
Executive function	Mean percentage declined	11.9	8.6	4.2	3.8	5.6	2.1
	Stroop interference	10.0	10.1	0.8	3.8	8.1	0.8
	Trail making test B	13.8	7.1	7.5	3.8	3.0	3.3
Manual motor speed	Mean percentage declined	10.0	7.6	4.8	5.0	5.6	1.65
	Finger tapping (dominant)	10.0	7.1	5.8	2.5	7.1	2.5
	Finger tapping (non-dominant)	10.0	8.1	3.8	7.5	4.0	0.8
Verbal fluency	Mean percentage declined	5.4	6.4	7.2	1.3	2.3	2.0
	Category – animals	7.5	7.1	6.7	3.8	4.0	1.7
	Category – professions	3.8	6.1	10.8	0.0	1.0	4.2
	Letters (D, A, T)	5.0	6.1	4.2	0.0	2.0	0.0
Reaction speed	Mean percentage declined	13.8	10.1	6.3	6.3	6.5	2.5
	Reaction speed (dominant)	18.8	14.1	7.5	8.8	10.1	1.7
	Reaction speed (non-dominant)	8.8	6.1	5.0	3.8	3.0	3.3
Working memory	WAIS-III letter-number sequencing	6.2	14.1	15.0	2.5	3.0	1.7

CI, confidence interval.