

**A Quantitative Analysis of Cognitive Impairments
Following Breast Cancer Treatment**

Lea Ann Maria Ouimet

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the PhD degree in Clinical Psychology

School of Psychology
Faculty of Social Sciences
University of Ottawa

©Lea Ann Maria Ouimet, Ottawa,
Canada, 2011

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor, Dr. Catherine Bielajew, for taking a chance on me and welcoming me into her lab, and into the graduate program. She has provided her guidance, support and editorial skills during my studies, but more than that, she has offered her friendship. I am also indebted to Angela Stewart for allowing me to use the data that she tirelessly collected so that I could undertake the studies in this thesis. I appreciate the thoughtful feedback and suggestions from my committee members, Dr. Claude Messier, Dr. Andra Smith, and Dr. Sophie Lebel; they served to strengthen my work.

I am very grateful to the women who participated in this research. Without their help, none of this would have been possible. They have selflessly helped to further research in this area for the benefit of future breast cancer survivors.

I thank my lab-mates for their camaraderie and my friends for their enduring presence and care. I appreciate the cheerleading from my brothers, Gregg and Gary Ouimet, and my step-daughter Emily Thorne, who had faith that I could do this.

I am lucky to have the loving support of my parents, Art and Pat Ouimet. They have stood proudly by me and offered me encouragement throughout this process. They are my safe harbor. To Doug Thorne, my partner and friend, thank you for your patience and understanding and for everything you do to make my life a little bit easier as I pursue my goals. And finally, to my children, Brittany, Blake and Kyle Reid, you are my driving force. Your hugs, kisses and laughter – and especially your love - can get me through anything.

ABSTRACT

One in nine North American women will be diagnosed with breast cancer in their lifetime and most will receive chemotherapy as part of their treatment. Although advances in treatment have increased survivorship, some research suggests chemotherapy results in cognitive deficits in a subset of recipients, a condition known as chemo-fog, thereby compromising quality of life. However, inconsistencies in methodology and neuropsychological assessment have complicated comparison of findings.

The first objective of this thesis was to review the methodological issues with an emphasis on the quantitative techniques typically employed. A comparison of group and individual based analyses found negligible effects for both univariate and multivariate approaches while individual based analyses identified severe declines in function in a subset of participants. A standardized-regression based (SRB) approach was recommended as the method of choice. Furthermore, it was recommended that the number of tests be limited since comprehensive batteries can complicate identification due to increased risk of misclassification.

Therefore, the second goal of the thesis was to evaluate the sensitivity of a reduced battery to the declines associated with chemo-fog. A comprehensive neuropsychological battery comprising 23 tests was compared to a subset of nine tests. SRB analyses demonstrated that a more selective battery was equally useful and may be appropriate for identification of chemo-fog.

Given the variability in the composition of neuropsychological test batteries, the final aim of this thesis was to compare the structure of the theoretical cognitive domains with ones identified through exploratory factor analyses (principle axis factoring) to evaluate the

convergence between the two. The results demonstrated there is statistical support for the conceptual framework that underlies the composition of the domains.

The contributions of this thesis include providing methodological guidelines for those conducting future research in this area to ensure that results are comparable across studies and are meaningful, and evaluating the utility of a screening battery to facilitate identification of chemo-fog. In addition, it was demonstrated that despite the lack of professional guidelines informing the selection and construction of neuropsychological test batteries, there is statistical evidence to support the practice of grouping tests into domains based on theoretical grounds.

LIST OF TABLES AND FIGURES

STUDY 1

Table 1.	Methodologies in review papers	23
Table 2.	Test battery identified by cognitive domain	31
Table 3.	Descriptive statistics of groups	35
Table 4.	Means and standard deviations associated with each measure	37
Table 5.	Probability levels associated with results of mixed ANOVAs	38
Figure 1.	Typical interaction pattern observed (ANOVA)	38
Table 6.	Probability levels associated with results of MANOVAs	39
Table 7.	Probability levels associated with results of MANCOVAs	40
Figure 2.	Percentage of participants declining on 0, 1, and 2 or more neuropsychological tests based on comparison with the healthy control group	41
Figure 3.	Percentage of participants declining on 0, 1, and 2 or more neuropsychological tests based on comparison with the hormonal group	43
Figure 4.	Percentage of participants declining on 0 – 9 neuropsychological tests based on comparison with the healthy control group	45

STUDY 2

Table 8.	Clinical characteristics of treatments groups	61
Table 9.	Means and standard deviations associated with each measure in the reduced battery	64
Table 10.	Rational for test selection	66
Figure 5.	Percentage of participants meeting criteria for cognitive decline on 0 – 9 tests based on comparison with a healthy control group	70
Figure 6.	Percentage of participants who met the cut-off criterion of significant decline on three or more tests in study 1 and two or more tests in study 2	71
Table 11.	Comparison of individual decline from study 1 to study 2	72

STUDY 3

Table 12.	Statistical factors at baseline and post-treatment	87
Table 13.	Comparison between conceptual and statistical factors at baseline and post-treatment	88

LEGEND

Analysis of variance	ANOVA
Beck Depression Inventory-II	BDI-II
High Sensitivity Cognitive Screen	HSCS
Multivariate analysis of variance	MANOVA
Multivariate analysis of covariance	MANCOVA
Predictive Analytic Software	PASW
Profile of Mood States	POMS
Reliable change index	RCI
Standard deviation	SD
Standard error of the mean	SEM
Standard Regression Based	SRB
Statistical Package for the Social Sciences	SPSS
Wechsler Adult Intelligence Scale – Third edition	WAIS-III
Wechsler Memory Scale – Third edition	WMS-III

Introduction	19
Operational definition of impairment	25
Test batteries	25
Designs	26
Groups	26
Types of analyses conducted	27
Model	28
Method	28
Data Analysis	32
Results	35
Discussion	46
Conclusion	52
STUDY 2	54
Abstract	56
Introduction	57
Model	60
Method	61
Data Analysis	67
Results	69
Discussion	73
STUDY 3	77
Abstract	79
Introduction	80

Method	83
Data Analysis	85
Results	86
Discussion	89
GENERAL DISCUSSION	92
Study 1	92
Study 2	93
Study 3	94
Quantitative aspects of chemo-fog research	95
Limitations and future directions	102
Conclusion	104
REFERENCES	106
APPENDICES	
A. Contributions of the collaborators	130
B. Description of neuropsychological and psychological measures	131
C. RCI syntax	141
D. SRB syntax	145

INTRODUCTION

Breast cancer is the most common cancer among women, both in Canada and the United States, affecting approximately one in nine women each year (American Cancer Society, 2010; Canadian Cancer Society, 2010). Advances in treatment have led to an increase in survival rates in recent years. However, the current standard in treatment continues to be chemotherapy with adjuvant hormonal (e.g. Tamoxifen) and/or radiation therapy. While effective, there are both physical and cognitive side effects associated with these drug therapies that impact quality of life (Ahles et al., 2005; Fallowfield, Leaity, Howell, Benson, & Cella, 1999; Ferguson et al., 2007; Hafner, 2009; Kayl, Wefel, & Myers, 2006; Mar Fan et al., 2005; Mulrooney, 2008; Reid-Arndt, Yee, Perry, & Hsieh, 2009; Taillibert, Voillery, & Bernarnd-Marty, 2007; Weis, Poppelreuter, & Bartsch, 2009). In fact, cognitive impairments are one of the most commonly voiced complaints in those treated with chemotherapy (Taillibert et al., 2007), and may create issues related to treatment adherence in cancer patients and follow-up care in survivors (Garofalo & Baum, 2001).

Ten years ago, only a handful of studies had examined the cognitive impairments reported by the recipients of chemotherapy treatment (Brezden, Phillips, Abdoell, Bunston, & Tannock, 2000; Schagen, van Dam, Muller, Boogerd, Lindeboom, & Bruning, 1999; van Dam et al., 1998; Wieneke & Dienst, 1995). In the last few years, as mortality rates have decreased and quality of life for survivors has become more of an issue, the phenomenon that has come to be known as “chemo-brain” or “chemo-fog” is receiving more attention (Abraham et al., 2008; Collins, Mackenzie, Stewart, Bielajew, & Verma, 2008; Debess, Riss, Engebjerb, & Ewertz, 2010; Ferguson et al., 2007; Hafner, 2009; Hermelink et al., 2007; Hurria & Lachs, 2007; Jansen, Miaskowski, Dodd, & Dowling, 2007; Jim et al., 2009;

Kesler, Bennett, Mahaffey, & Spiegel, 2009; Mulrooney, 2008; Ouimet, Stewart, Collins, Schindler, & Bielajew, 2009; Raffia & Tallarida, 2010; Reid-Arndt et al., 2009; Shilling & Jenkins, 2007; Stewart, Collins, Mackenzie, Tomiak, Verma & Bielajew, 2008a; Tager et al., 2010; Taillibert et al., 2007; Vardy, Rourke, & Tannock, 2007; Weis et al., 2009).

The impairments that characterize chemo-fog can include difficulties with memory, attention, concentration and perseverance. Although some studies suggest these are acute effects of chemotherapy that will resolve within approximately one year of treatment completion (Collins et al., 2008; Debess et al., 2010; Mar Fan et al., 2005; Wefel, Lenzi, Theriault, Davis, & Meyers, 2004b), others conclude that the effects are long-term and can persist for two to ten years post-treatment (Ahles et al., 2005; Brezden et al., 2000; Kesler et al., 2009; Kreukels et al., 2006; Schagen et al., 2002a; Tannock, Ahles, Ganz, & van Dam, 2004; van Dam et al., 1998).

Other studies report no significant declines in neuropsychological test results subsequent to chemotherapy, but do report an attenuation of practice effects from baseline to post-treatment that in all likelihood represents a decline in function (Collins et al., 2008; Tager et al., 2010).

There are three main objectives of this thesis. The first is to review in detail the methodological issues in this research area, with an emphasis on the impact of statistical methods on obtained results, and to provide recommendations that will enhance interpretability and comparison of findings (Study 1). The second goal is to evaluate the sensitivity of a reduced neuropsychological test battery to detect chemo-fog (Study 2). The aim of the third study is to compare the theoretical cognitive domains in a comprehensive neuropsychological test battery with the statistical ones identified through factor analyses to

determine if there is support for the conceptual framework underlying the composition of the domains (Study 3).

What is chemotherapy?

Chemotherapy or “chemical treatment” has been used since the days of the ancient Greeks and is the term coined to describe substances that are toxic to various forms of parasites but not to the human hosts. It wasn’t until the 1940s that physicians began to use it in the treatment of cancer. Today, the term is typically used to refer to the drug treatments used to kill cancer cells. These drug regimens may also be referred to as “anti-cancer” drugs or “antineoplastics.” There are many different types of chemotherapy and the mechanisms by which they target cancer cells vary depending on the type of drugs used. Typically, different agents are combined in a treatment regimen in order to maximize the effect (New, 2005; Verstappen, Heimans, Hoekman, & Postma, 2003).

Chemotherapeutic agents can be used to prevent cancer cells from multiplying with the goal of either curing the cancer, shrinking a tumour prior to surgery, controlling growth, or killing any cells that may be left after a tumour has been removed. However, these agents on their own cannot target cancer cells exclusively, and healthy cells are also damaged in the process (New, 2005; Verstappen et al., 2003).

Neurotoxic damage can occur both peripherally and centrally, as a function of the agents being administered (New, 2005; Verstappen et al., 2003). Methotrexate is one of the more commonly used treatments for breast cancer and is among the most detrimental to the central nervous system (New, 2005; Verstappen et al., 2003). Chemotherapy has also been shown to damage the visual system (Raffa & Tallarida, 2010) and induce anemia

(O'Shaughnessy, 2003; Kayl et al., 2006), conditions that may exacerbate impairments in cognitive functioning.

A newer treatment option called monoclonal antibodies is available for some forms of cancer and can be used to selectively target cancer cells. They show promising results in clinical trials for the treatment of colon cancer, and more recently for recurrent and metastatic breast cancer (Health Canada, 2009). However, they are administered with adjuvant chemotherapy; therefore the potential side effects of chemotherapy treatment are not eliminated.

Neuroanatomical correlates of chemo-fog

It was initially thought that chemotherapeutic agents could not penetrate the blood–brain barrier but there is recent evidence to suggest that these agents do in fact enter the cerebrospinal fluid and result in neural changes (Abraham et al., 2008; New, 2005; Troy et al., 2000; Tuxen & Werner, 1994; Verstappen et al., 2003). On occasion, these changes can produce complications including acute or chronic leukoencephalopathy, a change in the white matter that results in cerebral, cerebellar, and/or brainstem dysfunction (Filley, 1999; New, 2005; Verstappen et al., 2003).

For example, the risk increases from less than 2% in patients receiving intravenous methotrexate (typically used in the treatment of metastatic breast cancer) alone to approximately 45% in patients receiving intrathecal (injected into the spine) and intravenous methotrexate in addition to cranial radiation (Glass, Lee, Bruner & Fields, 1986). Over a 2-4 month period, symptoms can develop such as personality changes and decreased intellectual functioning. Chronic leukoencephalopathy can eventually lead to a subcortical type of

dementia with symptoms of apathy, cognitive/memory loss, frontal features, sleep disorders, incontinence, gait disorders, and possible seizures. In addition, chemotherapy has been associated with cerebrovascular stroke-like episodes (Posner in Saykin, Ahles, McDonald, 2003).

One study using MRI technology found brain lesions (referred to as high intensity areas) in children who had received intrathecal chemotherapy. These lesions were evident 13 and 16 weeks after treatment with some degree of change still evident after 52 weeks (Asato et al., 1992). Ahles et al. (2002) and Wieneke & Dienst (1995) have also found a significant positive correlation between the number of treatment cycles and degree of impairment. van Dam et al. (1998) found that recipients of high-dose chemotherapy had a 3.5 times higher risk of cognitive impairment compared to those who received standard-dose chemotherapy, and an 8.5 times higher risk than healthy controls and Kreukels et al. (2006) observed that breast cancer patients treated with high-dose chemotherapy have electrophysiologic alterations not seen in standard-dose controls.

A recent functional magnetic resonance imaging study by Kesler et al. (2009) examined differences in brain activation during memory tasks in breast cancer patients three years post chemotherapy with age and education matched healthy controls. They found that patients had different activation patterns and significantly more neural recruitment was required to achieve the same test scores. They reason that this increased activation may lead to cognitive fatigue and feelings of frustration when performing cognitive tasks.

Finally, a 2008 study by Abraham et al., using diffusion tensor imaging, which measures the diffusion of water to assess white matter integrity, revealed that breast cancer patients who received chemotherapy had reduced integrity in the genu of the corpus

collosum which was associated with significantly poorer performance on a test of processing speed.

Hypothesized mechanisms

The specific mechanisms that lead to cognitive and memory deficits are still not well understood but there are several that have been hypothesized including direct neurotoxic injury to the microglia, oligodendrocytes, and neuronal axons, resulting in demyelination or altered water content; secondary inflammatory response or allergic hypersensitivity; microvascular injury that causes obstruction of small- and medium-sized blood vessels, inducing ischemia/infarction and necrosis; and altered neurotransmitter levels (Tuxen & Werner, 1994).

Adjuvant hormonal therapies: Antiestrogens

Currently, antiestrogens are frequently given to breast cancer patients in conjunction with chemotherapy to decrease or stop breast cancer cell proliferation, especially in cases of tumours found to be estrogen positive. These drugs bind to the estrogen receptor site on cancer cells and block estrogen from entering the cancer cell, interfering with cell growth and eventually leading to cell death. However, estrogen is thought to offer some protection against the negative effects of chemotherapy by enhancing neuronal growth, and can serve to improve cognitive function (Eberling, Wu, Tong-Turnbeaugh, & Jagust, 2004; Garofalo & Baum, 2001). These effects have been found to be negated with Tamoxifen treatment (Castellon, Silverman, & Ganz, 2005; Eberling et al., 2004; Jenkins, Shilling, Fallowfield, Howell, & Hutton, 2004; Paganini-Hill & Clark, 2000; Schilder et al., 2009) and may in fact

produce impairments beyond those resulting from chemotherapy alone (Bender et al., 2006; Castellon et al., 2005). A study by van Dam et al. (1998) also noted that high dose chemotherapy is known to significantly impair ovarian function (Meirow, 1999; 2000; Lo Presti, Ruvolo, Gancitano & Cittadini, 2004; Torrents, Boiso, Barri & Veiga, 2003) thereby resulting in even lower levels of estrogen.

Summary of current research findings

Two recent meta-analyses have provided summaries of the research findings in this area. The first was a review of 29 studies that evaluated cognitive problems in chemotherapy patients with a variety of cancers (Anderson-Hanley, Sherman, Riggs, Agocha & Compas, 2003). They found that there were often significant reductions in cognitive functioning, particularly in the areas of executive function (*Cohen's d* = - .93) and verbal memory (*Cohen's d* = - .91).

The second meta-analysis was conducted in our laboratory (Stewart, Bielajew, Collins, Parkinson, & Tomiak, 2006) and was restricted to breast cancer patients. The results also showed that adjuvant chemotherapy treatment was associated with declines in cognitive functioning, with language, short-term memory, and spatial abilities being the areas most affected. Working and long-term memory, processing speed, and motor abilities were also negatively impacted, although to a lesser extent. Generally, the magnitude of the decline was in the order of -.24 (*Hedges' d*).

Taken together, the results indicate that there are subtle and enduring cognitive deficits in some breast cancer survivors. It is important to note that although these effects may appear small from a statistical standpoint, the nature of the impairments may have

negative impacts on the ability to function optimally in home, social, educational, and/or work settings.

However there are various methodological issues that make it difficult to draw clear conclusions from the literature in this area. These include inconsistencies in the operational definitions of key constructs including “cognitive impairment”, differing design types, group composition and selection of control groups, variability in the statistical methods used, and the lack of consistency in the composition of neuropsychological test batteries.

Neuropsychological test batteries

There are several issues that arise with respect to the composition of neuropsychological test batteries. Variability in the choice of tests used to represent the various cognitive domains make comparison of research results arduous, leading to difficulties in determining which areas of cognitive functioning are most affected in chemo-fog. This is due in part to the inability to arrive at a professional consensus or adopt policies regarding test selection, and the paucity of any theoretical findings on which to base such decisions. There is also some controversy regarding the preferential use of flexible test batteries (tests are chosen based on presenting problem and patient performance) versus fixed test batteries (an assembled, normed set of tests) (Wong, 2006).

Given the lack of appropriate guidelines, decisions about what tests to include in a battery may be based more on what is available in a particular setting or traditionally used, what is familiar, or simply personal preference as opposed to evidence of reliability and validity or statistical support. This is particularly troublesome because neuropsychological

tests vary greatly with respect to length, complexity, and what they are purporting to measure, and many do not have established reliability and validity.

A related issue is the decision regarding when to adopt the newest version of a revised test. As with other testing issues, there are no clear professional guidelines to inform choices other than the ethical responsibility to keep training up to date and use reliable and valid, evidence-based techniques and materials. However, given the economic benefits of revisions for the test publishers, it is not always clear that switching to a newer version of an instrument will improve clinical decision making (Bush, 2010; Silverstein & Nelson, 2000; Strauss, Spreen, & Hunter, 2000).

In addition, it is a challenge to choose tests that are clean (i.e. measure one specific construct) because all neuropsychological tests capture different facets of cognitive functioning and it is not always possible to look at the final score on a test and understand where the problem lies. For example, on the Boston Naming Test, a low score could be due to one of three things: an actual naming problem - the individual knows the object but can't access the word; a perceptual problem - the individual does not recognize what they are looking at; or a loss of semantic knowledge - the individual has lost this information altogether, as if it were never known. Testing the limits would resolve this issue but this would be difficult to do in a research setting and would lead to unstandardized test administration across participants, thus introducing confounds. Although any impairment is relevant, regardless of the exact nature of the problem, providing interventions becomes difficult without a clear understanding of the specific functions affected.

A recent study by Downie, Mar Fan, Tchen, Yi, & Tannock, (2006) of 21 women receiving adjuvant chemotherapy for breast cancer included the self-report Functional

Assessment of Cancer Therapy – General questionnaire (FACT-G; Cella et al., 1993) as well as additional subscales on fatigue (FACT-F; Yellen, Cella, Webster, Blendowski, & Kaplan, 1997) and endocrine symptoms (FACT-ES; Fallowfield et al., 1999). These measures assess quality of life in cancer patients and include items addressing physical, emotional, family/social and functional domains (FACT-G), the effects of fatigue (FACT-F), and symptoms related to changes in endocrine function (FACT-ES). The results of the self-report measures were then compared with the results of the High Sensitivity Cognitive Screen (HSCS; Faust & Fogel; 1989; Fogel, 1991), a neuropsychological measure designed to detect cognitive impairment, to determine whether subjective complaints of cognitive impairment coincided with objective neuropsychological test results.

The authors found discrepancies between the self-reports and the cognitive assessment, particularly in the area of attention and concentration, in that 90% of the patients reported difficulties but only 10% were categorized as abnormal on the HSCS. Language and memory were also reported as problematic in 78 and 95% of patients respectively, whereas the HSCS indicated abnormalities in only 61 and 48% of patients. The authors conclude that the HSCS may not be sensitive enough to capture the subtle cognitive declines experienced by this population.

A 2007 study by Jansen, et al. had similar findings. They conducted a meta-analysis of thirteen studies to identify which neuropsychological tests were more sensitive to the effects of chemotherapy on cognition. They found that of the 32 tests evaluated, only six were sensitive to chemotherapy induced impairments, one in the domain of language, two in motor function, two in visuospatial skills, and one in verbal memory. They concluded that

further investigation is needed to identify those instruments that are most valid and reliable for the detection of chemo-fog.

Although limited sensitivity to chemo-fog related impairments may explain the obtained results, it is equally possible that there is some divergence between theoretical cognitive domains and statistical ones, and that this could account for the apparent lack of sensitivity. One way to address this is to explore it further through the use of factor analysis.

Factor analysis and the study of cognitive functioning

Factor analysis was first used in 1904 by psychologist Charles Spearman in his research on mental abilities. He was interested in knowing if mental abilities correlated with each other and he set up a table of correlations in a hierarchy. Based on the results, he believed that there was an underlying factor in each of the variables that could be explained by the overall factor that he called “g”, and that each variable had its own saturation level (i.e. the maximum possible variance accounted for by any given variable) in “g” (Carroll, 1993).

The model that he used was refined several times over the years and in its current form, it is used as a means of finding patterns in the relationships among variables in order to determine if they can be explained by a smaller number of variables called factors. There are different types of factor analyses that can be performed, the most common being exploratory and confirmatory analyses. Exploratory factor analysis is used to examine which factors share the most variance statistically whereas confirmatory factor analysis involves imputing the variables based on established criteria in order to confirm whether they fit the theoretical model. Both models are used as a means of evaluating the fit between the hypothesized

structure and the actual data in order to determine whether or not the theories can be substantiated. Principal components analysis is another data reduction method used to produce a smaller number of factors. It is typically used prior to performing other statistical analyses when there are no specific hypotheses regarding the underlying structure of the data.

Both confirmatory and exploratory factor analysis have been used extensively in the study of cognitive functioning. For example, Newby, Hallenbeck and Embretson (1983) used confirmatory factor analysis to test the explanatory power of four proposed models using the modified Halstead-Reitan Battery in a neuropsychiatric population. They found that by making several modifications to the best fitting initial model, they achieved a relatively good fit. Pedraza et al. (2005) used it to replicate the 5-factor model previously identified in a sample of white participants from the Mayo's Older Americans Normative Studies (Ivnik et al., 1992; Lucas et al., 2005) with a sample of African American participants (Mayo's Older African Americans Normative Studies; Lucas et al. 2005). They were able to confirm the 5-factor model from the original study in this new population.

Exploratory factor analysis has been used by Mertens, Gagnon, Coulombe & Messier (2006) to determine which cognitive functions were recruited in the interference condition of the Brown-Peterson Task. Performance in this condition was shown to load primarily on a factor made up of the subtests of the Working Memory Index of the Wechsler Memory Scale-III and the Wechsler Adult Intelligence Test-III. Strauss & Fritsch (2004) performed exploratory factor analysis to identify the number of factors underlying the Consortium to Establish a Registry for Alzheimer's Disease neuropsychological test battery (CERAD). They found that dementia severity accounted for the shared variance of the CERAD

measures and that naming, praxis, and verbal fluency could effectively discriminate between the three subtypes of Alzheimer's disease. Based on these results, they recommended adding tests to improve reliability.

Principal components analysis was employed by Holtzer, Stern & Rakitin (2005) to evaluate which cognitive functions are compromised in dual-task performance in an older population. Their results suggest that attention and executive function are the strongest predictors of age-related decline in dual-task performance. O'Donnell, Drachman, Lew & Swearer (1988) used it to identify deficits associated with Alzheimer's Disease and to develop a brief screening tool (Brief Battery for Dementia) that was then used to evaluate the impact of disease severity on the discriminating power of the battery. They found that disease severity may be an important moderating variable in the discriminating power of intelligence tests. Similarly, McCracken & Franzen (1992) used principal components analysis to assess the equivalence of alternate forms of the Trail Making Test (B and D) to determine if they were as reliable as the originals (A and B). Based on their findings, the alternate forms of the test appear to be as reliable as the original version.

Overview of thesis articles

As mentioned, this thesis will consist of a series of three articles, the purpose being to examine cognitive impairments associated with chemotherapy from a quantitative perspective. The aim of the first paper (Study 1) was to propose a model of statistical analysis for application in this and related areas and to address various methodological issues commonly seen in this literature. In the second paper (Study 2), we will use the data obtained in the first study to construct and analyze a smaller neuropsychological test battery

for screening chemotherapy-induced cognitive impairments. In the final paper of this thesis (Study 3), factor analysis will be conducted on our data at baseline and post-treatment to evaluate whether the theoretical cognitive domains (i.e. group of tests chosen to measure a specific cognitive function such as language, attention, memory, etc.) are consistent with the statistically derived ones. A statement regarding contribution of collaborators and co-authors can be found in Appendix A.

Study 1 – Methodological Review

The objective of the first study of this thesis was to conduct a review of a representative subset of the literature in this area with the goal of identifying difficulties associated with the methodologies and statistical analyses typically employed. As an illustration, all methods were applied to a subset of data from our laboratory (Ouimet et al., 2009; Stewart et al., 2008a) to demonstrate both the applied and theoretical advantages and disadvantages of the various statistical procedures currently being employed. A model was then proposed that aimed to balance methodological and statistical considerations with the objective of obtaining results that are clinically relevant. In addition, there were several methodological issues that required clarification to ensure some degree of standardization among researchers in this area. These include the operationalization of cognitive impairment and the criteria with which participants are classified as impaired as well as design type, choice of appropriate control group(s), the selection of neuropsychological tests used to assess cognitive function, and the inclusion of covariates.

Study 2 – Investigation of a sensitive neuropsychological screen for chemo-fog

The aim of the second study of this thesis was to improve the clinical utility of our neuropsychological battery by reducing the number of tests, therefore, the original battery comprising 23 tests was compared with a reduced battery consisting of a subset of these tests. The tests were chosen to represent the domains previously identified as being most susceptible to the effects of chemotherapy (Anderson-Hanley et al., 2003; Stewart et al., 2006). Standard Regression Based (SRB; McSweeney, Naugle, Chelune & Lunders, 1993) analyses were applied to evaluate whether the scores of participants in the chemotherapy group declined across time. Performance across variables was summed and participants classified based on degree of decline. Results were compared to those obtained in Study 1 and it was found that the reduced battery was equally able to identify the women previously classified as significantly impaired.

Study 3 – Factor Analysis

The objective of the final study of this thesis was to compare the theoretical cognitive domains of a neuropsychological test battery with those identified through the use of factor analytic techniques in order to evaluate whether there was statistical support for the conceptualization of the cognitive domains. Factor analyses (principal axis factoring) were performed on each separate group (chemotherapy, hormonal, and healthy control) using the baseline and post-treatment results of 23 neuropsychological tests. Since the factors were expected to be inter-related to some extent, a non-orthogonal (oblimin) rotation was employed to enhance interpretability of the factors.

STUDY 1

Published as: Ouimet, L. A., Stewart, A., Collins, B., Schindler, D., & Bielajew, C. (2009).

Measuring neuropsychological change in chemo-fog: An analysis of statistical models.

Journal of Clinical and Experimental Neuropsychology, 31, 73-89.

**Measuring neuropsychological change following breast cancer treatment:
an analysis of statistical models**

Ouimet, L. A.; Stewart, A.; Collins, B.*; Schindler, D.; & Bielajew, C.

School of Psychology, University of Ottawa, Ontario, Canada.

*Ottawa Hospital, Civic Campus, Ottawa, Ontario, Canada.

ABSTRACT

This article considers the quantitative techniques currently in use in the evaluation of cognitive impairments associated with chemotherapy treatment for breast cancer. To illustrate differences among analytical approaches, all analyses were applied to baseline and post-treatment scores on neuropsychological tests obtained from Stages I and II breast cancer patients receiving either chemotherapy or hormonal therapy; a healthy control group with similar demographics to those of the treatment groups was also included. Conventional group analyses were compared with individual based analyses (standardized regression-based and reliable change methods). Both univariate and multivariate techniques with and without covariates produced negligible effects. In contrast, results of the individual based analyses identified a subset of participants in the chemotherapy group who experienced a severe decline in function on two or more tests. Differences between the control and treatment groups were greater than differences between the treatment groups alone. The standardized regression-based approach was more sensitive than the reliable change index in detecting chemotherapy and hormonal therapy subjects whose performance was different from baseline scores on two or more tests (roughly 80% vs. 50% of participants). From a clinical perspective, the degree of impairment determined on the basis of the individual-based methodologies could have a major impact on quality of life for those affected. On the whole, we argue that the standardized-regression based approach, allowing for the assessment of individual practice effects and evaluation of moderator variables, is the method of choice in this context.

Introduction & background

Breast cancer is the leading cause of malignancy in women. Approximately 1 in 9 women will develop breast cancer in their lifetime, and despite advances in medical treatments roughly 1 in 25 will die of it (National Cancer Institute, 2007; Public Health Agency of Canada, 2007). The current standard in treatment of breast cancer is chemotherapy, adjuvant hormonal (e.g. tamoxifen) therapy, and radiation therapy, or some combination of this depending on cancer stage and other prognostic indices such as patient age and cancer history. Although all of these treatments are known to be associated with physical side-effects, more recently it has been recognized that they may also produce cognitive disturbances that in some cases are long-lasting.

Five years ago, there were only a handful of articles reporting cognitive impairments related to chemotherapy treatment. Today, as survival rates have increased, and quality of life for survivors is now becoming an important treatment-related consideration, the phenomenon that has been coined “chemo-brain” or “chemo-fog”, - that is, chronic memory and attention problems post-treatment - is receiving more attention (Ahles et al., 2002; Bender et al., 2006; Castellon et al., 2004; Donovan et al., 2005; Gottschalk, Holcombe, Jackson, & Bechtel, 2003; Hermelink et al., 2007; Hurria & Lachs, 2007; Jansen, Miaskowski, Dodd, & Dowling, 2005; Jenkins et al., 2006; Kreukels et al., 2005; Mar Fan et al., 2005; Saykin et al., 2003; Schagen, Hamburger, Muller, Boogerd, & van Dam, 2001, Schagen et al., 2002a, 2002b; Scherwath et al., 2006; Servaes, Verhagen, & Bleijenberg, 2002; Shilling & Jenkins, 2007; Shilling, Jenkins, Fallowfield, & Howell, 2003; Shilling, Jenkins, Morris, Deutsch, & Bloomfield, 2005; Stewart et al., 2006; Stewart et al., 2008; Tchen et al., 2003; Vardy et al., 2006; Wefel et al., 2004a; 2004b).

A meta-analysis of 29 studies reported significant reductions in cognitive functioning following chemotherapy treatment for a variety of cancers, particularly in the areas of executive function (*Cohen's d* = - .93) and verbal memory (*Cohen's d* = - .91) (Anderson-Hanley et al., 2003). Using the same approach but restricted to breast cancer studies, we found that adjuvant chemotherapy treatment was associated with declines in cognitive functioning, particularly in areas of language, short-term memory, and spatial abilities (Stewart et al., 2006). Working and long-term memory, processing speed, and motor abilities were also affected, although to a lesser extent. Generally, the magnitude of the decline was in the order of -.24 (*Hedges' d*). Taken together, the results suggest that there are subtle and enduring cognitive deficits in some breast cancer survivors. It is important to note that although these effects may appear small from a statistical standpoint, the nature of the impairments may influence the ability to function optimally in home, social, educational, and/or work settings. However, the clinical relevance of these impairments has yet to be investigated.

To compound matters, selective estrogen receptor modulators (e.g. tamoxifen) are frequently given to breast cancer patients in conjunction with chemotherapy to inhibit breast cancer cell proliferation, especially in estrogen-positive tumors. However, there are studies indicating that Tamoxifen treatment may in fact produce impairments beyond those resulting from chemotherapy alone (Bender et al., 2006; Castellon et al., 2004; van Dam et al., 1998) and may therefore increase the likelihood of cognitive decline. There is little by way of data yet on the effects in this context of other classes of compounds that interfere with estrogen production such as the aromatase inhibitors; indeed, in many institutions, these are now the

preferred choice of hormonal adjuvant treatment. For more detail, we recommend the recent review of hormonal treatments on cognitive functioning by Schilder and Schagen (2007).

Despite the mounting evidence that chemotherapy, and possibly adjuvant tamoxifen, produce declines in neuropsychological functioning (Ahles et al., 2002; Ahles, Tope, Furstenberg, Hann, & Mills, 1996; Bender et al., 2006; Brezden et al., 2000; Castellon et al., 2004; Kreukels et al., 2005; Schagen et al., 1999, 2001; Scherwath et al., 2006; Shilling et al., 2005; Tchen et al., 2003; van Dam et al., 1998; Wefel et al., 2004a), differences in methodological approaches across studies have made it difficult to assess level of impairment in this population. These differences range from relatively minor – design types, inclusion/exclusion criteria for participants, demographics – to those serious enough to compromise the ability to compare results in any meaningful way, such as major discrepancies in the definition of key constructs - for example, cognitive impairment.

In a review of studies in this context by Anderson-Hanley et al. (2003), design types were mainly one time (posttreatment) cross-sectional between-subject (treatment and healthy control) assessments that failed to take into account premorbid level of cognitive functioning. Sample sizes were frequently too small to provide results that could be considered meaningful or generalizable. In addition, other factors known to potentially have an impact on cognitive function such as age, education, estimated IQ, depression, anxiety, and fatigue were frequently not assessed. Although we found no relationship between these variables and neuropsychological test performance in a recent meta-analytical review of this literature, their potential as predictors should be routinely evaluated (Stewart et al. 2006).

The statistical methods used to evaluate data also vary widely from study to study leading to questions about the meaningfulness of significant results. For example, a brief

report by Shilling, Jenkins, and Trapala (2006) determined that the degree of cognitive impairment reported in breast cancer patients was to a large extent a function of the method of analysis adopted, which influenced judgment of significance levels. This is particularly salient when the impairments are subtle, as appears to be the case in “chemo-brain”, and may be overlooked depending on the analytical methods employed.

The lack of consistency across studies with respect to methodology and statistical approach poses difficulties both for researchers studying the phenomenon and clinicians searching for advice that is of value in their practice. More uniform methodologies across studies would allow researchers to readily replicate, verify, and conduct meta-analyses on research findings and help clarify areas of focus for clinicians. The aim of this paper is to develop a model of statistical analysis for application in this and related areas. It will begin with a review of the methodologies used in 22 studies based on retrospective and prospective assessments of the long-term cognitive deficits associated with chemotherapy in breast cancer patients. The articles included are a small subset of those found via a search of PubMed using the search terms “breast cancer”, “chemotherapy”, “cognitive impairment”, and “neuropsychological function” and viewing all related articles. They were chosen to represent the broad range of group and design types and quantitative methods used in this area. Table 1 below contains details for each of the 22 studies.

Table 1. Methodologies in review papers

<i>Author(s)/year</i>	<i>n</i>	<i>Definition of impairment</i>	<i>Group type</i>	<i>Design type</i>	<i>Analyses conducted</i>
Ahles et al., 1996	54	Decline in average performance by group & from T1 to T2	Treatment (2) ^c	Repeated Measures	Manova, Anova
Ahles et al., 2002	128	Performance falling in the lower quartile as a function of treatment type	Treatment (2)	Between Groups	Mancova
Bender et al., 2006	46	Decline in average performance by group & from T1 to T2	Treatment (2) & control	Mixed Repeated Measures	Mancova, Ancova
Brezden et al., 2000	107	Significantly lower mean score on neuropsychological screening measure than controls	Treatment (2) & healthy control	Between Groups	Wilcoxon, Ancova
Castellon et al., 2004	72	Significantly lower mean performance on domain or global score than controls	Treatment (2) & healthy control	Between Groups	MANOVA, ANOVA, Correlation
Donovan et al., 2005	143	$Z \geq 2$ SDs below relevant published norms	Treatment (2)	Between Groups	Anova, Chi-square
Jenkins et al., 2006	177	1. Significantly lower mean score on subtest than control or from T1 to T2 2. Reliable decline (CI 90%) on subtests from T1 to T2 compared to controls	Treatment (2) & healthy control	Mixed Repeated Measures	Anova, RCI
Kreukels et al., 2005	49	Significantly lower mean score on information processing task than no-chemo group	Treatment (2)	Between Groups	Anova, Correlation, (Chi-square)
Mar Fan et al., 2005	164	Classification of moderate or severe on the High-Sensitivity Cognitive Screen ^a	Treatment (2) & healthy control	Mixed Repeated Measures	Wilcoxon, Trend, Correlation, Regression
Schagen et al., 1999	73	1. Number of tests with $Z \geq 2$ SD below mean of published normative data 2. 5 th percentile of number of tests 2 SD below mean	Treatment (2)	Between Groups	ANOVA, Correlation, Regression
Schagen et al., 2001	47	Same as above	Treatment (2) & control	Mixed Repeated Measures	Anova, Correlation
Schagen et al., 2002	103	3 or more subtests with $Z \geq 2$ SD below the mean of control group	Treatment (3) & control	Mixed Repeated Measures	Correlation, Regression
Scherwath et al., 2006	76	1. $Z \geq 1.4$ SD below mean 2. 5 th percentile of control group	Treatment (2) & control	Between Groups	Anova, Kruskal-Wallis (Chi-square)
Servaes, Verhagen, & Bleijenberg, 2002	228	Significantly lower mean score on subtest than nonfatigue or control group	Treatment (2) & healthy control	Between Groups	Chi-square, GLM-General Factorial
Shilling et al., 2003	129	Significantly lower mean score on subtest than controls	Treatment & healthy control	Between Groups	Anova, <i>t</i> Tests

Shilling et al., 2005	93	Reliable decline (CI 90%) on 2 or more subtests from T1 to T2 compared to controls	Treatment & healthy control	Mixed Repeated Measures	Anova, Chi-square, Mann-Whitney <i>U</i> , Correlation, RCI
Stewart et al., 2006	112	$\geq 2 SD$ below predicted score in 2 or more subtests	Treatment & healthy control	Mixed Repeated Measures	Anova, Chi-Square, SRB
Tchen et al., 2003	200	Classification of moderate or severe on the High-Sensitivity Cognitive Screen ^a	Treatment & healthy control	Between Groups	Wilcoxon, Chi-square, Correlation
van Dam et al., 1998	104	1. Number of tests with $Z \geq 2 SD$ below mean of published normative data 2. 5 th percentile of number of tests $2 SD$ below mean for control	Treatment (2) & control	Between Groups	Anova, Correlation, Regression
Wefel et al., 2004a	18	1. Reliable decline (CI 90%) on subtests from T1 to T2 & T3 2. Significantly lower mean score on tests from T1 to T2 & T3	Treatment	Repeated Measures	RCI, Correlation, <i>t</i> Tests
Wefel et al., 2004b	84	$Z \geq -1.5 SD^b$ on two or more subtests <i>or</i> $Z \geq -2.0 SD^b$ on one subtest	Treatment	Between Groups	Chi-Square, Correlations, <i>t</i> Tests
Wieneke & Dienst, 1995	28	1. Mild = $> -1 SD$ on at least two measures 2. Moderate = mild <i>and</i> $> -2 SD$ on at least one measure	Treatment	Repeated Measures ^c	Manova, <i>t</i> Tests

Note. ANOVA = analysis of variance. ANCOVA = analysis of covariance. MANOVA = multivariate analysis of covariance. GLM = general linear model. RCI = reliable change index. SRB = standardized regression-based model.

^aControl (diseased) versus healthy control. ^bBased on published normative data. ^cCompared obtained scores with estimated pre-morbid functioning.

Operational definition of impairment

As noted recently by Vardy et al. (2007), an immediate concern is related to the definition of impairment, which differs widely across studies. For example, some researchers have defined cognitive impairment as a decline in average group performance on neuropsychological tests (Ahles et al., 1996; Bender et al., 2006). Others define it as lower mean scores than those for a control group (Castellon et al., 2005; Kreukels et al., 2005; Servaes et al., 2002; Shilling et al., 2003), while some (Mar Fan et al., 2005; Tchen et al., 2003) have used a classification of “moderate” or “severe” on the High Sensitivity Cognitive Screen (Faust & Fogel, 1989; Fogel, 1991). Indeed, the operational definition of impairment is clearly indicated in only a few studies. As a result, the information reported in Table 1 was derived largely based on the types of analyses conducted.

Tests/batteries

Another issue pointed out by Freeman and Broshek (2002) and McQuellon (comment in Olin, 2001) relates to the lack of consistency in the composition of test batteries commonly employed. Vardy et al. (2007) and Hurria and Lachs (2007) also point out the need for shorter batteries to eliminate fatigue that may arise from completion of a 2-3 hour test battery. Freeman and Broshek’s study (2002) focused on the construction of a more precise, less time-consuming battery consisting of neuropsychological tests that are most sensitive in detecting cognitive decline in breast cancer patients. This would enable other researchers and clinicians to readily compare results across studies and would serve to clarify the domains and abilities affected. In addition, it would provide information to clinicians as to the selection of appropriate tests needed to quickly and accurately identify the subset of women who experience significant cognitive decline.

Designs

Design types vary widely across studies and include cross-sectional and retrospective between-groups designs and retrospective or prospective (longitudinal) repeated-measures approaches, or mixed designs. While cross-sectional and retrospective studies can be very useful for exploring areas of research to determine which variables merit more scrutiny, many researchers are now conducting prospective research instead (Ahles et al., 1996; Bender et al., 2006; Jenkins et al., 2006; Mar Fan et al., 2005; Schagen et al., 2001; Schagen et al., 2002a; Shilling et al., 2005; Stewart et al., 2008a; Wefel et al., 2004b). Prospective studies are the preferred choice because they provide a baseline comparison to posttreatment results.

Groups

The composition of groups also varies across studies with some including treatment groups only (i.e. chemotherapy, hormonal, or adjuvant), some including a control group consisting of cancer patients who do not receive any chemotherapy or hormone therapy treatments, and some using healthy controls (see Table 1).

The inclusion of a disease control group may rule out effects that are not treatment specific but, rather, related to the diagnosis of the disease (e.g. stress). Ideally, this group would comprise breast cancer patients who are not receiving chemotherapy (e.g. those treated with radiation therapy) or adjuvant hormonal therapies, as they may also be associated with impairments in cognitive function. Since the overwhelming majority of breast cancer patients in western countries do undergo chemotherapy, it would prove extremely difficult to recruit this group. As a result, this comparison group often consists of breast cancer patients who have not yet started chemotherapy or those diagnosed with an equally serious but non-malignant disease. However, using a disease control group as a

reference group creates the possibility of underestimating cognitive decline from a healthy state, whereas the inclusion of a comparable healthy control group facilitates this comparison. In addition, the data gathered from a healthy control group can be used to assess the practice effects associated with repeated testing since alternate forms of instruments may or may not be available (Freeman & Broshek, 2002). Although a disease control group could also be used for this purpose, there is a risk that practice effects might be attenuated by the impact of the disease, causing the actual degree of cognitive decline to be misjudged.

Types of analyses conducted

The statistical approaches used to analyze the neuropsychological data in these studies ranged from group to individual-based analyses, including the computation of *z*-scores using published norms, *t* tests, analyses of variance (ANOVAs), multivariate analyses of variance (MANOVAs), and the assessment of reliable change. None of the investigators used a standardized regression-based model, which takes into account regression to the mean and practice effects as well as allowing for the inclusion of moderator variables (Sawrie, Marson, Boothe, & Harrell, 1999). In addition to the wide range of statistical tests employed, the type of data used to compute neuropsychological impairment varies, with some researchers using subtest scores, some domain scores, and some overall scores (see Table 1).

While there has been a tendency to explore “group” as opposed to “individual” differences, declines in cognitive function are thought to affect only a subset of those who have received chemotherapy (Stewart et al., 2008a). Conducting group analyses makes it more likely that these subtle effects will be missed as group data evaluation obscures individual differences, and the appropriateness of group comparisons is questionable if the results are to have any clinical utility.

Sample sizes are often adequate for any individual statistical test performed on a data set; however, it is frequently the case that multiple tests are evaluated without any adjustment to the family size error rate (e.g., Bonferroni). While this appears to be an accepted practice, it is only appropriate if exploratory analyses are being conducted as a means to flesh out variables of interest. In the case of confirmatory analyses, it is necessary to control for the Type I error rate, or risk reporting significant findings that are more likely to be due to chance than to the interventions being employed (Kirk, 1968). Unadjusted error rates can have serious consequences for clinicians if they follow recommendations based on these findings.

Model

As an illustration, all methods will be applied to a subset of recently published data from our laboratory (Stewart et al., 2008a) as well as the inclusion of new unpublished data (disease- free control group) on this phenomenon to demonstrate both the applied and theoretical advantages and disadvantages of the various statistical procedures currently being employed. Finally, a model is proposed aimed to balance methodological and statistical considerations with the objective of obtaining results that are clinically relevant.

METHOD

Participants

Data were obtained from 123 participants who were recruited as part of a larger longitudinal study conducted out of the Ottawa Regional Cancer Centre to investigate the neuropsychological effects of adjuvant chemotherapy. They consisted of stage I or II breast cancer patients who had undergone mastectomy or lumpectomy and who were receiving chemotherapy (with or without hormonal treatment, $n = 49$) or hormonal treatment alone ($n = 46$). The majority of patients in the chemotherapy group received anthracycline-based

chemotherapy regimen (FEC = Fluorouracil (5FU), Epirubicin, and Cyclophosphamide ; CEF = Cyclophosphamide, Epirubicin, and Fluorouracil; FAC = Cyclophosphamide, Adriamycin, and 5-Fluorouracil).

Given the often subtle nature of the observed cognitive impairments, it was considered important to have a comparison group that exhibited very similar characteristics to the treatment group aside from their status as cancer patients. While the tests in the battery are standardized and have established norms, they do not allow for this level of precision. Therefore, a healthy control cohort ($n = 28$) was also included as a comparison group to examine practice effects and to provide a basis on which to evaluate these effects in the treatment group.

All women were between the ages of 50 and 66 ($M = 57.89$, $SD = 4.13$). The healthy control sample was recruited through advertisements posted at the Ottawa Hospital. Participants received \$50 for each test session completed.

Sample selection

In order to minimize the potential confounding effects of hormonal status and age-related cognitive decline, only postmenopausal women aged 65 years or younger (generally >50 years) were included in the study. Fluency in English was required in order to complete the test battery. Exclusion criteria for all groups included a history of previous cancer and chemotherapy or radiation treatment. Finally, participants presenting with serious psychiatric disorder (e.g. major depression, schizophrenia), neurological illness, or significant substance abuse were also excluded due to their potential negative effects on cognition. This study was approved by the board of ethics at the Ottawa Hospital, and written informed consent was obtained from all participants.

Measures

The Quick Test (Ammons & Ammons, 1962) is a measure of receptive vocabulary that provides an IQ equivalency score. It was included in the battery in order to provide an estimate of pre-morbid cognitive function.

The Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1992) was designed to assess psychological distress and changes in mood over time. It contains six subscales including Anger-Hostility, Confusion-Bewilderment, Depression-Dejection, Fatigue-Inertia, Tension-Anxiety, and Vigor-Activity as well as a Total Mood Disturbance score. Participants are asked to use a 5-point Likert scale to rate their mood over the past 7 days.

A total of 23 neuropsychological tests with strong psychometric properties were chosen to represent the major cognitive domains. They included executive function, language function, motor skills, processing speed, verbal learning and memory, visual learning and memory, visuospatial function, and working memory. A list of the tests by domain is provided in Table 2.

Table 2. Test battery identified by cognitive domain

<i>Tests</i>	<i>Measure abbreviation</i>
<i>Executive Function</i>	
PASAT	PASAT 2.4s, total correct
Trail Making Test Part B	Trails B
WCST	WCST Trials administered
<i>Language function</i>	
Boston Naming Test	Boston Naming Test total
Controlled Oral Word Association Test	FAS, total correct
<i>Motor</i>	
Grooved Pegboard	Grooved Pegboard, D & ND
<i>Processing speed</i>	
Digit-Symbol Coding from the WAIS-III	Digit Symbol Coding
Symbol Search from the WAIS-III	Symbol Search
Trail Making Test Part A	Trails A
<i>Verbal learning and memory</i>	
CVLT-II, List A Trial 1	CVLT, List A Trial 1
CVLT-II, Long-Delay Free Recall	CVLT, Long Delay Free
CVLT-II, Long-Delay Recognition	CVLT, Long Delay Recog
Logical Memory II from the WMS-III	Logical Memory II
<i>Visual learning and memory</i>	
RVLT Free Recall Trial 1	RVLT, Trial 1
RVLT Long Delay Free Recall Total	RVLT, Long Delay Free
RVLT Long Delay Recognition	RVLT, Long Delay Recog
Family Pictures II from the WMS-III	Family Pictures II
<i>Visuospatial function</i>	
Block Design from the WAIS-III	Block Design
<i>Working memory</i>	
Arithmetic from the WAIS-III	Arithmetic
Consonant Trigrams	CCC Total
Digit Span from the WAIS-III	Digit Span (F & B)
Letter-Number-Sequencing from the WAIS-III	Letter-Number Sequencing
Spatial Span from the WMS-III	Spatial Span

Note. PASAT = Paced Auditory Serial Addition Task. WCST = Wisconsin Card Sorting Test. WAIS-III = Wechsler Adult Intelligence Scale-III. WMS-III = Wechsler Memory Scale-III. CVLT-II = California Verbal Learning Test II. RVLT = Rey Visual Learning Test.

Procedure

After informed consent had been obtained, demographic information and past medical history were collected. The Quick Test (Ammons & Ammons, 1962) was first administered to estimate IQ, followed by the neuropsychological test battery and the POMS (McNair et al., 1992). All tests were administered at baseline (after surgery but before chemotherapy was initiated) and following the last chemotherapy cycle for the women in the relevant group, roughly six months after diagnosis. The hormonal and healthy control participants completed the test sessions at equivalent time points. All tests were administered by experienced individuals. Sessions lasted an average of three hours and took place at either the participant's home or the hospital, according to preference.

Data analyses

All analyses were conducted using the Statistical Package for the Social Sciences (SPSS; Version 15.0). Note that exact probability values are reported in all tables, and those denoted as significant (*) have not been adjusted for multiple comparisons.

Correlations. In order to evaluate redundancy among the 23 neuropsychological variables, we examined their intercorrelation matrix. This procedure was applied to each of the individual groups (healthy controls, hormonal, chemotherapy) at each time point (baseline and post-treatment phases).

A correlation of 0.9 and above suggests that the pair of variables is measuring the same construct. In such cases, eliminating one is recommended in order to strengthen the analysis (Tabachnick and Fidell, 2006). We adopted this criterion to maximize the variance accounted for. Correlations below this value indicate that the two tests measure different characteristics of the same cognitive function.

ANOVAs. A total of 23 mixed 3 x 2 ANOVAs were conducted to evaluate group differences over time on each of the neuropsychological measures.

MANOVAs. MANOVAs were performed to determine if there were group differences across domains.

Multivariate analyses of covariance (MANCOVAs). Included in these analyses as covariates were factors that are believed to influence cognitive functioning – age, education, estimated IQ, fatigue, and depression. The other POMS subscales were excluded from these analyses in the interest of reducing error.

Reliable change index. A modified version of the reliable change index (RCI; Jacobson & Truax, 1991), taking into account practice effects, was used to evaluate decline in neuropsychological function in the chemotherapy treatment group. Typically, this approach is used to identify the individuals who have shown a significant change in performance across time. The computation involves determining whether the individual's score falls outside the 90% confidence interval surrounding the baseline score adjusted for practice effects. This adjustment is made by adding the mean change score for the normative group to the baseline score. The 90% confidence interval is obtained by multiplying the standard deviation of the baseline – posttreatment difference by the appropriate *z*-score cutoff value (± 1.645). If the posttreatment score is greater than the upper limit of the confidence interval, then the individual is considered to have shown a significant improvement, and if less than the lower limit a significant decline.

This approach was applied to each of the 23 neuropsychological variables. Two separate sets of analyses were conducted, the first using the means and standard deviations of the healthy control group as an estimate for evaluating measurement error and practice effects and the second using the means and standard deviations of the hormonal group.

Performance across variables was summed, and participants were then classified according to the number of tests on which they had significantly declined (0 = no change; 1 = change on 1 test; 2 = change on 2 or more tests).

Standardized regression-based model. A standardized regression-based model (SRB; McSweeney et al., 1993), was used to evaluate decline in neuropsychological function in the chemotherapy group. This model not only provided adjustments for measurement error and practice effects, but assessed several covariates as well (age, education, estimated IQ, POMS subscales). The POMS subscale scores were included as covariates in the analyses when their correlation with any given neuropsychological measure was equal to or greater than .30 (Frigon & Laurencelle, 1993).

Like the RCI, this approach is used to identify which individuals have shown a significant change in performance across time. The method involves regressing the posttreatment performance on the baseline and the set of covariates to derive a predicted test score. A standardized change score is then obtained by subtracting the predicted score from the obtained score and dividing by the standard error of estimate. If the standardized score is greater than ± 1.64 , it is considered to fall in the extreme 5% at either end of the normal distribution and as such denotes a significant improvement or decline in performance.

This procedure was followed to assess whether the individuals in the chemotherapy group showed a significant decline on each measure. As for the RCI, two separate sets of analyses were conducted, one based on the healthy control group and one based on the hormonal group as an estimate for evaluating measurement error and practice effects. Performance across variables was summed, and participants were then classified according to the number of tests on which they had significantly declined (0 = no change; 1 = change on 1 test; 2 = change on 2 or more tests).

Following these individual-based analyses (RCI and SRB), chi-square tests of independence were used to evaluate whether the distribution of frequencies across the three categories varied as a function of group.

Note that all of the above methods were also used to measure improvement scores - that is, scores falling in the top 5% of the distribution. Because negligible results were obtained, these findings are not reported or discussed further here.

RESULTS

Descriptive statistics relating to age, education, estimated IQ at baseline, and test-retest interval are provided in Table 3. There were significant group differences between groups on education and estimated IQ.

Table 3. Descriptive statistics of groups

	<i>Healthy control</i>	<i>Hormonal group</i>	<i>Chemotherapy group</i>	<i>p value</i>
Age at baseline ^a	59.4 (4.1)	57.5 (4.2)	57.5 (3.9)	.088
Range	51-66	50-65	50-66	
Education at baseline ^a	16.0 (2.5)	14.0 (2.9)	14.4 (3.0)	.014
Range	12-21	9-20	8-23	
Estimated IQ at baseline	46.0 (2.4)	44.0 (3.1)	44.7 (3.1)	.013
Range	40-50	35-50	36-49	
Test-retest Interval ^b	151.9 (30.1)	158.9 (31.4)	144.6 (34.9)	.107
Range	98-218	123-245	91-245	

Note. IQ = intelligence quotient as measured by the Quick Test (Ammons & Ammons, 1962). Mean values shown, with standard deviations in parentheses.

^aIn years. ^bIn days.

Correlations

Of the 23 variables assessed, none of the pair-wise correlations was found to be higher than the cut-off value of .9. In addition to this criterion as a method for eliminating redundant variables, the same relationship for any pair of variables had to be observed across groups and test sessions. As neither of these requirements was met, all the variables were retained.

ANOVA

Means and standard deviations for all neuropsychological measures are provided in Table 4. The results of these analyses are reported in Table 5, with partial eta squared values in brackets. A total of 23 mixed ANOVAs were conducted to analyze group differences over time. The uncorrected alpha level yielded 7 main effects of group, 12 main effects of time, and 4 interactions: Family Pictures; Paced Auditory Serial Addition Test (PASAT); Letter-Number-Sequencing; California Verbal Learning Test-II (CVLT-II) List A, Trial 1. The Bonferroni correction reduced the number of significant results to 4 group effects, 8 time effects, and no interactions. Although there were instances of heterogeneity of variance (significant Levene's test results), the data were not normalized given the illustrative purposes of this paper. Figure 1 shows the typical pattern observed in the case of an interaction (refers to CVLT Long Delay Free Recall), improved performance in the control and hormonal groups on test repetition, and no evidence of a "practice effect" in the chemotherapy group.

Table 4. Means and standard deviations associated with each measure

<i>Test</i>	<i>Baseline</i>			<i>Postchemotherapy</i>		
	<i>Healthy control</i>	<i>Hormonal</i>	<i>Chemotherapy</i>	<i>Healthy control</i>	<i>Hormonal</i>	<i>Chemotherapy</i>
Spatial Span	15.5 ± 2.3	15.2 ± 2.5	15.4 ± 2.6	15.4 ± 2.5	15.4 ± 2.4	15.5 ± 2.5
Logical Memory II	29.6 ± 7.2	23.4 ± 7.4	26.8 ± 7.2	32.7 ± 7.2	26.8 ± 6.9	30.7 ± 5.4
Family Pictures	42.9 ± 9.4	41.9 ± 9.6	45.6 ± 9.2	48.7 ± 7.0	47.0 ± 8.4	46.3 ± 9.7
Trails A	25.3 ± 5.4	27.2 ± 9.1	27.6 ± 8.2	25.1 ± 5.4	26.3 ± 6.8	27.1 ± 9.4
Trails B	65.2 ± 15.8	70.2 ± 22.9	69.9 ± 25.8	59.8 ± 15.9	70.0 ± 21.8	67.1 ± 27.0
FAS, number of correct words	43.1 ± 13.5	37.6 ± 10.5	41.5 ± 12.4	44.4 ± 11.6	39.4 ± 10.8	40.3 ± 12.7
CVLT-II, Trial 1 Free Recall	6.75 ± 1.7	6.2 ± 1.3	7.2 ± 2.2	8.5 ± 2.4	7.3 ± 2.2	7.4 ± 2.3
CVLT-II, Long Delay Free Recall	13.5 ± 2.1	12.0 ± 2.6	12.4 ± 3.0	14.5 ± 1.6	12.6 ± 2.5	13.1 ± 2.5
CVLT-II, Long Delay Yes-No	15.2 ± 1.2	15.0 ± 1.3	15.2 ± 1.1	15.5 ± 0.9	15.0 ± 1.3	15.4 ± 1.0
PASAT 2.4s, total correct	41.8 ± 8.2	39.2 ± 9.9	41.6 ± 9.9	47.1 ± 7.3	41.2 ± 9.3	43.4 ± 9.5
RVLT, Trial 1 Free Recall	5.4 ± 1.7	3.9 ± 1.5	4.7 ± 1.6	6.1 ± 1.8	4.9 ± 1.8	5.1 ± 1.8
RVLT, Long Delay Free Recall	8.9 ± 2.5	7.1 ± 2.3	8.1 ± 2.4	9.6 ± 2.2	8.1 ± 2.5	8.7 ± 2.5
RVLT, Long Delay Recognition	13.1 ± 1.2	12.5 ± 1.4	13.1 ± 1.1	13.5 ± 1.1	13.1 ± 1.2	13.1 ± 1.4
Digit Symbol Coding	72.3 ± 11.3	68.9 ± 11.3	67.7 ± 12.5	75.6 ± 10.2	68.7 ± 11.9	68.2 ± 13.4
Block Design	38.6 ± 10.5	22.4 ± 9.2	37.2 ± 11.4	40.1 ± 9.9	34.4 ± 9.5	38.4 ± 11.8
Arithmetic	15.5 ± 2.9	13.7 ± 2.9	13.7 ± 3.6	15.8 ± 2.9	13.8 ± 3.1	14.0 ± 3.6
Digit Span, Forward + Backward	18.1 ± 3.9	17.1 ± 4.2	17.0 ± 4.1	17.6 ± 3.5	18.0 ± 3.4	17.2 ± 4.0
Symbol Search	30.5 ± 5.9	30.3 ± 5.0	30.9 ± 6.6	32.1 ± 5.3	30.5 ± 4.9	31.3 ± 6.9
Letter-Number Sequencing	10.4 ± 2.6	10.5 ± 2.2	10.8 ± 2.8	11.3 ± 2.5	11.0 ± 2.3	10.3 ± 2.6
WCST, no. of trials, raw score	93.8 ± 22.4	104.4 ± 23.8	99.7 ± 22.0	92.3 ± 23.7	101.7 ± 23.7	97.4 ± 21.9
CCC, total	49.6 ± 5.8	43.0 ± 7.3	43.4 ± 7.6	51.3 ± 5.1	44.7 ± 7.7	43.1 ± 8.7
Boston Naming Test, total	56.9 ± 3.0	54.3 ± 5.4	55.3 ± 4.7	57.8 ± 2.5	55.2 ± 5.0	56.2 ± 4.6
Groove Pegboard, combined	151.4 ± 17.2	155.4 ± 23.3	153.3 ± 42.1	163.9 ± 28.2	153.2 ± 26.0	151.0 ± 34.4

Note. Abbreviations are explained in Table 2.

Table 5. Probability levels associated with results of mixed ANOVAs

<i>Test</i>	<i>Group</i>	<i>Time</i>	<i>Group x Time</i>
Logical Memory II	< .001 (.128)	< .001 (.286)	<i>ns</i>
Family Pictures	<i>ns</i>	< .001 (.193)	.007 (.080)
PASAT 2.4s, total correct	<i>ns</i>	< .001 (.216)	.050 (.062)
RVLT, Trial 1	.001 (.108)	< .001 (.157)	<i>ns</i>
RVLT, Long Delay Free Recall	.006 (.081)	< .001 (.137)	<i>ns</i>
RVLT, Long Delay Recognition	<i>ns</i>	.006 (.062)	<i>ns</i>
Digit-symbol Coding	<i>ns</i>	.032 (.038)	<i>ns</i>
Block design	.050 (.049)	.046 (.033)	<i>ns</i>
Arithmetic	.028 (.058)	<i>ns</i>	<i>ns</i>
Letter-Number Sequencing	<i>ns</i>	<i>ns</i>	.010 (.073)
Boston Naming Test, total	<i>ns</i>	< .001 (.170)	<i>ns</i>
CVLT-II List A, Trial 1*	<i>ns</i>	< .001 (.170)	.013 (.070)
CVLT-II Long Delay Free Recall*	.014 (.070)	< .001 (.129)	<i>ns</i>
CCC total*	< .001 (.157)	.031 (.040)	<i>ns</i>

Note. Only significant effects reported; values in parentheses represent partial eta squared. ANOVA = analysis of variance. PASAT = Paced Auditory Serial Addition Task. RVLT = Rey Visual Learning Test. CVLT-II = California Verbal Learning Test II.

* Significant Levene's *F* test, $p < .05$

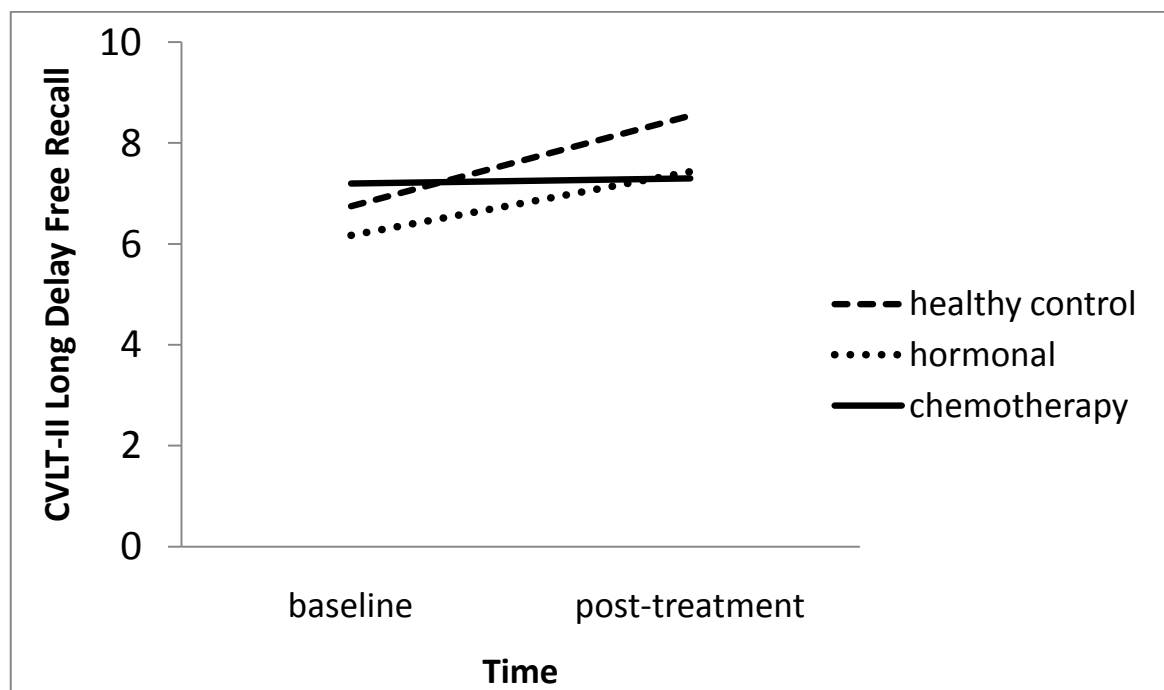


Figure 1. Typical interaction pattern observed (ANOVA).

MANOVA/MANCOVA

The significant results of these analyses are presented in Tables 6 and 7; as above, values in brackets represent partial eta squares. Using an adjusted alpha level ($p \leq .006$), the MANOVA produced significant group differences in verbal learning and working memory, four significant time effects, and no interactions. Adjusted alpha levels rendered all effects for the MANCOVA nonsignificant. The IQ covariate was significant for language and visual learning, education for verbal learning and working memory, and age for executive function, language, processing speed, and visual learning.

Table 6. Probability levels associated with results of MANOVAs

<i>Factor</i>		<i>Group</i>	<i>Time</i>	<i>G x T</i>
1. Executive function	PASAT 2.4s, total correct, Trails B, WCST Trials administered	ns	< .001 (.265)	NS
2. Language function	Boston Naming Test, total FAS, total correct	NS	< .001 (.177)	NS
3. Motor	Grooved Pegboard, D & ND	NS	NS	NS
4. Processing speed	Digit-Symbol Coding, Symbol Search, Trails A	NS	.028 (.074)	NS
5. Verbal learning and memory	CVLT-II, List A, Trial 1, CVLT-II, Long Delay Free Recall, CVLT-II Long Delay Recognition, Logical Memory II	.006 (.090)	< .001 (.372)	NS
6. Visual learning and memory	RVLT, Trial 1, RVLT, Long Delay Free Recall, RVLT, Long Delay Recognition, Family Pictures	.035 (.069)	< .001 (.300)	NS
7. Visuospatial function	Block Design	NS	.046 (.033)	NS
8. Working memory	Arithmetic, CCC total, Digit Span (F & B), Letter-Number Sequencing, Spatial Span	.002 (.117)	NS	NS

Note. Significance values based on Wilks' λ ; values in parentheses represent partial eta squared. ns refers to $p \geq .05$. MANOVA = multivariate analysis of variance. PASAT = Paced Auditory Serial Addition Task. WCST = Wisconsin Card Sorting Test. CVLT-II = California Verbal Learning Test II. RVLT = Rey Visual Learning Test.

Table 7. Probability levels associated with results of MANCOVAs

<i>Factor</i>		<i>Group</i>	<i>Time</i>	<i>Group x Time</i>	<i>Covariates</i>		
					<i>Age</i>	<i>IQ</i>	<i>Education</i>
1. Executive function	PASAT 2.4s, total correct Trails B, WCST, Trials administered	<i>ns</i>	<i>ns</i>	<i>ns</i>	.048 (.095)		
2. Language function	Boston Naming Test, total FAS, total correct	<i>ns</i>	<i>ns</i>	<i>ns</i>	.016 (.074)	<.001 (.295)	
3. Motor	Grooved Pegboard, D & ND	<i>ns</i>	.025 (.056)	<i>ns</i>	<i>ns</i>	<i>ns</i>	
4. Processing speed	Digit-Symbol Coding Symbol Search, Trails A	<i>ns</i>	<i>ns</i>	<i>ns</i>	.013 (.096)		
5. Verbal learning and memory	CVLT-II, List A, Trial 1, CVLT-II, Long Delay Free Recall, CVLT-II, Long Delay Recognition Logical Memory II	.032 (.077)	<i>ns</i>	<i>ns</i>			.004 (.140)
6. Visual learning and memory	RVLT, Trial 1, RVLT, Long Delay Free Recall, RVLT, Long Delay Recognition Family Pictures	<i>ns</i>	<i>ns</i>	<i>ns</i>	.020 (.105)	.040 (.091)	
7. Visuospatial function	Block Design	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>		
8. Working memory	Arithmetic, CCC total, Digit Span (F & B), Letter-Number Sequencing, Spatial Span	<i>ns</i>	<i>ns</i>	<i>ns</i>		.031 (.116)	.014 (.133)

Note. Significance values based on Wilks' λ ; values in parentheses represent partial eta squared. *ns* refers to $p \geq .05$. MANOVA = multivariate analysis of variance. PASAT = Paced Auditory Serial Addition Task. WCST = Wisconsin Card Sorting Test. CVLT-II = California Verbal Learning Test II. RVLT = Rey Visual Learning Test.

RCI and SRB

Figure 2 presents the percentage of participants in each group who declined on neuropsychological tests using the healthy control group as a comparison; Panel a shows the results based on the RCI analyses, and Panel b shows the results of the SRB analyses. Visual inspection of the figures indicates that the RCI and SRB results produce different patterns. Decline on two or more tests occurred more frequently (roughly 25% higher in the chemotherapy group and 46% in the hormonal group) using the SRB compared to the RCI technique. Chi-square tests of independence produced a significant difference across groups as a function of the number of tests on which decline was observed (0, 1, or 2 tests; RCI $\chi^2 = 12.03, p = .017$; SRB $\chi^2 = 24.57, p < .001$).

Figure 3 presents the percentage of participants in each group who declined on neuropsychological tests using the hormonal group as a comparison. As above, Panel a shows the results based on the RCI analyses, and Panel b shows the SRB analyses. These figures illustrate that using the hormonal group as a comparison, the results of the RCI and SRB are relatively consistent. Roughly 62% of the participants in the hormonal group declined on 0 or 1 test compared to approximately 50% in the chemotherapy group (RCI $\chi^2 = 20.24, p < .001$; SRB $\chi^2 = 3.89, p = .42$). Relative to the RCI analyses, the SRB increased the number of participants who declined on two or more tests from 33 to 39% in the hormonal group and decreased the number from 55% to 48% in the chemotherapy group.

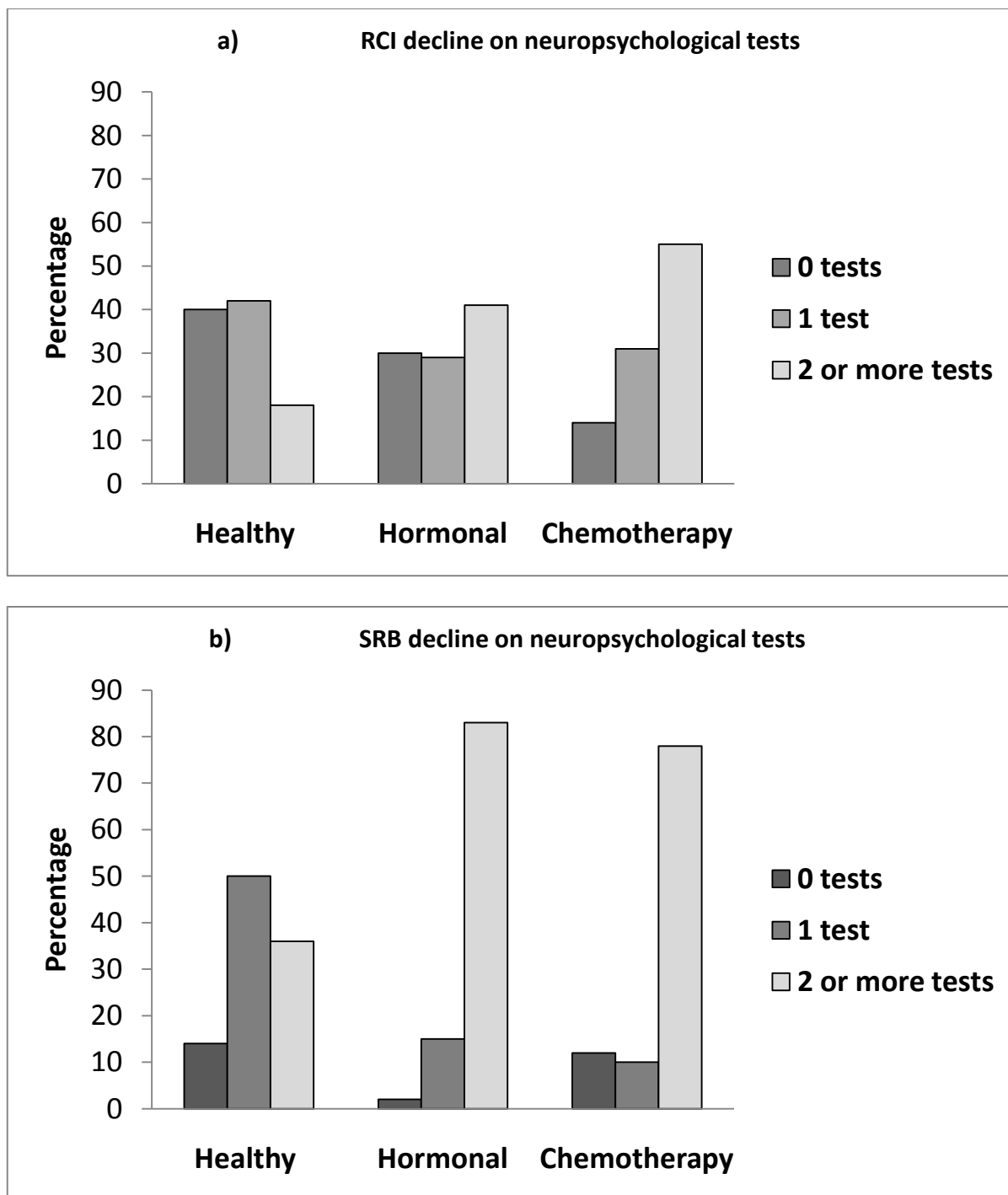


Figure 2. Percentage of participants declining on 0, 1 and 2 or more neuropsychological tests based on comparison with the healthy control group. The upper panel (a) presents the results based on the RCI analyses and the lower panel (b) the SRB analyses.

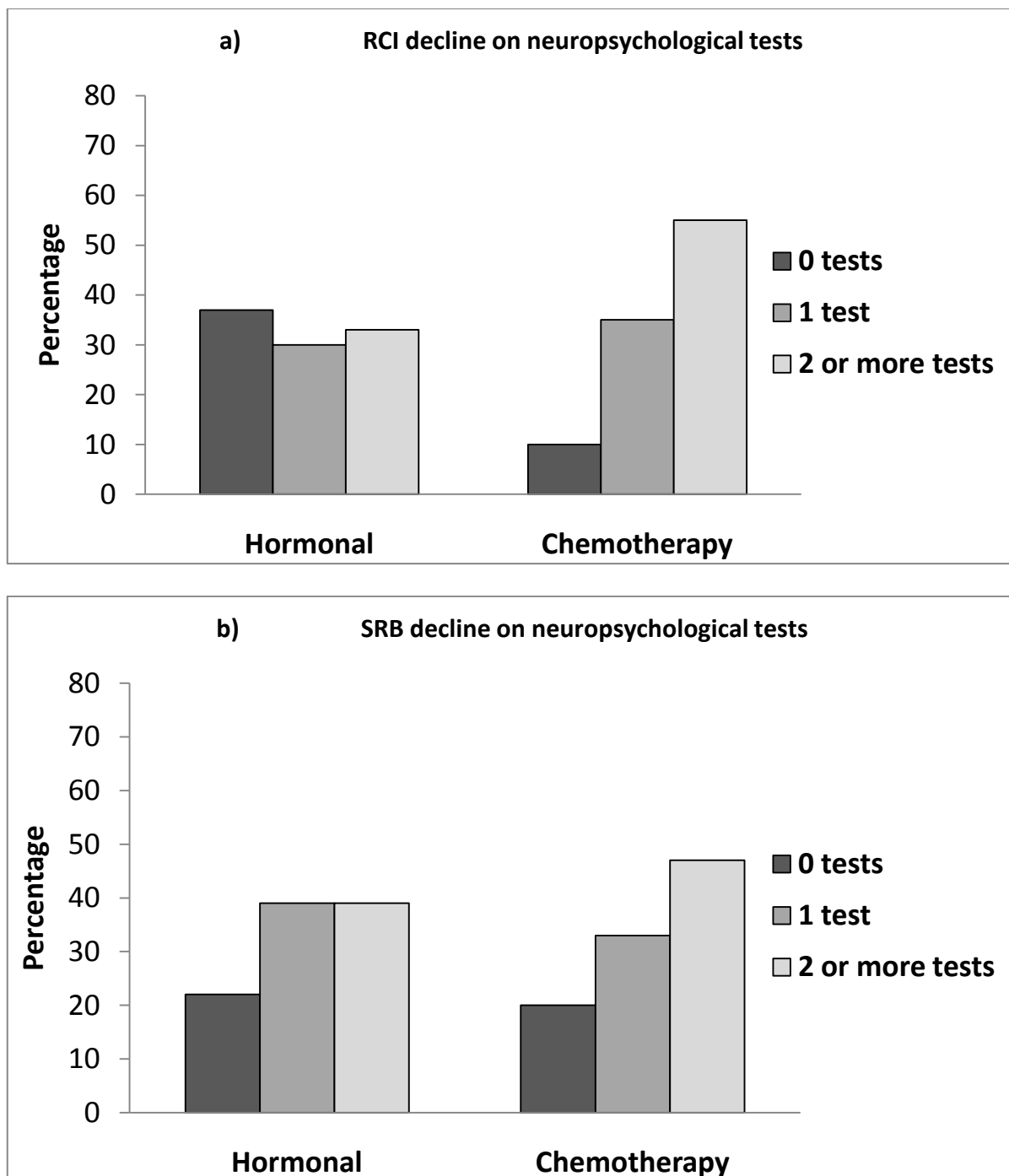


Figure 3. Percentage of participants declining on 0, 1 and 2 or more neuropsychological tests based on comparison with the hormonal group. The upper panel (a) presents the results based on the RCI analyses and the lower panel (b) the SRB analyses.

The covariates (age, education, estimated IQ, and POMS subscales) used in the SRB analyses were entered into the model first, followed by the post-treatment score. They accounted for a significant proportion of the variance in the Logical Memory, $R^2=.29$, $F\Delta(3,24)=3.22$, $p=.04$, and Symbol Search, $R^2=.29$, $F\Delta(3,24)=3.20$, $p=.04$, subtests of the WAIS-III. Using the hormonal group as a comparison, the covariates accounted for a significant proportion of the variance in the CVLT-II Long delay free recall, $R^2=.26$, $F\Delta(5,39)=2.75$, $p=.03$; Rey Visual Learning Test (RVLT) long delay free recall, $R^2=.24$, $F\Delta(3,42)=4.39$, $p=.01$; RVLT long delay recognition, $R^2=.26$, $F\Delta(3,42)=4.91$, $p=.01$; FAS test, $R^2=.28$, $F\Delta(3,42)=5.51$, $p=.003$; Wisconsin Card Sorting Test (WCST), $R^2=.25$, $F\Delta(5,37)=2.51$, $p=.047$; and Boston Naming Test (BNT) tests, $R^2=.44$, $F\Delta(4,41)=8.02$, $p=.001$.

Figure 4 presents the breakdown of participants in each group who declined on each of 0 to 9 tests in both the RCI (left side) and SRB (right side) analyses based on comparison with the healthy control group. This figure illustrates the sensitivity of the SRB method in identifying patients who decline on as many as 9 tests. Indeed, just over 8% of chemotherapy and 2% of hormonal therapy participants declined on 35% of the measures (8 or 9 out of 23). The majority of these participants failed to be detected using RCI methodology.

Note that due to space limitations, the SPSS syntax for performing the standardized regression-based analysis was not included but is available upon request.

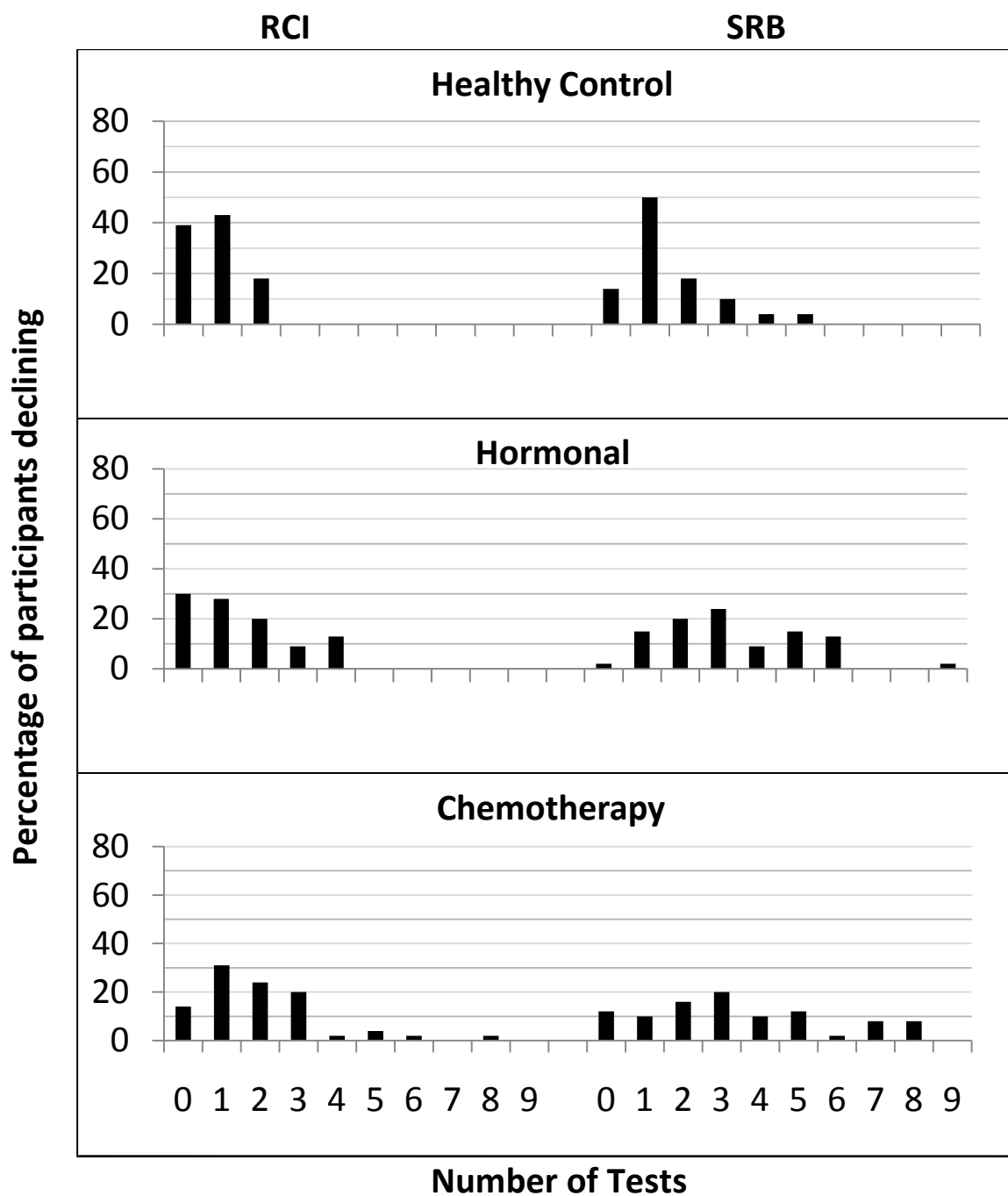


Figure 4. Percentage of participants declining on 0 – 9 neuropsychological tests based on comparison with the healthy control group. The left side of each graph presents the results based on the RCI analyses and the right side, the SRB analyses.

DISCUSSION

The goal of this paper was to evaluate the diversity of statistical methods used to detect the effects of chemotherapy treatment on cognitive function based on neuropsychological assessments, while taking into account methodological issues and the clinical relevancy of the findings. To illustrate the strengths and weaknesses of these methods, we applied them to our investigations on the phenomenon of “chemo-fog” (Stewart et al., 2008a). Univariate methods produced several group differences, mainly in memory-related domains but no interactions once the adjustment for multiple comparisons was applied. Similarly, multivariate analyses resulted in a group difference in working memory but no indication that changes in performance differed across groups from baseline to posttreatment evaluations. It is important to note that the results of the hormonal group were similar to those obtained in the chemotherapy group, although fewer hormonal participants showed decline on two or more tests. This supports prior research that has found cognitive impairments related to the use of adjuvant hormonal treatment (Bender et al., 2006; Castellon et al., 2004; Jenkins et al., 2004; van Dam et al., 1998).

Overall, our results are reasonably consistent with other investigations in this area – generally small effects in memory-related tests (Ahles et al., 2002; Brezden et al., 2000; Castellon et al., 2004; Schagen et al., 1999; van Dam et al., 1998; Wefel et al., 2004a; Wieneke & Dienst, 1995). The effect sizes associated with these results were in the order of 10% in ANOVA tests. In addition, although the education, age, and IQ covariates were found to be significant in some analyses (see Table 7), they had relatively low effect sizes indicating minimal contribution to the overall variance.

In contrast, both RCI and SRB methods generated significant differences between groups in the frequency of cases classified as declined. Recall that decline was defined as a change in performance between the predicted and observed score (divided by the standard error of prediction) that was greater than the criterion value (i.e. falling outside the lower boundary of the 90% confidence interval). Compared to the control group, in which an average of 25% (18% RCI vs. 35% SRB) of participants were identified with the decline category, the percentage of frequencies generated by the treatment arms (hormonal and chemotherapy) in this category ranged from 40% to 83% (Figure 4). Furthermore, there were notably more participants who scored below the criterion on two or more tests using the SRB method (hormonal group, 40% RCI vs. 82% SRB; chemotherapy group, 56% RCI vs. 78% SRB).

There are several types of analysis at the individual level that have been proposed (Heaton et al., 2001; Temkin, Heaton, Grant, & Dikmen, 1999) but the most commonly used is the RCI. First introduced over 20 years ago to evaluate meaningful change in psychotherapy outcome research (Jacobson, Follette, & Revenstorf, 1984), the method has been applied in one of two ways: with and without a correction for practice effects. The latter is preferred because practice effects can mask declines on measures. However, a weakness of this approach is that it employs a mean practice effect based on the comparison group, which assumes that all participants are experiencing the same degree of benefit from repeated test exposure (Sawrie, 2002).

Three of the articles reviewed for this paper were based on RCI analyses (Jenkins et al., 2006; Shilling et al., 2005; Wefel et al., 2004a). All three found that a small subset of

women experience significant cognitive decline after chemotherapy treatment, particularly in the areas of memory, attention and concentration.

The SRB method, a more recent advancement (McSweeney et al., 1993), has been used in the medical literature to statistically assess cognitive change in epilepsy patients (LoGalbo et al., 2005; Martin, Griffith, Sawrie, Knowlton, & Faught, 2006; Martin et al., 2002), cardiac patients (Raymond, Hinton-Bayre, Radel, Ray, & Marsh, 2006) and sports-related concussion (McCrea et al., 2005), and in the neuropsychological literature to assess age-related cognitive decline (Frerichs & Tuokko, 2006; Isella et al., 2003) and in our group to evaluate cognitive impairment related to chemotherapy treatment for breast cancer (Stewart et al., 2008a).

Unlike the RCI method, this technique does not incorporate an equal practice effect for all participants. Due to factors such as regression towards the mean two individuals could have the same raw change score but, depending on their baseline score, different SRB change scores. The method also allows for the inclusion of covariates, in order to evaluate factors (e.g., fatigue, education, IQ, age) that might influence test-retest scores (Daly, Kormaroff, Bloomingdale, Wilson, & Albert, 2001; Tiersky, Johnson, Lange, Natelson, & DeLuca, 1997). We found that these covariates were significantly related to decline on four tests (Block Design, FAS, WCST and BNT; see results section). Although they were not significant for the remaining tests, such variables have been shown to be important in other groups (Lezak, Howieson, & Loring, 2004) and should continue to be included in future analyses.

Frigon and Laurencelle (1993) suggest that even correlations of .30 between covariates and the dependent variable(s) may have an impact on the final model of an

analysis, even if they are small and nonsignificant. The method by which covariates are entered in the regression analysis affects the results obtained in the final model. Some researchers employ a stepwise approach where the inclusion or exclusion of a covariate is determined on a statistical basis, with nonsignificant covariates being dropped out of the final model. Other researchers used a forced-entry method where covariates are retained in the model regardless of their statistical significance.

The disadvantage of the stepwise approach is that the results are specific to the data set under evaluation and therefore less generalizable. Another issue is that the number of covariates can vary widely from one variable to another, leading to questions about the model's theoretical underpinnings. For example, fatigue might be found to be a significant covariate for one test but nonsignificant for a subsequent one, posing difficulties for understanding how fatigue could affect performance on one test, but not the next. Forced entry of the covariates eliminates these problems, but causes a decrease in power because one degree of freedom is lost for each covariate entered. This may be problematic when sample sizes are very small, but should not be a major deterrent for larger samples.

No analysis is perfect, a truism for both RCI and SRB methodologies. One major drawback as noted by Sawrie (2002) is the family-wise Type I error rate when a large number of measures are used. In these circumstances, it is highly probable that there will be at least some misclassifications using this model; the more measures there are in a battery, the greater the number of participants who will be misclassified. Given this, it will be important for future research to clarify the cognitive domains that are most affected by chemotherapy treatment and to determine which tests will best detect cognitive decline. Through the selection of instruments with strong psychometric properties and appropriate

norms, neuropsychological batteries could be limited to a smaller number of measures for each domain to avoid redundancy, while reducing the risk of misclassification.

Additional methodological issues that need to be clarified in order to improve the degree of standardization in studies of this nature include the following: a clear statement of the definition of impairment, ideally the use of prospective designs, and suitable control groups.

One of the key issues that requires consensus is the operationalization of cognitive impairment and the criteria used to classify participants as impaired. We suggest that individual scores be compared to those of an appropriate control group or, in its absence, normative test data, and categorized based on a criterion – for example, a score that falls in the lower 5% or 1% of the distribution. Regardless of the definition, the way in which cognitive impairment is defined for the purposes of statistical analyses needs to be clearly stated in the Methods section.

As noted in the Introduction, prospective repeated measures designs are becoming the preferred choice because they provide researchers with the baseline data needed to evaluate cognitive decline. Although standardized tests often include normative data that could be used for this purpose, age-appropriate norms may not be available, and normative samples are unlikely to closely match the demographics of the participants in a research study and therefore are not the ideal reference group. For example, a high-functioning participant's score might drop substantially from baseline to post-treatment, but may still not be considered impaired because the score remains within the average range based on comparison with normative data.

In the absence of serial testing, some investigators use a measure of premorbid IQ as a baseline reference to measure subsequent decline. For instance, the National Adult Reading Test has been used for many years to assess cognitive decline in a variety of contexts (Dick et al., 2004; Joyce, Hutton, Mutsatsa, & Barnes, 2005; Nelson & McKenna, 1975; Peace, Orme, Padayatty, Godfrey, & Belchetz, 1998). Although measures such as this were developed in part as a means of providing neuropsychologists with a way to compare an individual's current test performance with pre-morbid functioning, they are not ideal because they extrapolate based on very limited information. In addition, they preclude the use of detailed individual based analyses because they do not provide sufficient information for computing decline in specific cognitive domains.

Related to this issue is the nature of the control group. Sawrie (2002) notes that the quality of RCI analyses is largely dependent on how closely the control group matches the characteristics of the treatment group; careful matching ensures that any significant differences in performance are not due to preexisting factors. On this basis, it might be argued that a healthy control group is not the best choice for comparison to evaluate cognitive decline in chemotherapy recipients and that some form of cancer control group would be better. However, this scenario can result in an underestimation of the degree of impairment, as participants are not being evaluated based on a healthy state. From a methodological perspective, randomized groups would be ideal but are not feasible due to ethical considerations. Barring this, the inclusion of both healthy and disease control groups would allow investigators to distinguish disease from treatment effects. In practice, issues related to recruitment, time, and costs may be prohibitive. A compromise would be to include for example, a cancer control group along with a measure of premorbid IQ.

CONCLUSIONS

Many researchers in this area continue to perform group analyses on their data. Although appropriate in exploratory investigations, they prove less useful once parameters have been established. It is now known that there are subtle group differences in cognitive function between chemotherapy recipients and diseased or healthy control groups (Ahles et al., 2002; Ahles et al., 1996; Bender et al., 2006; Brezden et al., 2000; Castellon et al., 2004; Kreukels et al., 2005; Schagen et al., 1999, 2001; Shilling et al., 2005; Stewart et al., 2006, 2008a; Tchen et al., 2003; van Dam et al., 1998; Wefel et al., 2004a; Wieneke & Dienst, 1995). At this point, group comparisons are less useful because they do not provide adequate power to detect interactions, the effects of interest, particularly once an adjustment for multiple comparisons has been applied. From a clinical perspective, the degree of impairment in a subset of subjects identified by these individual-based methods might be expected to have a major impact on quality of life for those affected and provide clinicians with some guidelines for treating individual cases.

Work is still needed to evaluate subjective measures of cognitive impairment in recipients of chemotherapy and hormonal therapy in order to assess the clinical relevance of any declines in function and impact on quality of life. It would also be useful to confirm whether the domains that have been identified as impaired based on neuropsychological testing and statistical analyses are consistent with those reported by the participants. Only one study to our knowledge has evaluated the correspondence between self-reported symptoms and scores on neuropsychological tests in breast cancer patients (Shilling & Jenkins, 2007). They found little association between objective and subjective measures and

recommended that a neuropsychological battery that captures the cognitive capacity associated with everyday functioning be developed.

Finally, there are several methodological issues that need to be considered. These include a clear definition of cognitive impairment to facilitate comparison and replication across studies; selection of an appropriate control group to minimize confounding variables; a psychometrically sound test battery that focuses on functions previously identified in the literature as impaired in order to minimize statistical error; and prospective designs that assess premorbid function as a baseline for measuring decline. More rigorous research methods will ensure that results are not only clinically relevant, but valid, reliable and meaningful as well.

Acknowledgements

We would like to offer our sincere thanks to the women who generously contributed their time to participate in this study, to the reviewers who provided valuable feedback, and to the Canadian Breast Cancer Foundation for their support.

Note

Examples of the syntax used for the reliable change index and standardized regression-based model are included in Appendixes C and D respectively.

STUDY 2

Submitted to the Archives of Clinical Neuropsychology: Ouimet, L. A., Stewart, A.,

Collins, B., Schindler, D., & Bielajew, C. Investigation of a Neuropsychological Screen for
Chemotherapy-Related Cognitive Impairment.

**Investigation of a Neuropsychological Screen for
Chemotherapy-Related Cognitive Impairments**

Lea Ann Ouimet, Angela Stewart¹, Barbara Collins²,
Dwayne Schindler, Catherine Bielajew

School of Psychology, University of Ottawa, ON, Canada.

¹Royal Ottawa Health Care Group, Ottawa, ON, Canada

²Ottawa Hospital, Civic Campus, Ottawa, ON, Canada.

ABSTRACT

Research on chemotherapy-induced cognitive impairment (the term “chemo-fog” is used by many investigators) supports the occurrence of subtle declines in function for a subset of recipients. Identification of vulnerable individuals via comprehensive neuropsychological batteries is complicated due to their lack of clinical utility and increased risk of misclassification. The goal of this paper was to evaluate the ability of a reduced battery to detect chemotherapy-related cognitive impairments. Data from our previous study (Ouimet et al., 2009) were used to compare a comprehensive neuropsychological test battery comprising 23 tests with a reduced battery consisting of a subset of nine tests. A standardized regression-based approach revealed that a comparable numbers of participants were identified by both batteries, suggesting that individuals vulnerable to chemotherapy-induced cognitive impairment can be identified by a more selective battery. Further work is needed to clarify the neuropsychological tests most sensitive to detecting impairments associated with chemotherapy so that assessment batteries can be limited to these tests.

INTRODUCTION

Breast Cancer strikes approximately one in nine women each year in North America (Canadian Cancer Society, 2010; American Cancer Society, 2010) and the majority will receive some form of chemotherapy as part of their treatment regimen. In the last few years, there has been increasing evidence to suggest that a subset of these women suffer from chemotherapy-related cognitive impairments that primarily affect memory, attention, concentration, and perseverance (Abraham et al., 2008; Collins et al., 2008; Debess et al., 2010; Ferguson et al., 2007; Hermelink et al., 2007; Hurria & Lachs., 2006; Jansen et al., 2007; Jim et al., 2009; Kesler et al., 2009; Reid-Arndt et al., 2009; Stewart et al., 2008a; Tager et al., 2010; Tallibert et al., 2007).

While from a statistical standpoint the effects have generally been small; from a clinical perspective, the degree of impairment noted in both self-report research and those using objective measurement might reasonably be expected to have major impacts on quality of life. In fact, deleterious impact of cognitive decline from chemotherapy on quality of life in breast cancer patients has been reported by various researchers (Ahles et al., 2005; Downie et al., 2006; Fallowfield et al., 1999; Ferguson et al., 2007; Kayl et al., 2006; Mar Fan et al., 2005; Mehnert et al., 2007; Reid-Arndt et al., 2009; Taillibert et al., 2007; Von Ah, Russell, Storniolo, & Carpenter, 2009; Weis et al., 2009).

Given the incidence of breast cancer, and the standard inclusion of some form of chemotherapy as part of the treatment regimen, if quality of life is negatively impacted, then the identification of chemotherapy-related cognitive impairments becomes an issue of clinical relevance. Unfortunately, various issues in the area of neuropsychological

assessment complicate this process, making it difficult to accurately identify the subgroup of women affected.

Although there is a substantial literature on test theory and professional guidelines for the selection of individual tests, a salient issue affecting neuropsychology is the lack of theory underlying the optimal composition of a neuropsychological test battery. As a result, there is wide variability in the choice of tests used to measure various cognitive domains, making it difficult to accurately identify the areas of functioning that are most compromised following chemotherapy treatment. Compounding the problem is the very nature of cognitive functions; they are complex and overlapping, making it a considerable challenge to choose tests that are as “clean” as possible (i.e. measuring one specific construct).

Given these limitations, it is not surprising that there are few professional guidelines or policies available to instruct clinicians on the most effective and reliable means of constructing a neuropsychological test battery. As a result, tests may be chosen as a function of availability, tradition, familiarity, or personal preference rather than evidence of reliability, validity, or statistical support.

Another important consideration is the historical impetus for the development of neuropsychological tests, which was in fact, to assist in diagnostics, not in the prediction of outcome. However, assessments are used commonly today for aiding in the determination of expected future performance, achievement, or recovery and to evaluate whether or not cognitive changes occur over time. Whether or not or to what extent neuropsychological tests designed for diagnostics retain their validity and reliability when used for predictive purposes can only be answered through continued research on these aspects.

There is also some controversy regarding the choice of fixed versus flexible test batteries (Wong, 2006). A fixed battery refers to an assembled, normed set of tests, in contrast to flexible batteries, which are chosen and adjusted based on a dynamic hypothesis regarding the nature of the presenting problem, as well as the patient's performance throughout the assessment. A third approach, the intermediate position (Strauss, Sherman & Spreen, 2006), refers to a core battery that is supplemented by tests relevant to the diagnosis of a particular population who show specific deficits (e.g. Alzheimer's patients) as well as the patient's performance.

A related issue that further complicates matters is the number of tests used in any given neuropsychological test battery. From a research perspective, as the number of tests multiplies, so does the likelihood of redundancy and multi-collinearity. Adding more tests also reduces power and the probability of finding statistical significance, thereby escalating Type II errors. Practically, a longer battery may decrease participation overall, and raise attrition in longitudinal studies. But perhaps the most critical disadvantage of a large battery, from both a research and clinical perspective, is the risk of misclassification.

Viewed from a purely clinical standpoint, a larger battery increases the time required for administration and introduces potentially conflicting explanations for poor test performance such as fatigue, loss of attention and focus, and decreased motivation.

Although all of these issues pertain to the area of breast cancer research, one that is of particular clinical relevance in this area is the size of the battery used. They typically consist of a large number of tests, and while this may be appropriate in a research context, they have limited clinical utility because they are too big to be of practical use for screening in health care settings. Consequently, the objective of this paper is to use the data obtained in our

previous study (Ouimet et al., 2009) to aid in the construction of a smaller neuropsychological test battery for screening “chemo-fog”. In the earlier study we found that a subset of participants in both the chemotherapy and hormonal groups experienced a significant decline in function (greater than two standard deviations) on two or more tests in a neuropsychological test battery.

Specifically, this paper compares a neuropsychological test battery comprising 23 tests representing the major cognitive domains (executive, language, motor, processing speed, verbal learning and memory, visual learning and memory, visuospatial, working memory) with a subset of nine tests, in order to evaluate the sensitivity of the reduced battery to chemotherapy-induced cognitive impairments. For the remainder of this paper, our original 2009 data and results (Ouimet et al., 2009) will be referred to as study 1 and the current paper will be referred to as study 2.

Model

To determine if the reduced neuropsychological test battery was sensitive in detecting impaired performance, standard regression-based analyses (SRB; McSweeney et al., 1993) were applied to assess whether the individuals in the hormonal and chemotherapy groups show a significant decline across time. SRB analyses were chosen as opposed to the reliable change index because SRB takes into account regression and individual practice effects. This method also allows for the inclusion of covariates.

Whereas the previous study consisted of two separate sets of SRB analyses using the means and standard deviations of both a healthy control group and a hormonal group as an estimate for evaluating measurement error and practice effects, the current study used only the healthy control group.

This decision was based on an evaluation of the results obtained in the first study in that the differences between the chemotherapy and hormonal groups were less than the differences between either of these groups and a healthy control reference. As a result, the use of the hormonal group as a basis for comparison in this context is limited because of the risk of underestimating cognitive decline, thereby increasing the risk of misclassification. To minimize this risk, the healthy control group was used for comparison.

METHOD

Participants

Data were obtained from 123 stage I or II breast cancer patients who were recruited as part of a larger longitudinal study conducted out of the Ottawa Regional Cancer Centre to investigate the neuropsychological effects of adjuvant chemotherapy. All participants had undergone mastectomy or lumpectomy and were receiving chemotherapy (with or without hormonal treatment, $n = 49$) or hormonal treatment alone ($n = 46$). Details regarding disease stage, treatment regimens, and number of chemotherapy cycles can be found in Table 8 below.

Table 8. Clinical characteristics of treatment groups

Characteristic	Chemotherapy (N = 49)	Hormonal (N = 46)	<i>p</i>
Stage of disease			<.001
I	(n = 13)	(n = 39)	
II	(n = 34)	(n = 5)	
III	(n = 2)	(n = 0)	
Type of chemotherapy			
FEC-100	(n = 27)	NA	NA
AC/AC-Taxol	(n = 13)		
CEF	(n = 3)		
FAC	(n = 3)		
Other	(n = 3)		

Number of chemotherapy cycles			
Mean (SD)	5.69 (1.17)	NA	NA
Min/Max	4 - 10		
FEC-100 = Fluorouracil, epirubicin, cyclophosphamide; AC = doxorubicin, cyclophosphamide; AC-Taxol = doxorubicin, cyclophosphamide, taxol; CEF = cyclophosphamide, epirubicin, fluorouracil; FAC = cyclophosphamide, doxorubicin, fluorouracil.			

In addition, a healthy control group ($n = 28$) was recruited through advertisements posted on hospital bulletin boards. This group was included to examine practice effects and to facilitate evaluation of these effects in the treatment groups as well as to provide a means of evaluating decline from a healthy state. Participants received \$50 for each test session.

Sample Selection

The sample was limited to postmenopausal women aged 65 years or younger as a means of minimizing the potential confounds of hormonal status and age-related cognitive decline. Fluency in English was required in order to complete the test battery. Exclusion criteria for all groups included a history of previous cancer and chemotherapy or radiation treatment and serious psychiatric disorder (e.g. major depression, schizophrenia), neurological illness, or significant substance abuse due to their potential negative effects on cognitive function. The board of ethics at the Ottawa Hospital approved this study and written informed consent was obtained from all participants.

Measures

Our original study included a total of 23 psychometrically sound neuropsychological tests representing the major cognitive domains: executive function, language function, motor skills, processing speed, verbal learning and memory, visual learning and memory, visuospatial function, and working memory (please see Ouimet et al., 2009 for a detailed

list). Nine subtests were retained for the present paper, including: Trails A & B, FAS (Controlled Oral Word Association Test), Wechsler Adult Intelligence Scale-III Digit Symbol Coding, Symbol Search and Letter-Number Sequencing subtests, California Verbal Learning Test-II Long Delay Free Recall and Long Delay Recognition and Consonant Trigrams. Means and standard deviations for each test are reported in Table 9 (see page 6).

Table 9. Means and standard deviations associated with each measure in the reduced battery

<i>Test</i>	<i>Baseline</i>			<i>Post-chemotherapy</i>		
	<i>Healthy control</i>	<i>Hormonal</i>	<i>Chemotherapy</i>	<i>Healthy control</i>	<i>Hormonal</i>	<i>Chemotherapy</i>
Trails A	25.3 ± 5.4	27.2 ± 9.1	27.6 ± 8.2	25.1 ± 5.4	26.3 ± 6.8	27.1 ± 9.4
Trails B	65.2 ± 15.8	70.2 ± 22.9	69.9 ± 25.8	59.8 ± 15.9	70.0 ± 21.8	67.1 ± 27.0
FAS, number of correct words	43.1 ± 13.5	37.6 ± 10.5	41.5 ± 12.4	44.4 ± 11.6	39.4 ± 10.8	40.3 ± 12.7
CVLT-II, Long Delay Free Recall	13.5 ± 2.1	12.0 ± 2.6	12.4 ± 3.0	14.5 ± 1.6	12.6 ± 2.5	13.1 ± 2.5
CVLT-II, Long Delay Yes-No	15.2 ± 1.2	15.0 ± 1.3	15.2 ± 1.1	15.5 ± 0.9	15.0 ± 1.3	15.4 ± 1.0
WAIS-III, Digit Symbol Coding	72.3 ± 11.3	68.9 ± 11.3	67.7 ± 12.5	75.6 ± 10.2	68.7 ± 11.9	68.2 ± 13.4
WAIS-III, Symbol Search	30.5 ± 5.9	30.3 ± 5.0	30.9 ± 6.6	32.1 ± 5.3	30.5 ± 4.9	31.3 ± 6.9
WAIS-III, Letter-Number Sequencing	10.4 ± 2.6	10.5 ± 2.2	10.8 ± 2.8	11.3 ± 2.5	11.0 ± 2.3	10.3 ± 2.6
CCC, total	49.6 ± 5.8	43.0 ± 7.3	43.4 ± 7.6	51.3 ± 5.1	44.7 ± 7.7	43.1 ± 8.7

Note. CVLT-II = California Verbal Learning Test II. CCC = Consonant Trigrams. WAIS-III = Wechsler Adult Intelligence Scale-III. FAS = Verbal Fluency.

Procedure

Two commonly used texts on clinical neuropsychological assessment were consulted in order to select the most appropriate tests from the 23 used in our original study for inclusion in the reduced battery: Lezak et al. (2004) *Neuropsychological Assessment 4th Edition* and Strauss et al. (2006) *A Compendium of Neuropsychological Tests, 3rd Edition* . An emphasis was placed on tests purported to measure the functions previously identified as impaired in this population (i.e. language, verbal memory, executive function, processing speed and working memory). In addition, since the objective was to evaluate a battery that could be clinically useful as a screening tool, administration time for individual tests was an important characteristic. Finally, because there are very few studies to date that have examined the sensitivity of individual neuropsychological tests to chemotherapy-related impairments in cognition (Jansen et al., 2007; Freeman & Broshek, 2002), tests were also considered for inclusion if they have been recommended for the evaluation of mild traumatic brain injury because the areas found to be most affected are similar to those noted in chemotherapy-related cognitive impairment (Lucas & Addeo, 2005). With these resources and considerations in mind, careful evaluation led to the elimination of 14 tests. Table 10 (see page 65) provides an explanation and/or rationale for the decision made regarding each test.

Table 10. Rational for test selection

<i>Test</i>	<i>Status</i>	<i>Rational</i>
Trails A & B	Retained	Shown to be highly sensitive to cognitive deficits in traumatic brain injury (TBI); minimal practice effects over longer intervals (1 year), quick administration
FAS, number of correct words	Retained	High reliability; sensitive to detection of TBI; good measure of verbal abilities; quick administration
CVLT-II, Long Delay Free Recall & Long Delay Yes/No Recognition	Retained	Solid internal consistency, good test-retest reliability
Symbol Search & Digit Symbol Coding	Retained	Low specificity on its own, but solid measure of processing speed when combined with digit symbol coding; shown to be sensitive to TBI severity; quick administration
Letter-Number Sequencing	Retained	Good reliability; sensitive to TBI severity; quick administration
CCC, total	Retained	Sensitive to detection of TBI
Spatial Span	Removed	Not validated as a measure of visual working memory
Logical Memory II	Removed	Not shown to be measuring a different construct than Logical Memory I; may not measure delayed memory
Family Pictures	Removed	Has not been shown to necessarily capture visual learning; highly influenced by verbal abilities
CVLT-II, Trial 1 Free Recall	Removed	More interested in delayed vs. immediate recall
PASAT 2.4s, total correct	Removed	Despite solid psychometrics, was very difficult for participants and was often discontinued; too much missing data as a result; long administration
RVLT, Trial 1 Free Recall RVLT, Long Delay Free Recall RVLT, Long Delay Recognition	Removed	More interested in verbal vs. visual memory as this has been shown to be more vulnerable to the effects of chemo-fog
Block Design	Removed	Good psychometrics but administration time is too long
Arithmetic	Removed	Performance has been shown to depend on math skills as much as Working Memory; test has been removed from WAIS-IV for this reason
Digit Span, Forward + Backward	Removed	Not found to be sensitive to degree of severity in TBI
WCST, no. of trials, raw score	Removed	Concerns regarding the aspect of “novelty” required for this test given the test/re-test nature of the study
Boston Naming Test, total	Removed	Very highly correlated with IQ, education, and age; does not differentiate well when scores are high
Grooved Pegboard, combined	Removed	too much missing data

Note. PASAT = Paced Auditory Serial Addition Task. WCST = Wisconsin Card Sorting Test. WAIS-III = Wechsler Adult Intelligence Scale-III. WMS-III = Wechsler Memory Scale-III. CVLT-II = California Verbal Learning Test II. RVLT = Rey Visual Learning Test. FAS = Controlled Oral Word Association Test. CCC = Consonant Trigrams

Sources: Lezak, 2004; Strauss, Sherman, & Spreen, 2006.

Data Analysis

Criterion cut-off scores were calculated using the resampling add-in for Excel (Resampling Stats for Excel, Version 4.0). All other analyses were conducted using the Predictive Analytics Software version 18.0 from SPSS Inc.

Standardized regression-based model. A standardized regression-based model (SRB; McSweeney et al., 1993) was used to evaluate decline in neuropsychological function in the chemotherapy and hormonal groups. This model provided adjustments for measurement error and practice effects and permitted the inclusion of several covariates including age, education, and estimated IQ.

Briefly, the SRB approach is used to identify participants who have experienced a significant change in performance across time. For each test, the post-treatment performance is regressed on the baseline and covariates in order to obtain a predicted test score. This value is then subtracted from the obtained score and divided by the standard error of estimate in order to derive a standardized change score. A score greater than ± 1.64 (falling within the 5% area at either end of a normal distribution) is indicative of a significant change in performance.

Following the SRB analyses, performance across the nine test variables was summed and participants classified according to the cut-off criterion of significant decline on two or more tests. Once the results were obtained, they were compared to those from study 1 to assess whether the reduced battery was equally able to identify the women who were previously found to be significantly impaired. Three comparisons were made: (1) the percentage of participants who declined on 0 to 9 tests were calculated, (2) the percentage of

participants who met the cut-off criterion for each study were calculated, and (3) the percentage of participants who met a more stringent cut-off criterion of significant decline on 4 or more tests in study 1, and 3 or more tests in study 2 were calculated.

Resampling. Once individual decline by test was established using SRB techniques, a cut-off criterion to signify an overall decline on the neuropsychological test battery from study 1 (Ouimet et al., 2009) was calculated. Rather than using an arbitrary cut-off, a criterion value was derived through the use of resampling with replacement. A data set was established such that the probability of showing a decline on any one test of the 23 in the battery, simply as a function of chance, was 2.5% (chosen to represent two standard deviations below the mean). Twenty-three samples were drawn to simulate results on the entire battery for one participant, a procedure that was repeated by the program 65,000 times. Frequencies were then calculated and based on the results, it was determined that the probability of showing significant decline on three or more tests out of 23, due to chance alone, was slightly less than 2%. Thus a cut-off value of three or more was considered to warrant further clinical investigation. Using the same approach with the reduced battery from study 2, a cut-off of decline on two or more tests was considered sufficient to warrant further clinical investigation.

These values were chosen as they are indicative of a pattern of decline that would only be seen in slightly less than 2% of the normal population. However, the actual rates of chemotherapy-related cognitive decline are difficult to ascertain; therefore, a more stringent cut-off for decline on four or more tests in study 1 and three or more tests in study 2 is also included for comparison. A pattern of this sort would be expected to occur less than 1% of the time in a normal population.

RESULTS

Descriptive statistics relating to age, education, estimated IQ at baseline, and test-retest interval are provided in Table 3 (see page 35). There were significant group differences between groups on education and estimated IQ. When these covariates were entered into the model first in the SRB analyses, followed by the post-treatment score, they accounted for a significant proportion of the variance in the Symbol Search ($R^2 = .29$, $F\Delta(3,24) = 3.20$, $p = .04$) subtest of the WAIS-III.

Figure 5 presents the breakdown of participants in each group who declined on each of 0 to 9 tests based on the reduced battery (left side) and the long battery (right side). This figure illustrates the variability in the obtained results as a function of the number of tests included in the neuropsychological battery.

Figure 6a shows the percentage of participants in each group who met the cut-off criterion of significant decline on three or more tests in study 1 and two or more tests in study 2. Approximately 39% of both the hormonal and chemotherapy group met criterion as compared to 7% of the healthy control in both studies (light gray column). Figure 6b shows the effect of using a more stringent cut-off criterion - significant decline on four or more tests in study 1 and three or more tests in study 2. Not surprisingly, the percentage of participants meeting this more stringent criterion decreases to 24% in the hormonal group and 20.5% in the chemotherapy group. None of the healthy controls are identified using this criterion.

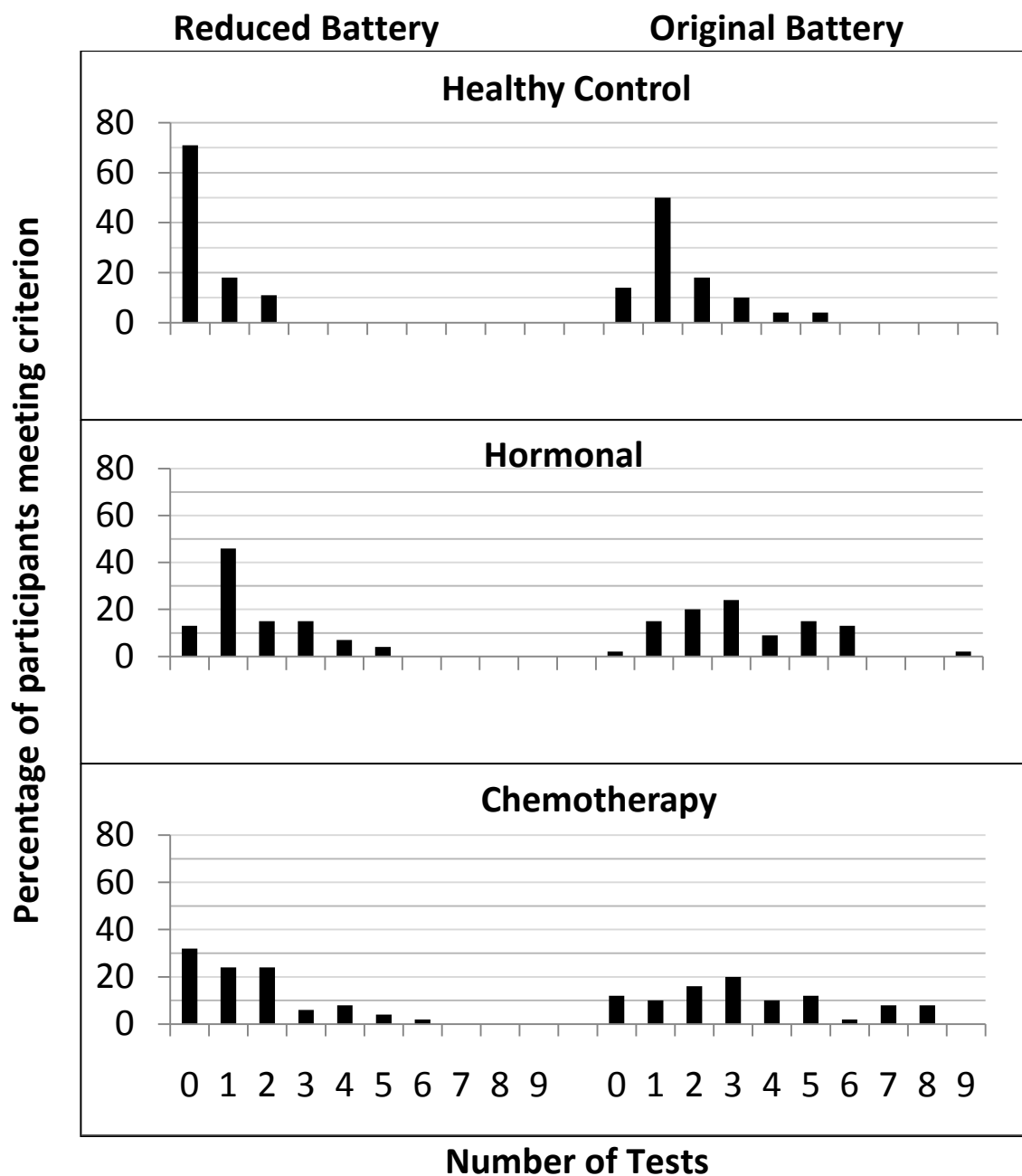


Figure 5. Percentage of participants meeting criterion for cognitive decline on 0 – 9 tests based on comparison with a healthy control group. The left side of each graph presents the results from the reduced battery (study 2) and the right side, the comprehensive battery (study 2). The three horizontal panels represent the results of the healthy control, hormonal, and chemotherapy groups respectively.

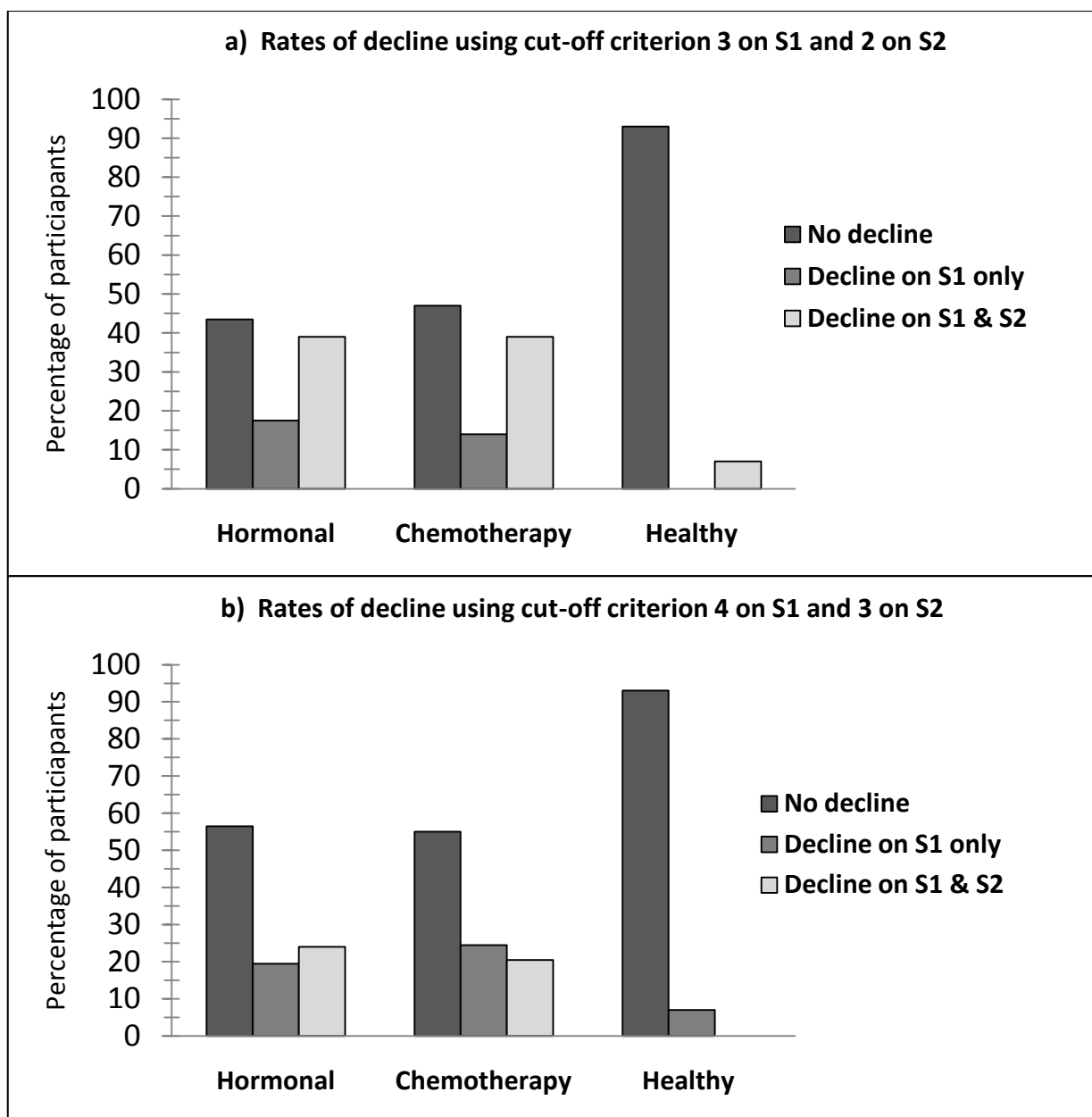


Figure 6. Percentage of participants who met the cut-off criterion of significant decline on three or more tests in study 1 (S1) and two or more tests in study 2 (S2) are presented in the upper panel (a). The lower panel (b) illustrates the percentage of participants who met a more stringent cut-off criterion of significant decline on four or more tests in S1 and three or more tests in S2.

Table 11 provides a comparison of individual participants' performance between our previous study and the current one.

Table 11. Comparison of individual decline from study 1 to study 2

Hormonal group					Chemotherapy group				
Participant number	# tests declined study 1	% tests declined study 1	# tests declined study 2	% tests declined study 2	Participant number	# tests declined study 1	% tests declined study 1	# tests declined study 2	% tests declined study 2
259	5	22	3	33	168	8	35	4	44
253	6	26	4	44	165	4	17	2	22
251	6	26	4	44	164	3	13	2	22
250	3	13	2	22	157	3	13	2	22
246	3	13	2	22	151	7	30	2	22
244	5	22	5	55	149	7	30	4	44
241	3	13	2	22	147	4	17	2	22
240	5	22	3	33	145	5	22	5	55
237	9	39	5	55	137	5	22	4	44
230	4	17	3	33	136	5	22	2	22
228	6	26	3	33	135	3	13	2	22
223	6	26	3	33	131	4	17	4	44
219	5	22	3	33	130	7	30	3	33
217	6	26	2	22	120	8	35	3	33
216	6	26	4	44	119	8	35	6	66
212	3	13	2	22	118	4	17	2	22
207	3	13	2	22	116	7	30	5	55
204	3	13	3	33	114	3	13	2	22
					102	6	26	3	33
260*	4	17	1	11	163*	5	22	1	11
256*	5	22	1	11	162*	4	17	1	11
254*	4	17	1	11	159*	4	17	1	11
248*	4	17	1	11	158*	8	35	1	11
242*	4	17	1	11	128*	5	22	0	0
236*	5	22	1	11	121*	5	22	1	11
221*	6	26	1	11	105*	5	22	1	11
208*	4	17	0	0					
203†	2	9	2	22	140†	2	9	2	22
					127†	2	9	2	22
					124†	2	9	2	22
Healthy control group									
423	5	22	2	22					
409	4	17	2	22					
401†	2	9	2	22					

Bold-Participants showing decline on 3 or more tests in study 1 and identified by the reduced battery in study 2.

*Participants showing decline on 3 or more tests in study 1 but not identified by the reduced battery in study 2.

†Participants showing decline on fewer than 3 tests in study 1 but identified by the reduced battery in study 2.

In the hormonal group, 56.5% of participants met the cut-off criteria of significant decline on three or more tests using the long battery. Of these, 39% were also identified by

the reduced screening battery, while 17.5% did not meet the cut-off criteria of decline on two or more tests on the same battery. One participant in this group did not meet the cut-off criterion for the larger battery, but was identified by the reduced battery.

Results were similar in the chemotherapy group; 53% of participants met the cut-off criterion on the large battery, of which 39% were identified by the screening battery and 14% did not meet criterion. Of the 47% of participants in the chemotherapy group who did not meet the cut-off criterion on the large battery, 6% (three participants) were still identified by the reduced battery.

In the healthy control group, only two participants (7%) showed significant decline on four or more tests in study 1, and they were also identified by the reduced battery in study 2. One additional participant who did not meet the long battery cut-off criterion was identified by the reduced battery.

DISCUSSION

The goal of this study was to compare the sensitivity of a small battery of neuropsychological tests to detect significant decline in cognitive function following chemotherapy treatment with that of a comprehensive battery. Comparisons were made using a healthy control group as a reference and two criteria levels. The results were similar in that comparable numbers of participants were identified by both batteries, suggesting that individuals vulnerable to chemo-fog can be identified as easily by a more selective battery emphasizing specific domains. This has great clinical utility.

Nevertheless, there were some participants (roughly 19%) who were identified by the comprehensive battery and not the reduced battery using both the relaxed and more stringent criterion. This could be a result of the cut-offs being too low although there were reasonable

statistical grounds for the selected values. Another possibility is that certain tests are more sensitive to age-related cognitive decline than the effects of chemotherapy.

There are certain limitations to our current study. Declines in function in this population may not be entirely attributable to chemotherapy. Various other explanations have been proposed, including disease-related stress, fatigue and depression. Although our study found no evidence that fatigue or depressive symptoms played a role in performance (i.e. non-significant as covariates) they should still be routinely evaluated.

Some researchers have remarked that baseline assessments are often undertaken at a time of great stress and may therefore not provide an accurate comparison point post-treatment. This is useful to bear in mind because if baseline scores are underestimated due to stress, then the degree of subsequent impairment will also be underestimated, and apparent improvement may be overestimated.

Adjuvant hormonal therapies may also lead to cognitive declines. In fact, although we are attempting to develop a screening tool that is sensitive to the cognitive deficits specifically associated with chemotherapy, the results suggest that the hormonal group meets screening criterion almost as frequently as the chemotherapy group, despite having received no chemotherapy treatment. One possible explanation for this is differing patterns of tests results such that chemotherapy patients may be declining on a distinct combination of tests; therefore the battery may need to be modified to reflect the declines more particular to chemotherapy treatment versus hormonal therapies. Current research also suggests that there are cognitive impairments associated with hormonal treatment (Bender et al., 2006; Castellon et al., 2004; van Dam et al., 1998) and it may be difficult to distinguish these effects from those resulting from chemotherapy alone.

In this study, our comparison was a healthy control group as opposed to a hormonal or another treatment group. It has been demonstrated that the nature of the control group can influence the quality of analysis (Sawrie, 2002) and that careful matching ensures that pre-existing factors do not overly influence results. Based on this observation, it could be argued that our healthy control group does not share the characteristics associated with having cancer. However, given the aim of the study, underestimation of decline and misclassification were considered to be of primary importance and the healthy control group was considered the best comparison group for decline in function from a healthy state.

Lastly, there was a statistically significant difference in education and estimated IQ between the healthy control group and the two treatment groups, suggesting the healthy control group may have more cognitive reserve; this could account for some of the differences observed between groups. However, it should be noted that based on standardized IQ scores (as measured by WAIS instruments), all participants fell within +/- one standard deviation from the mean (i.e. 100 +/- 15). With respect to education, approximately two thirds of the hormonal and chemotherapy participants completed post-secondary studies, as compared to slightly over half of the healthy control participants.

Despite these significant differences at a group level, when the covariates were evaluated at an individual level using the SRB analyses, they only accounted for a significant proportion of the variance for one of the nine variables, suggesting that they are not exerting a great influence on the majority of decline scores.

Although several of the limitations in neuropsychological assessment described in the introduction cannot be readily addressed in this paper - lack of clean constructs, fixed versus flexible batteries, limited theory - there are some that can be effectively minimized.

These include: redundancy & misclassification and issues with poor psychometrics, which can be dealt with through the use of a reduced battery and careful selection of tests based on available empirical resources and research. Professional guidelines or policies would also be very beneficial to instruct clinicians on the most effective and reliable means of constructing a neuropsychological test battery, and would reduce test selection based solely on availability, tradition, familiarity, or personal preference.

Finally, studies evaluating subjective measures of cognitive impairment in recipients of chemotherapy and hormonal therapy are still very limited. This type of research will be important in determining which aspects of cognitive functioning have the most negative impact on quality of life for the breast cancer survivor. Further work is also still needed to clarify which neuropsychological tests are most sensitive to detecting cognitive decline associated with chemotherapy treatment so that assessment batteries can be limited to fewer tests. Preliminary research in this area has identified a limited number of tests that appear to be sensitive to declines in this population (Jansen et al., 2007; Freeman & Broshek, 2002) but more research of this nature must be done in order to improve clinical efficiency and reduce the likelihood of misclassification.

STUDY 3

Submitted to the Journal of Clinical and Experimental Neuropsychology: Ouimet, L. A., Stewart, A., Collins, B., Schindler, D., & Bielajew, C. Neuropsychological assessment of chemo-fog in breast cancer: A comparison of conceptual versus statistical cognitive domains using factor analytic techniques

Neuropsychological assessment of chemo-fog in breast cancer: A comparison of conceptual versus statistical cognitive domains using factor analytic techniques

Lea Ann Ouimet, Angela Stewart¹, Barbara Collins²,

Dwayne Schindler, Catherine Bielajew

School of Psychology, University of Ottawa, ON, Canada.

¹Royal Ottawa Health Care Group, Ottawa, ON, Canada

²Ottawa Hospital, Civic Campus, Ottawa, ON, Canada.

ABSTRACT

There is a great deal of variability in the composition of neuropsychological test batteries used in the assessment of chemo-fog in breast cancer patients; tests vary in length, complexity, and what they purport to measure, and many have not established reliability and validity indices. These differences may contribute to creating divergence between theoretical cognitive domains (i.e. group of tests chosen to measure a specific cognitive function such as language, attention, memory, etc.) and statistically defined ones. Our objective was to compare the conceptual domains in a neuropsychological test battery with those identified through factor analysis, an issue that has not been addressed in the breast cancer literature. This was accomplished using data obtained from our previous studies investigating the neuropsychological effects of adjuvant chemotherapy. To determine whether there was statistical support for the conceptual framework underlying the composition of the domains, factor analyses (principal axis factoring) were performed using the baseline and post-treatment results of 22 neuropsychological tests. As factors were expected to be inter-related to some extent, a non-orthogonal (oblimin) rotation was employed to enhance interpretability of the factors. For baseline data, the Kaiser-Meyer-Olkin was .82 and Bartlett's $\chi^2(231, N = 123) = 789.97, p < .001$. A six factor solution explained 53% of the common variance. For the post-treatment data, the Kaiser-Meyer-Olkin was also .80 and Bartlett's $\chi^2(231, N = 123) = 755.23, p < .001$ and a six factor solution explained 56% of the common variance. Two of the conceptual domains showed a high degree of consistency with the statistically defined factors at both baseline and post-treatment. The remaining seven conceptual domains differed in varying degrees, but some overlap was evident in all domains, with the baseline data showing greater convergence than data obtained post-treatment.

INTRODUCTION

Breast Cancer strikes approximately 1 in 9 women each year and the majority will receive some form of chemotherapy as part of their treatment regimen. However, a proliferation of literature on survivors of breast cancer has demonstrated that a subset of these women suffer from “chemo-fog”, a decline in cognitive function associated with chemotherapy (Ahles et al., 2002; Ahles et al., 1996; Bender et al., 2006; Brezden et al., 2000; Castellon et al., 2004; Donovan et al., 2005; Gottschalk et al., 2003; Jenkins et al., 2006; Kreukels et al., 2005; Mar Fan et al., 2005; Ouimet et al., 2009; Saykin et al., 2003; Schagen et al., 2002a, 2002b, 2001, 1999; Scherwath et al., 2006; Servaes et al., 2002; Shilling et al., 2003, 2005; Stewart et al., 2006; Tchen et al., 2003; Wefel et al., 2004b; van Dam et al., 1998; Weis et al., 2009).

Meta-analytic reviews of the literature have revealed declines in the areas of executive function and verbal memory (Anderson-Hanley et al., 2003) and language, short-term memory, and spatial abilities Stewart et al. (2006). However, despite increasing evidence to support the assertion that chemotherapy can result in impaired cognitive functioning, various neuropsychological assessment issues complicate the identification of chemo-fog. These issues include but are not limited to: the composition of neuropsychological test batteries and test selection, the number of tests administered, and the overlap of cognitive functions.

The composition of test batteries is quite varied in chemo-fog research (Freeman and Broshek, 2002; McQuellon, comment in Olin, 2001). This lack of consistency across studies leads to great variability in the choice of tests used to represent the various cognitive domains, making comparison of results onerous. Moreover, although a substantive literature

on test theory and guidelines for the selection of individual tests is available (Hebben & Milberg, 2009; Lezak et al., 2004; Mitrushina, Boone, Razani & D'Elia, 2005; Strauss et al., 2006), there is a lack of theory specifically addressing the optimal composition of a neuropsychological battery or the choice of tests by cognitive domain. This creates difficulty in accurately identifying the areas of functioning that are most compromised subsequent to chemotherapy treatment.

The number of tests administered also impacts the results obtained, both directly and indirectly, with longer assessments being more problematic. In the area of breast cancer and chemo-fog research, the batteries typically employed comprise a large number of tests. Clinically, a larger battery increases administration time, thereby limiting utility. Also, as pointed out by Hurria & Lachs (2007) and Vardy et al. (2007), participant fatigue may result after two to three hours of testing, introducing potentially conflicting explanations for poor test performance.

Long batteries may also lead to a loss of attention and focus, decreased motivation, etc., and makes clear interpretation of the results more problematic. However, perhaps the most critical disadvantage is the risk of misclassification, a risk that increases with each additional test. In a worst case scenario, misclassification could lead to recommendations that are less than optimal or inappropriate, or a lack of intervention altogether.

From a research perspective, each added test reduces power and the probability of obtaining statistically significant results. Additional tests also increase the likelihood of redundancy, multi-collinearity, and Type II errors. Practically, a longer battery may decrease participation overall, and raise attrition in longitudinal studies.

Results from a preliminary study evaluating the ability of a reduced battery to detect chemo-fog (Ouimet, Stewart, Collins, Schindler & Bielajew, 2010) suggest that a comprehensive battery may not be necessary. A standardized regression based approach compared a battery including 23 tests and a more selective battery of 9 subtests and found the reduced battery was consistently able to identify cognitive impairments in chemotherapy recipients. This suggests that individuals vulnerable to chemo-fog can be identified by a more selective battery.

Compounding all of the issues previously noted, is the complexity and interrelatedness of cognitive functions. The way in which they overlap makes test selection challenging when the goal is to measure a specific construct since all neuropsychological tests capture multiple facets of cognitive functioning.

Although these issues arise to a greater or lesser extent irrespective of the presenting problem, they become more salient in areas where the research findings are inconsistent and where the approach to neuropsychological assessment is highly variable across studies (i.e. differences in test selection and composition of cognitive domains) as is the case in the breast cancer/chemo-fog research. These obstacles in assessment can eventually be manifested in studies where the subjective reports from survivors experiencing chemo-fog do not coincide with objective measures of cognitive function.

For example, a 2006 study by Downie et al. measured self-reported quality of life in a group of women receiving adjuvant chemotherapy for breast cancer and compared the results with a neuropsychological measure designed to detect cognitive impairment, the High Sensitivity Cognitive Screen (HSCS; Faust & Fogel, 1989, Fogel, 1991). Their goal was to

assess whether subjective complaints of cognitive impairment corresponded with objective neuropsychological test results.

The authors found discrepancies between the self-reports and the cognitive assessment, particularly in the area of attention and concentration, in that 90% of the patients reported difficulties but only 10% were categorized as abnormal on objective measures. Smaller inconsistencies were also found in the areas of language and memory (78% and 95% self-reported vs. 61% and 48% identified). They concluded that the HSCS may not be sensitive enough to capture the subtle cognitive declines experienced by this population. However, given the issues related to neuropsychological testing, it is equally possible that there is some divergence between theoretical cognitive domains and statistical ones, and that this could account for the lack of consistency between self-reports and objective measures of cognitive impairment. Therefore, the objective of this paper was to explore this further through the use of factor analysis.

Model

Using data obtained from our previous studies on the neuropsychological effects of adjuvant chemotherapy in breast cancer patients (Ouimet et al. 2009; Stewart et al. 2006; 2008a), factor analytic techniques were used to evaluate the conceptual framework underlying the composition of the cognitive domains. The goal was to determine whether there was statistical support for the conceptual grouping of these particular tests by domain.

METHOD

Participants

Participants were recruited for a larger longitudinal study conducted out of the Ottawa Regional Cancer Centre investigating the neuropsychological effects of adjuvant

chemotherapy. They consisted of 95 stage I or II breast cancer patients aged 50 to 66 ($M = 57.89$, $SD = 4.13$), who had undergone mastectomy or lumpectomy and who were receiving chemotherapy (with or without hormonal treatment, $n = 49$) or hormonal treatment alone ($n = 46$). A healthy control group ($n = 28$) was also included as a comparison group to examine and evaluate practice effects within the treatment group. The healthy control sample was recruited through advertisements posted on bulletin boards at the Ottawa Hospital. All participants received \$50 for each test session they completed.

Sample Selection

The study was limited to postmenopausal women aged 65 years or younger (generally > 50 years) as a means of minimizing the potential confounding effects of hormonal status and age-related cognitive decline. Fluency in English was required in order to complete the test battery. Exclusion criteria for all groups included a history of previous cancer and chemotherapy or radiation treatment. Participants presenting with serious psychiatric disorder (e.g. major depression, schizophrenia), neurological illness, or significant substance abuse were excluded due to the potential negative effects on cognition. This study was approved by the board of ethics at the Ottawa Hospital and written informed consent was obtained from all participants.

Measures

A battery consisting of 23 neuropsychological tests with strong psychometric properties was administered to all participants at baseline and at a six month follow-up. The tests were chosen to represent the major cognitive domains including executive function, language function, motor skills, processing speed, verbal learning and memory, visual

learning and memory, visuospatial function, and working memory. A list of the tests by domain is provided in Table 2 (see page 31).

Of the original 23 tests, the arithmetic subtest of the Wechsler Adult Intelligence Scale-III (WAIS-III) was not included in the analyses because it had many significant correlations with other variables in the correlation matrix, making it redundant when considered in combination with other components.

Procedure

After obtaining informed consent, demographic information and past medical history were collected. Next, the Quick Test (Ammons & Ammons, 1963) was administered to estimate IQ, followed by the neuropsychological test battery. All tests were administered at baseline and following the last chemotherapy cycle for participants in the chemotherapy group. The hormonal and healthy control participants completed the test sessions at time points comparable to the chemotherapy group, roughly six months after diagnosis. All tests were administered by experienced individuals. Sessions lasted an average of three hours, and were conducted at either the participant's home or the hospital, according to her preference.

Data Analysis

Analyses were conducted using the Predictive Analytic Software (PASW) statistical package (version 18.0; SPSS Inc., 2009). To determine whether there was statistical support for the conceptual framework underlying the composition of the domains, factor analyses (principal axis factoring) were performed on the 22 neuropsychological test results at baseline and post-treatment. A non-orthogonal (oblimin) rotation was chosen to enhance interpretability due to the expected inter-relatedness of the factors.

RESULTS

Descriptive statistics relating to age, education, estimated IQ at baseline, and test-retest interval can be found in Table 3 (see page 35). There were significant differences between the healthy control group and the two treatment groups on education and estimated IQ, with the healthy control having higher levels of both education and estimated IQ.

Factor Analysis

In evaluation of the baseline data using factor analytical techniques, the Kaiser-Meyer-Olkin measure of sampling adequacy was .82 and would be considered in current terms to be meritorious (Kaiser & Rice, 1974). The Bartlett's test of sphericity was significant (approx. $\chi^2(231, N = 123) = 789.97, p < .001$). A six factor solution was found that accounted for 53% of the common variance. Individually, the first factor accounted for 29% of the variance, the second factor for 8%, the third factor for 5%, the fourth factor for 4%, the fifth factor for 4%, and the sixth factor for 3%.

Evaluation of the post-treatment data also resulted in a Kaiser-Meyer-Olkin measure of sampling adequacy of .80 and a Bartlett's test of sphericity that was significant (approx. $\chi^2(231, N = 123) = 755.23, p < .001$). In this case a six factor solution was found that explained 56% of the common variance. Individually, the first factor accounted for 33% of the variance, the second factor for 7%, the third factor for 5%, the fourth factor for 5%, the fifth factor for 3%, and the sixth factor for 3% of the common variance. Details of the variables found for each factor at baseline and post-treatment are included in Table 12.

Table 12. Statistical factors at baseline and post-treatment

Statistical factors at baseline	Statistical factors at post-treatment
<i>Factor 1</i> Symbol Search Block Design Grooved Pegboard Spatial Span PASAT WCST	<i>Factor 1</i> RVLT, Trial 1, Free Recall RVLT, Long-Delay Recognition RVLT, Long-Delay Free Recall Boston Naming Test
<i>Factor 2</i> Trails A Trails B Digit Symbol Coding	<i>Factor 2</i> Digit Symbol Coding Trails A Trails B Grooved Pegboard WCST Family Pictures II
<i>Factor 3</i> CVLT-II, Long-Delay Free Recall CVLT-II, Long-Delay Recognition CVLT-II, Trail 1, Free Recall Logical Memory II	<i>Factor 3</i> Digit Span Letter-Number Sequencing
<i>Factor 4</i> RVLT, Long-Delay Free Recall RVLT, Trial 1, Free Recall RVLT, Long-Delay Recognition Family Pictures II	<i>Factor 4</i> CVLT-II, Long-Delay Recognition CVLT-II, Long-Delay Free Recall Logical Memory II CVLT-II, Trail 1, Free Recall
<i>Factor 5</i> Letter-Number Sequencing Digit Span	<i>Factor 5</i> FAS Consonant Trigrams PASAT
<i>Factor 6</i> FAS Boston Naming Test Consonant Trigrams	<i>Factor 6</i> Symbol Search Spatial Span Block Design

The overlap between the conceptual and statistical domains is illustrated in Table 13.

Table 13. Comparison between conceptual and statistical factors at baseline and post-treatment

Conceptual factors	Statistical factors at baseline	Statistical factors at post-treatment
Visual learning and memory RVLT Free Recall, Trial 1 RVLT Long-Delay Free Recall RVLT Long-Delay Recognition Family Pictures II	Factor 4 RVLT, Long-Delay Free Recall RVLT, Trial 1, Free Recall RVLT, Long-Delay Recognition Family Pictures II	Factor 1 RVLT, Trial 1, Free Recall RVLT, Long-Delay Recognition RVLT, Long-Delay Free Recall <i>Boston Naming Test</i>
Processing speed Digit Symbol Coding Symbol Search Trails A	Factor 2 Trails A <i>Trails B</i> Digit Symbol Coding	Factor 6 Symbol Search <i>Spatial Span</i> <i>Block Design</i>
Verbal learning and memory CVLT-II, List A, Trial 1 CVLT-II, Long-Delay Free Recall CVLT-II, Long-Delay Recognition Logical Memory II	Factor 3 CVLT-II, Long-Delay Free Recall CVLT-II, Long-Delay Recognition CVLT-II, Trial 1, Free Recall Logical Memory II	Factor 4 CVLT-II, Long-Delay Recognition CVLT-II, Long-Delay Free Recall Logical Memory II CVLT-II, Trial 1, Free Recall
Working memory Consonant Trigrams Digit Span Letter-Number Sequencing Spatial Span Arithmetic*	Factor 5 Letter-Number Sequencing Digit Span	Factor 3 Digit Span Letter-Number Sequencing
Language function Boston Naming Test FAS	Factor 6 FAS Boston Naming Test <i>Consonant Trigrams</i>	Factor 5 FAS <i>Consonant Trigrams</i> PASAT
Executive function PASAT Trails B WCST Motor Grooved Pegboard Visuospatial function Block Design	Factor 1 <i>Symbol Search</i> Block Design Grooved Pegboard <i>Spatial Span</i> PASAT WCST	Factor 2 <i>Digit Symbol Coding</i> <i>Symbol Search</i> <i>Trails A</i> Trails B Grooved Pegboard WCST <i>Family Pictures II</i>

As can be seen, although there are eight conceptual factors, only six were identified in the statistical models. The conceptual domain “verbal learning and memory” was identical

to one of the statistically defined factors at both baseline and post-treatment and “visual learning and memory” matched completely at baseline and retained three of the four subtest variables at post-treatment. All other conceptual domains differed from the statistical ones to varying degrees; however some agreement was evident for all domains.

Overall, the statistical factors identified from the baseline data were more consistent with the conceptual factors than those found with the post-treatment data. The greatest discrepancies were in the domains for executive, motor and visuo-spatial functioning.

DISCUSSION

The main objective of this paper was to evaluate the framework underlying the conceptual domains in a neuropsychological test battery to determine if they were consistent with statistical ones derived through exploratory factor analysis.

The results indicate that there is a strong statistical basis for the inclusion of particular groupings of subtests into the conceptual domains. Overall, the results for the baseline data were found to be a slightly better match than the post-treatment data. This could be a result of the decline in test scores over time observed in participants in the treatment groups and reported in our previous studies (Ouimet et al., 2009; 2010), but is most likely due to the interrelatedness of cognitive functions described in the introduction.

For example, the conceptual domain “processing speed” comprises the subtests Digit Symbol Coding, Symbol Search and Trails A, whereas the most comparable statistical factor in the baseline data includes Digit Symbol Coding, Trails A and Trails B. However, it is not surprising that these subtests were grouped together under one statistical factor because speed is a crucial aspect of performance for all of these tests. Similarly, the most comparable

statistical factor for the post-treatment data includes Symbol Search, Spatial Span and Block Design, all of which utilize visual stimuli.

At a more basic level, some cognitive functions can be expected to affect all other domains. For instance, compromised attention could potentially reduce scores on tests measuring memory, processing speed, language, etc. As a result of this complexity in cognitive functioning, there are few absolutes when it comes to organizing neuropsychological tests into domains. Different researchers or clinicians may include the same subtests under varying domains depending on their perception of what aspect of cognitive functioning is recruited most by a given task. If researchers were able to arrive at a consensus regarding test selection, comparison of results would be simplified and more appropriate interventions could be developed more readily.

In this study the sample size was 123. Although there are no agreed upon criteria for appropriate sample sizes when employing factor analysis, generally speaking, smaller n 's are more subject to sampling error and therefore may produce results that do not generalize to the population under investigation (Costello & Osborne, 2005).

Also noted in our previous studies (Ouimet et al., 2009; 2010) were the group differences in education and IQ, with the treatment groups being more highly educated and scoring higher on estimated pre-morbid IQ. Although the differences between the treatment groups and the healthy control group was significant for both variables, all participants fell within +/- one standard deviation from the mean based on standardized IQ scores (i.e. 100 +/- 15 as measured by WAIS instruments).

Finally, although some degree of overlap was observed amongst all of the statistical versus conceptual domains, it is possible that the results would have been improved with a

smaller battery. As previously noted, a preliminary study by our group (Ouimet et al., 2010) found that a reduced battery can be equally sensitive in detecting cognitive impairments associated with chemo-fog. One way to achieve this, while at the same time limiting the risk of misclassification, is to carefully select tests by examining correlations to ensure that tests within a given domain have low inter-correlations but correlate strongly with the criterion ability (Vanderploeg & Schnika, 2004). Future studies will be critical for establishing which neuropsychological tests are most sensitive in detecting and screening for chemo-fog.

Further research is also needed to establish what type of cognitive decline has the most negative impact on QOL for breast cancer survivors. Although self reports consistently indicate that the cognitive impairments experienced by chemo-fog sufferers negatively affect QOL (Downie et al., 2006; Kayl et al., 2006; Mehnert et al., 2007; Reid-Arndt et al., 2009; Wies et al., 2009), at least one study comparing subjective and objective measures found significantly more participants reported difficulties than were captured by the neuropsychological tests (Downie et al., 2006).

Appropriate interventions aimed at improving QOL can only be developed once the exact nature of the deficits is clarified, and sensitive screening tools are available to identify those affected. Poignant examples in the literature of the daily impact of chemo-fog on QOL (Hafner, 2009; Mulrooney, 2008) serve as a reminder of the importance of continued research in this area.

GENERAL DISCUSSION

The purpose of the studies undertaken in this thesis was to examine chemotherapy-related cognitive impairments using a quantitative approach. This was accomplished by conducting a review of the methodological issues in this research area, with an emphasis on statistical analysis. This was followed by a comparison of sensitivity between a comprehensive neuropsychological test battery and a reduced one. Finally, the convergence of the theoretical and statistical neuropsychological domains was assessed using factor analytic techniques.

This discussion is divided into three sections. The first section summarizes the findings from each study. The second section consolidates the conclusions and presents a general discussion on the quantitative aspects of research with this population and related issues in neuropsychological assessment. The third section examines the limitations of the three studies included in this thesis, and highlights areas that would benefit from further research.

Study 1

In Study 1, we conducted a thorough review of the methodological issues in the breast cancer literature, with a particular focus on statistical methods, and discussed how these issues impact results and comparison across studies.

It has been established through meta-analysis (Falleti, Sanfilippo, Maruff, Weih, Phillips, 2005; Stewart et al., 2006) that the cognitive deficits in this population are subtle in nature and may be washed out by the use of group analyses. We demonstrated this effect by running group and individual analyses on our data and comparing the results. Although

negligible group differences were observed using univariate and multivariate approaches, individual based analyses revealed a subset of participants who demonstrated significant declines in function across numerous tests.

Furthermore, use of a standardized-regression based approach was preferable to the reliable-change index because it accounts for individual practice effects, regression to the mean, but most important, permits the inclusion of covariates such as age, education, measures of depression, fatigue, etc., so that the impact of these potential confounds can be assessed (Bost, Wen, Basso, & Cates, 2008; Sawrie et al., 1999).

Study 2

The neuropsychological test batteries used in the area of breast cancer research typically consist of a large number of tests. This not only reduces the likelihood of detecting meaningful results, but limits clinical utility because the batteries are too big to be of practical use for evaluating cognitive impairment in clinical settings.

The second study of this thesis compared the sensitivity of a comprehensive neuropsychological test battery with a reduced battery composed of a subset of these tests. The goal was to determine if the participants identified as being significantly impaired using 23 measures would also be identified through the use of only nine measures.

It was found that a comparable number of participants were identified overall using the more selective battery (41% of chemotherapy group; 39% of hormonal group) versus the larger battery (53% of chemotherapy group; 56.5% of hormonal group). When performance was examined by individual participants, 19 out of 26 (73%) chemotherapy participants were

identified by both the large battery and the smaller one, as were 18 out of 26 (69%) of hormonal participants.

Based on these results, it was concluded that a small battery of neuropsychological tests could be equally sensitive in detecting the cognitive effects of chemotherapy, provided adequate attention is given to test selection. However, some participants were not identified by the smaller battery, and further work is needed to clarify which individual tests are most sensitive to declines in this population.

Study 3

The third study of this thesis compared the theoretical cognitive domains of the neuropsychological test battery used in the previous two studies, with the statistical groupings identified through factor analysis. The aim was to determine whether there was statistical support for this particular division of tests by domain.

Of the nine theoretical domains, two (visual learning and memory, verbal learning and memory) showed a high degree of overlap with statistically derived factors. The remaining seven domains were more variable, with baseline data demonstrating better convergence. Overall, all theoretical domains showed some degree of overlap.

It was concluded that due to the overlapping nature of cognitive functions, the choice of which test to include in a particular domain is largely a subjective one that is based on different perceptions of how any given task is broken down into its component parts. Given this, it was recommended that some consensus among researchers regarding battery composition by domain would simplify comparison of results.

Some of the most recent papers in this area maintain there are no adverse effects of chemotherapy on cognitive function (Debess et al., 2010; Jim et al., 2009; Tager et al., 2010) despite the fact that impaired cognitive function is the most commonly reported symptom affecting QOL in cancer survivors treated with chemotherapy (Tallibert et al., 2007). However, studies continue to have methodological limitations that may affect the results and the nature of the conclusions drawn from them. These include group versus individual analyses (Tager et al., 2010), cross-sectional designs (Jim et al., 2009), definition of impairment (Debess et al., 2010; Jim et al., 2009) and issues with test selection (Debess et al., 2010).

Results from our first study illustrate the importance of prospective, longitudinal designs and data analysis at the individual, as opposed to group level, in this population. It was demonstrated that use of group analyses serves to obscure significant declines in individual participants. In contrast, when individual analyses are employed to assess serial test results, it becomes possible to distinguish impaired from non-impaired women by examining the amount of change from baseline to post-treatment in each participant's performance across all tests in the battery.

The type of analysis chosen is closely linked with how impairment is defined, and this also plays an important role in the interpretation of results. A fairly conservative definition of cognitive decline based on differences between baseline and post-treatment scores was used in study 1 and 2, because although not all domains are negatively affected in chemo-fog, those that are tend to fall in the significantly impaired range. This is consistent with other studies analyzing changes in performance at the individual level (Collins et al., 2008; Hurria et al., 2006; Jenkins et al., 2006; Stewart et al., 2008; Wefel et al., 2004).

In contrast, Jim et al. (2009) defined impairment as a decline based on comparisons with normative data rather than baseline assessment. Although this approach does examine data at the individual level, it is problematic because it is important for researchers and clinicians not to conclude that no cognitive impairment is present simply because an examinees scores fall in the normal range.

In fact, high functioning participants may obtain scores that appear average, but still indicate a significant decline based on their pre-morbid level of functioning and these declines may in turn represent a significant limitation in their ability to resume their previous lifestyle. Given this, evaluation of individual decline based on baseline data collected prior to treatment is the preferred approach.

However, it should be noted that baseline assessments are often undertaken at a time of great stress and may therefore not provide an accurate comparison point post-treatment. If it is expected that baseline scores will be underestimated due to stress or other factors, then the degree of subsequent impairment will also be underestimated. Also, as previously noted (Collins et al., 2008; Tager et al., 2010), attenuation of practice effects may actually represent a decline in cognitive function.

The study by Debess et al., (2010) also examined individual change between baseline and post-treatment. However, their battery consisted of only four tests, with five derived scores, purported to measure declines in performance across six domains (memory, attention, processing speed, mental flexibility, visuospatial functioning, and executive function). They defined impairment as difference scores below the 5th percentile on two of the five scores. There is no explanation as to how this cut-off criterion was chosen, but study 2 illustrated

how a more stringent cut-off increases the potential misclassification of participants who are significantly impaired.

Of greater concern is the small number of tests used to assess performance across so many domains. It has been suggested that more than one measure be administered for each domain because neuropsychological functions are so interrelated, and one test may tap into a slightly different configuration of cognitive demands than another test does (Strauss et al., 2000).

For example, in study 3 it was noted that the conceptual domain “processing speed” was composed of the Digit Symbol Coding, Symbol Search, and Trails A subtests. The statistical factor that was most comparable at baseline included the Digit Symbol Coding, Trails A and Trails B subtests. All of these tests measure speed, and this is a primary factor in performance on these tests. However, attention and visual scanning are also key components of performance, and all of these facets are recruited to a greater or lesser extent in each task.

Similarly, the most comparable statistical factor from the post-treatment data was composed of Symbol Search, Spatial Span, and Block Design. Again, all of these tests recruit visuospatial functions since they all use visual stimuli, but attention is crucial to performance, particularly for Symbol Search and Spatial Span, as is processing speed in Symbol Search and Block Design.

Additional considerations for test selection include a thorough understanding of the strengths and weaknesses of neuropsychological tests in order to make appropriate selections, with an emphasis on psychometrics and the adequacy of the normative sample (i.e. how closely does it match that of the examinee) (Brooks, 2009).

Use of the most up-to-date versions of revised tests should only be undertaken when it has been established that new revisions improve clinical reliability and validity of a test and the revised version is appropriate for the population being assessed (Bush, 2010; Silverstein & Nelson, 2000; Strauss et al., 2000). Attention must also be given to the number of tests selected, with consideration of the statistical impact of each additional test added.

Even after careful deliberation has been given to all of these factors, it is also important to remember that some low scores are to be expected, even in a healthy population, when a battery of tests has been administered. This is a function of a number of factors that affect performance including demographic variables, the number of tests completed, and the correlation between tests (Brooks, 2009), and shouldn't be interpreted as evidence for an impairment. However, a pattern of several low scores in a particular domain is cause for further consideration and investigation.

Some researchers have noted this apparent lack of agreement between subjective and objective reports of cognitive deficits (with subjective complaints typically higher than objective data) and have suggested that this is the result of survivors experiencing cognitive effects that are disease related (i.e. fatigue, stress, depressive symptoms) as opposed to treatment related, and that there are in fact no systematic effects of chemotherapy on cognitive function (Debess et al., 2010; Castellon et al., 2005; Hemelink et al., 2007; Jim et al., 2009; Schagen et al., 1999; Servaes et al., 2002; Shilling & Jenkins, 2006; Tager et al., 2010; Tchen et al., 2003).

As previously noted, fatigue and depressive symptoms were assessed in our studies using the POMS, and were included as covariates in our analyses. It was determined that they did not play a significant role in participant performance. However, the findings on the

relationship between fatigue, depression, and cognitive decline following chemotherapy for breast cancer are inconsistent, and it is preferable that these factors be evaluated routinely to ensure they are not overly influencing measured declines in function.

This discrepancy between self-reported versus objectively measured cognitive decline is not unique to this population. For example, in studies of multiple sclerosis, where early treatment-related cognitive deficits may also be very subtle in nature, concordance between subjective and objective functioning has also been found to be poor (Benedict et al., 2003) and one study found that reports from caregiver informants are more reliable indicators of actual functionality (Goverover, Chiaravalloti & DeLuca, 2005).

Studies of patients receiving coronary artery bypass grafts for coronary artery disease also report differences in subjective versus objective cognitive impairment. Although subjective complaints are commonly reported, objective measures do not always find corroborating evidence (Phillips-Bute et al., 2006; Selnes et al., 2004). However, Selnes et al. (2006) did find that coronary artery bypass graft patients were significantly more likely to report memory issues than a non-surgery control group with coronary artery disease at three and 12 month follow-ups. They conclude that deficits are likely subtle and difficult to capture with traditional neuropsychological tests.

Interestingly, a longitudinal study by Weis, Poppelreuter & Bartsch (2008) found nine out of 19 participants met objective criteria for significant cognitive decline roughly nine months post-treatment, but reported no subjective symptoms. This suggests that adverse effects of chemotherapy may be present, but may not impact functionality, possibly due to support systems, level of re-integration, compensatory strategies, or coping mechanisms.

In contrast, a large neuroimaging study by Stewart et al. (2008b) examining subjective memory deficits in a community population found that subjective deficits were associated with an increase in temporal and hippocampal volume irrespective of objectively measured deficits. The authors suggest that some participants may be aware of cognitive changes that are not yet measurable via neuropsychological testing. Together, these studies imply that subjective reports of cognitive decline are variable, and should not be taken as evidence for or against actual neuropsychological impairments without appropriate evaluation. As with any population, malingering must also be considered as a potential confound; however there is no evidence to suggest that this is an issue in breast cancer survivors experiencing chemo-fog.

Other factors have also been posited as contributing to the impairments in chemo-fog, including the effects of surgery and anesthesia. Although there are known short-term physical effects, including nausea and vomiting, pain, and fatigue (Montgomery & Bovbjerg; 2004), there are no studies of possible long-term physical effects in this population.

On the other hand, cognitive side effects of surgery and anesthesia known as post-operative cognitive dysfunction are common immediately after surgery, but typically resolve quickly (Hanning, 2005; Monk et al., 2007; Neubauer & Golden, 2005). Some lasting effects may be seen for up to three months or longer post-surgery, but are very rare, with the exception of patients who have undergone cardiac surgery (Hanning, 2005) or the elderly (Ancelin et al., 2001; Culley, Xie, & Crosby, 2007).

There may also be difficulties distinguishing the effects of adjuvant hormonal therapies from those of chemotherapy treatment. For example, a similar number of participants in the chemotherapy and hormonal groups were identified in study 1 and study 2

as being significantly impaired. However, although the number of participants was comparable, when individual test patterns were examined, it was noted that participants in the chemotherapy group were less likely to be significantly impaired on Trails B than those in the hormonal group, and more likely to be significantly impaired on Consonant Trigrams, Letter-Number Sequencing, and Digit Symbol Coding.

Although Trails B, Consonant Trigrams, and Letter-Number Sequencing all measure attention and working memory, Trails B also has an emphasis on set-shifting abilities that the other two measures do not have. Therefore, including tests that incorporate aspects of attention and working memory without a set-shifting component may help to identify those experiencing chemotherapy-related cognitive changes. Interestingly, Jansen et al. (2007) also found Trails B to have poor sensitivity to chemo-fog, and this test may prove useful in distinguishing between the effects of chemotherapy versus hormonal therapy.

Similarly, Symbol Search and Digit Symbol Coding are both measures of processing speed: however, only the latter differentiated between groups. This could be because fine motor skills are recruited more in the Digit Symbol Coding task. This is consistent with the findings of Jansen et al. (2007) who reported that the Grooved Pegboard test, a measure of fine motor skills, seemed to be highly sensitive to the effects of chemotherapy.

It should also be noted that almost half of the hormonal group ($n = 29$) received radiation therapy prior to post-treatment testing as compared with three in the chemotherapy group, and it has been suggested that radiation therapy can cause detrimental effects that are similar to, and in some cases beyond, those documented in chemo-fog (Garofalo & Baum, 2001; Packer & Mehta, 2002).

Limitations and future directions

As with all research, the findings in this thesis must be considered in light of the limitations of the design, some of which have already been addressed in each individual study. However, there are some general themes that merit discussion.

One of the limitations of all three studies is the choice to include only post-menopausal women as participants. Although this effectively eliminated the confound of hormonal status on cognitive function, it also means the women in the study were in an older age range. A healthy control group was included for comparison; however, they had higher levels of education than that of the two treatment groups. As a result, it is possible that declines in scores are at least in part attributable to age-related cognitive changes.

As noted earlier, another limitation is that a subset of the chemotherapy group received hormonal treatment prior to post-treatment assessment ($n = 12$) and a subset of both groups received radiation therapy (chemotherapy group, $n = 3$; hormonal group, $n = 29$). As discussed, because these adjuvant therapies may themselves contribute to functional impairments (Bender et al., 2006; Castellon et al., 2004; Garofalo & Baum, 2001; Jenkins et al., 2004; Packer & Mehta, 2002; Paganini-Hill & Clark, 2000; Shilling et al., 2003), it is difficult to disentangle the effects of chemotherapy alone.

A major limitation of study 2 is that a separate group was not administered the reduced battery alone. Instead, the data for the reduced battery were extracted from that obtained during the more comprehensive assessment. Therefore, there is no means to account for factors that may have impacted performance on the longer battery alone, such as administration time, test order, and fatigue. However, in order to meet the objective of the paper – i.e. comparing the sensitivity of a reduced battery to that of a comprehensive one,

this approach was considered valuable to establish that a screening tool has the potential to be equally sensitive to the effects of chemo-fog prior to testing it on an independent sample.

Finally, a concern of these and all other studies employing neuropsychological measures is the limited evidence for the ecological validity of neuropsychological tests in the prediction of functional outcomes. The tests simply may not be able to adequately capture the deficits reported by survivors, and this may account for some of the discrepancy between objective results and subjective reports. This is something that would benefit from further research efforts. At this time, there is a lack of studies comparing objective versus subjective functioning in chemotherapy recipients. A prospective study evaluating change at the individual level, using a limited battery of sensitive tests and appropriate control groups, would help to clarify if neuropsychological tests are useful predictors of functionality in this population.

In addition, a debriefing with each participant immediately following test administration may help to identify instances in which participants feel a test is extremely taxing even if performance appears normal. This type of feedback could serve to further inform test selection and improve ecological validity. Furthermore, adding a neuroimaging component with an emphasis on one or more of the previously identified neurophysiological correlates of chemo-fog would allow comparison between all three aspects of functioning. This would facilitate the development of a battery of tests that is sensitive to the impairments in this population. Once such a battery is established, the examination of errors at a more in depth level would enhance understanding as to the exact nature of the deficits so that appropriate interventions can be developed.

Further research into genetic predisposition to cognitive impairment would also be beneficial. For example, studies looking at the apolipoprotein E4 allele and its role in mild cognitive impairment (Farlow et al., 2004), Parkinson's disease (Li et al., 2004), Alzheimer's (McGeer & McGeer, 2001), cognitive impairment in middle age (Blair, 2005) and cancer survivors (Ahles et al., 2003) have found that carriers of this allele are more vulnerable to cognitive impairments than control participants. With more knowledge regarding genetic vulnerabilities, treatment choices could be made that may effectively reduce the number of cancer survivors who develop chemotherapy-related cognitive deficits.

Finally, as noted by Nail (2006), it is important for future cancer drug trials to monitor the impact on cognitive function throughout clinical trials to ensure that detrimental effects are identified and limited as much as possible.

CONCLUSION

From a clinical perspective, the degree of impairment noted in the studies conducted in this area to date might be expected to have major impacts on the quality of life of those affected. Despite all of the potential confounding variables and methodological limitations that might appear sufficient to account for the declines reported in this population, a growing body of literature has established neurophysiological correlates, including electrophysiologic alterations (Kreukels et al., 2006; Schagen et al., 2001) and reduced brain activation (Kessler et al., 2009). Of particular note are the studies that have identified white and grey matter changes (Abraham et al., 2008; Inagaki et al., 2006). Collectively, these findings would suggest that if there is evidence of brain alterations and structural changes, it might

reasonably be expected that there would be corresponding behavioural or cognitive manifestations that might be improved with appropriate management.

Several interventions have been proposed for those experiencing chemo-fog, including a behavioural skills program designed to teach cancer survivors techniques for keeping track of daily schedules along with mnemonic strategies used to compensate for difficulties with memory (Ferguson et al., 2007), remediation (e.g. concentration exercises to improve attention), compensatory strategies (e.g. using an agenda to track appointments), or the use of psycho-stimulants to reduce the effects of fatigue and increase attention (Nail, 2006).

However, until screening tools are developed to permit health care providers to quickly and accurately identify patients who may be cognitively compromised during or following chemotherapy treatment, no follow-up is possible. Once identified, these individuals could then be targeted for a more thorough neuropsychological evaluation and/or be provided with appropriate interventions.

All human research has inherent limitations because of the complicated nature of human participants and the fallibility of the available tools. However, the primary objective of this thesis was to illustrate how methodological limitations can be effectively minimized by careful consideration of the characteristics of the population under investigation, and of the strengths and weaknesses of differing design choices. It is only through the use of carefully constructed and consistent research designs that results across studies will be amenable to comparison, and real insights into the nature of the difficulties experienced in this population might become clear.

REFERENCES

- Abraham, J., Haut, M. W., Moran, M. T., Filburn, S., Lemiux, S., & Kuwabara, H. (2008). Adjuvant chemotherapy for breast cancer: effects on cerebral white matter seen in diffusion tensor imaging, *Clinical Breast Cancer*, 8, 88-91.
- Ahles, T. A., Saykin, A. J., Furstenberg, C. T., Cole, B., Mott, L. A., Skalla, K., et al. (2002). Neuropsychologic impact of standard-dose systemic chemotherapy in long-term survivors of breast cancer and lymphoma. *Journal of Clinical Oncology*, 20, 485-493.
- Ahles, T. A., Saykin, A. J., Furstenberg, C. T., Cole, B., Mott, L. A., Titus-Ernstoff, L., et al. (2005). Quality of life of long-term survivors of breast cancer and lymphoma treated with standard-dose chemotherapy or local therapy. *Journal of Clinical Oncology*, 23, 4399-4405.
- Ahles, T. A., Saykin, A. J., Noll, W. W., Furstenberg, C. T., Guerin, S., Cole, B., et al. (2003). The relationship of APOE genotype to neuropsychological performance in long-term cancer survivors treated with standard dose chemotherapy. *Psychology of Women Quarterly*, 27, 612-619.
- Ahles, T. A., Tope, D. M., Furstenberg, C., Hann, D., & Mills, L. (1996). Psychologic and neuropsychologic impact of autologous bone marrow transplantation. *Journal of Clinical Oncology*, 14, 1457-1462.
- American Cancer Society. (2010). *How many women get breast cancer?* Retrieved September 11, 2010, from <http://www.cancer.org/Cancer/BreastCancer/OverviewGuide/breast-cancer-overview-key-statistics>
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental*

- disorders (4th ed)*. Washington, DC: Author.
- Ammons, R. B., & Ammons, C. H. (1962). The Quick Test (QT): provisional manual. *Psychological Reports. Monograph Supplement 1-VII*, 111-161.
- Ancelin, M. L., Ke Roquefeuil, G., Ledesert, B., Bonnel, F., Cheminal, J. C., & Ritchie, K. (2001). Exposure to anaesthetic agents, cognitive functioning and depressive symptomatology in the elderly. *British Journal of Psychiatry*, *178*, 360-366.
- Anderson-Hanley, C., Sherman, M. L., Riggs, R., Agocha, V. B., & Compas, B. E. (2003). Neuropsychological effects of treatments for adults with cancer: a meta- analysis and review of the literature. *Journal of the International Neuropsychological Society*. *9*, 967-982.
- Army Individual Test Battery (1944). *Manual for directions and scoring*. Washington, D.C.: War Department, Adjutant General's Office.
- Asato, R., Akiyama, Y., Ito, M., Kubota, M., Okumura, R., Miki, Y., et al. (1992). Nuclear magnetic resonance abnormalities of the cerebral white matter in children with acute lymphoblastic leukemia and malignant lymphoma during and after central nervous system prophylactic treatment with intrathecal methotrexate. *Cancer*, *70*, 1997-2004.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory (2nd ed)*. & *Manual*. San Antonio, TX: Psychological Assessment Resources.
- Bender, C. M., Sereika, S. M., Berga, S. L., Vogel, V. G., Brufsky, A. M., Paraska, K. K., et al. (2006). Cognitive impairment associated with adjuvant therapy in breast cancer. *Psycho-Oncology*. *15*, 422-430.
- Benedict, R. H. B., Munschauer, F., Linn, R., Miller, C., Murphy, E., Foley, F., et al. (2003). Screening for multiple sclerosis cognitive impairment using a self-administered 15-

- item questionnaire. *Multiple Sclerosis*, 9, 95-101.
- Blair, C. K., Folsom, A. R., Knopman, D. S., Bray, M. S., Mosley, T. H., & Boerwinkle, E. (2005). APOE genotype and cognitive decline in a middle-aged cohort. *Neurology*, 64, 268-276.
- Bost, R. H., Wen, F. K., Basso, M. R., & Cates, G. R. (2008). Online tools for evaluating patient change: statistical foundations, clinical applications, research relevance. *Rehabilitation Psychology*, 53, 313-320.
- Brezden, C. B., Phillips, K., Abdolell, M., Bunston, T. & Tannock, I. F. (2000). Cognitive function in breast cancer patients receiving adjuvant chemotherapy. *Journal of Clinical Oncology*. 18, 2695-2701.
- Brooks, B. L., Strauss, E., Sherman, E. M. S., Iverson, G. L., & Slick, D. J. (2009) Developments in neuropsychological assessment: refining psychometric and clinical interpretive methods. *Canadian Psychology*, 50, 196-209.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal Of Experimental Psychology*, 10, 12-21.
- Bush, S. S. (2010). Determining whether or when to adopt new versions of psychological and neuropsychological tests: ethical and professional considerations. *The Clinical Neuropsychologist*, 24, 7-16.
- Canadian Cancer Society. (2010). Breast cancer statistics. Retrieved September 11, 2010, from http://www.cancer.ca/Canada-wide/About%20cancer/Cancer%20statistics/Stats%20at%20a%20glance/Breast%20cancer.aspx?sc_lang=en
- Carroll, J. B. (1993). *Human cognitive abilities: a survey of factor-analytic studies*.

Cambridge University Press: New York.

- Castellon, S. A., Ganz, P. A., Bower, J. E., Peterson, L., Abraham, L., & Greendale, G. A. (2004). Neurocognitive performance in breast cancer survivors exposed to adjuvant chemotherapy and tamoxifen. *Journal of Clinical and Experimental Neuropsychology*, *26*, 955-969.
- Castellon, S. A., Silverman, D. H. D., & Ganz, P. A. (2005). Breast cancer treatment and cognitive functioning: current status and future challenges in assessment. *Breast Cancer Research and Treatment*, *92*, 299-206.
- Cella, D. F., Tulsky, D. S., Gray, G., Safavian, B., Linn, E., Bonomi, A., et al. (1993). The functional assessment of cancer therapy scale: development and validation of the general measure. *Journal of Clinical Oncology*, *11*, 570-579.
- Collins, B., Mackenzie, J., Stewart, A., Bielajew, C., & Verma, S. (2008). Cognitive effects of chemotherapy in post-menopausal breast cancer patients 1 year after treatment. *Psycho-Oncology*, *18*, 134-143.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *10*, 1-9.
- Crawford, J. R., Obansawin, M. C., & Allan, K. M. (1988). PASAT and components of WAIS-R performance: convergent and discriminant validity. *Neuropsychological Rehabilitation*, *8*, 255-272.
- Culley, D. J., Xie, Z., & Crosby, G. (2007). General anesthetic-induced neurotoxicity: an emerging problem for the young and old? *Current Opinion in Anaesthesiology*, *20*, 408-413.

- Daly, E., Kormaroff, A. L., Bloomingdale, K., Wilson, S., & Albert, M. S. (2001). Neuropsychological function in patients with chronic fatigue syndrome, multiple sclerosis, and depression. *Applied Neuropsychology, 8*, 12-20.
- Debess, J., Riss, J. O., Engebjerg, M. C., & Ewertz, M. (2010). Cognitive function after adjuvant treatment for early breast cancer: a population-based longitudinal study. *Breast Cancer Research and Treatment, 121*, 91-100.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *The California Verbal Learning Test II Manual*. New York: Psychological Corporation.
- Dick, F., Semple, S., Soutar, A., Osborne, A., Cherrie, J. W., & Seaton, A. (2004). Is colour vision impairment associated with cognitive impairment in solvent exposed workers? *Occupational and Environmental Medicine, 61*, 76-78.
- Dikmen, S. S., Heaton, R. B., Grant, I., & Temkin, N. R. (1999). Test-retest reliability and practice effects of Expanded Halstead-Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society, 5*, 346-356.
- Donovan, K. A., Small, B. J., Andrykowski, M. A., Schmitt, F. A., Munster, P., & Jacobsen, P. B. (2005). Cognitive functioning after adjuvant chemotherapy and/or radiotherapy for early-stage breast carcinoma. *Cancer, 104*, 2499-2507.
- Downie, F. P., Mar Fan, H. G., Tchen, A., Yi, Q., & Tannock, I. F. (2006). Cognitive function, fatigue, and menopausal symptoms in breast cancer patients receiving adjuvant chemotherapy: evaluation with patient interview after formal assessment. *Psycho-Oncology, 15*, 921-930.
- Eberling, J. L., Wu, C., Tong-Turnbeaugh, R., & Jagust, W. J. (2004). Estrogen- and tamoxifen-associated effects on brain structure and function. *Neuroimage, 21*, 364-

371.

- Falletti, M. G., Sanfilippo, A., Maruff, P., Weih, L., & Phillips, K. (2005). The nature and severity of cognitive impairment associated with adjuvant chemotherapy in women with breast cancer: a meta-analysis of the current literature. *Brain and Cognition, 59*, 60-70.
- Fallowfield, L. J., Leaity, S. K., Howell, A., Benson, S. & Cella, D. (1999). Assessment of quality of life in women undergoing hormonal therapy for breast cancer: validation of an endocrine symptom subscale for the FACT-B. *Breast Cancer Research and Treatment, 55*, 189-199.
- Farlow, M. R., He, T., Tekin, S., Xu, J., Lane, R., & Charles, H. C. (2004). Impact of APOE in mild cognitive impairment. *Neurology, 63*, 1898-1901.
- Faust, D., & Fogel, B. S. (1989). The development and initial validation of a sensitive bedside cognitive screening test. *Journal of Nervous and Mental Disease, 177*, 25-31.
- Ferguson, R. J., Ahles, T. A., Saykin, A. J., McDonald, B. C., Furstenberg, C. T., Cole, B. F. et al. (2007). Cognitive-behavioral management of chemotherapy-related cognitive change. *Psycho-Oncology, 16*, 772-777.
- Filley, C. (1999). Toxic leukoencephalopathy. *Clinical Neuropharmacology, 22*, 249-260.
- Fogel, B. S., (1991). The High Sensitivity Cognitive Screen. *International Psychogeriatrics, 3*, 273-288.
- Franzen, M. D., Haut, M. W., Rankin, E., & Keefover, R. (1995). Empirical comparison of alternate forms of the Boston Naming Test. *The Clinical Neuropsychologist, 9*, 225-229.

- Freeman, J. R., & Broshek, D. K. (2002). Assessing cognitive dysfunction in breast cancer: what are the tools? *Clinical Breast Cancer, 3*, 591-599.
- Frerichs, R. J., & Tuokko, H. A. (2006). Reliable change scores and their relation to perceived change in memory: implications for the diagnosis of mild cognitive impairment. *Archives of Clinical Neuropsychology, 21*, 109-115.
- Frigon, J. Y., & Laurencelle, L. (1993). An analysis of covariance: a proposed algorithm. *Educational and Psychological Measurement, 53*, 1-18.
- Garofalo, J. P., & Baum, A. (2001). Neurocognitive sequelae of cancer therapies. *Current Opinion in Psychiatry, 14*, 575-583.
- Glass, J. P., Lee, Y., Bruner, Y., & Fields, W. S. (1986). Treatment-related leukoencephalopathy: a study of three cases and literature review. *Medicine, 65*, 154-162.
- Goodglass, H., Kaplan, E., & Weintraub, S. (1983). *The Boston Naming Test (2nd ed)*. Philadelphia: Lea & Febiger.
- Gottschalk, L. A., Holcombe, R. F., Jackson, D., & Bechtel, R. J. (2003). The effects of anticancer chemotherapeutic drugs on cognitive function and other neuropsychiatric dimensions in breast cancer patients. *Methods and Findings in Experimental and Clinical Pharmacology, 25*, 117-122.
- Goverover, Y., Chiaravalloti, N., & DeLuca, J. (2005). The relationship between self-awareness of neurobehavioral symptoms, cognitive functioning, and emotional symptoms in multiple sclerosis. *Multiple Sclerosis, 11*, 203-212.
- Gronwall, D. M. A. (1977). Paced Auditory Serial Addition Task. A measure of recovery from concussion. *Perceptual and Motor Skills, 44*, 367-373.

- Hafner, D. L. (2009). Lost in the fog: understanding “chemo brain”. *Nursing*, 39, 43-45.
- Hanning, C. D. (2005). Postoperative cognitive dysfunction. *British Journal of Anaesthesia*, 95, 82-87.
- Heaton, R. K. (1981). *Wisconsin Card Sorting Test Manual*. Odessa, FL: Psychological Assessment Resources.
- Heaton, R. K., Avitable, N., Grant, I., & Matthews, C. G. (1999). Further crossvalidation of regression-based neuropsychological norms with an update for the Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology*, 21, 572-582.
- Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: revised and expanded*. Odessa, FL: Psychological Assessment Resources.
- Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an expanded Halstead-Reitan Battery: demographic corrections, research findings and clinical applications*. Odessa, FL: Psychological Assessment Resources.
- Heaton, R. K., Temkin, N., Dikmen, S., Avitable, N., Taylor, M. J., Marcotte, T. D., et al. (2001). Detecting change: a comparison of three neuropsychological methods, using normal and clinical samples. *Archives of Clinical Neuropsychology*, 16, 75-91.
- Health Canada. (2009). *Authorization with conditions of avastin (bevacizumab) for use in combination with paclitaxel for the treatment of patients with metastatic HER2-negative breast cancer who are ECOG class 0-1*. Retrieved August 31, 2010 from http://www.hc-sc.gc.ca/dhp-mps/alt_formats/hpfb/dgpsa/pdf/prodpharma/avastin_dhcpl_lapds_107746-eng.pdf
- Hebben, N. & Milberg, W. (2009). *Essentials of neuropsychological assessment* (2nd ed.).

Hoboken, NJ, US: John Wiley & Sons Inc; US.

Hermelink, K., Untch, M., Lux, M. P., Kreienberg, R., Beck, T., Bauerfeind, I., et al. (2007).

Cognitive function during neoadjuvant chemotherapy for breast cancer. *Cancer*, *109*, 1905-1913.

Holtzer, R., Stern, Y., & Rakitin, B. C. (2005). Predicting age-related dual-task effects with individual differences on neuropsychological tests. *Neuropsychology*, *19*, 18-27.

Hurria, A., & Lachs, M. (2007). Is cognitive dysfunction a complication of adjuvant chemotherapy in the older patient with breast cancer? *Breast Cancer Research and Treatment*, *103*, 259-268.

Inagaki, M., Yoshikawa, E., Matsuoka, Y., Sugawara, Y., Nakano, T., Akechi, T., et al. (2007). Smaller regional volumes of brain gray and white matter demonstrated in breast cancer survivors exposed to adjuvant chemotherapy. *Cancer*, *109*, 146-156.

Isella, V., Atzeni, L., Iurlaro, S., Villa, M. L., Russo, A., Forapani, E., et al. 2003). Assessing clinically relevant cognitive decline: preliminary data on a new method. *Neurological Sciences*, *24*, 236-241.

Ivnik, R. J., Malee, J. F., Smith, G. E., Tangalos, E. G., Petersen, R. C., Kokmen, E., et al. (1992). Mayo's Older Americans Normative Studies: WAIS-R norms for ages 56-97. *Clinical Neuropsychologist*, *6*, 1-30.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, *15*, 336-352.

- Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12-19.
- Jansen, C. E., Miaskowski, C., Dodd, M., & Dowling, G. (2005). Chemotherapy-induced cognitive impairment in women with breast cancer: a critique of the literature. *Oncology Nursing Forum*, *32*, 329-342.
- Jansen, C. E., Miaskowski, C. A., Dodd, M. J., & Dowling, G. A. (2007). A meta-analysis of the sensitivity of various neuropsychological tests used to detect chemotherapy-induced cognitive impairments in patients with breast cancer. *Oncology Nursing Forum*, *34*, 997-1005.
- Jenkins, V., Shilling, V., Fallowfield, L., Howell, A., & Hutton, S. (2004). Does hormone therapy for the treatment of breast cancer have a detrimental effect on memory and cognition? A pilot study. *Psycho-Oncology*, *13*, 61-66.
- Jenkins, V., Shilling, V., Deutsch, G., Bloomfield, D., Morris, R., Allan, S., et al. (2006). A 3-year prospective study of the effects of adjuvant treatments on cognition in women with early stage breast cancer. *British Journal of Cancer*, *94*, 828-834.
- Jim, H. S. L., Donovan, K. A., Small, B. J., Andrykowski, M. A., Munster, P. N., & Jacobsen, P. B. (2009). Cognitive functioning in breast cancer survivors: a controlled comparison. *Cancer*, *115*, 1776-1783.
- Joyce, E. M., Hutton, S. B., Mutsatsa, S. H. & Barnes, T. R. (2005). Cognitive heterogeneity in first-episode schizophrenia. *British Journal of Psychiatry*, *187*, 516-522.
- Kaiser, H. F. & Rice, J. (1974). Little jiffy mark IV. *Educational and Psychological Measurement*, *34*, 111-117.

- Kayl, A. E., Wefel, J. S., & Meyers, C. A. (2006). Chemotherapy and cognition: effects, potential mechanisms, and management. *American Journal of Therapeutics, 12*, 362-369.
- Kesler, S. R., Bennett, F. C., Mahaffey, M. L., & Spiegel, D. (2009). Regional brain activation during verbal declarative memory in metastatic breast cancer. *Clinical Cancer Research, 15*, 6665-6673.
- Kirk, R. E. (1968). *Experimental design: procedures for the behavioral sciences*. Brooks/Cole Publishing Company: Belmont, California.
- Kreukels, B. P., Schagen, S. B., Ridderinkhof, K. R., Boogerd, W., Lamburger, H. L., & van Dam, F. S. (2005). Electrophysiological correlates of information processing in breast- cancer patients treated with adjuvant chemotherapy. *Breast Cancer Research and Treatment, 94*, 53-61.
- Kreukels, B. P. C., Schagen, S. B., Ridderinkhof, K. R., Boogerd, W., Hamburger, H. L., Muller, M. J., et al. (2006). Effects of high-dose and conventional-dose adjuvant chemotherapy on long-term cognitive sequelae in patients with breast cancer: an electrophysiologic study. *Clinical Breast Cancer, 7*, 67-78.
- Lezak, M.D., Howieson, D.B., & Loring, D.W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.
- Li, Y. J., Hauser, M. A., Scott, W. K., Martin, E. R., Booze, M. W., Qin, X. J. et al. (2004). Apolipoprotein E controls the risk and age at onset of Parkinson disease. *Neurology, 62*, 2005-2009.
- LoGalbo, A., Sawrie, S., Roth, D. L., Kuzniecky, R., Knowlton, R., Faught, E., et al. (2005). Verbal memory outcome in patients with normal preoperative verbal memory and left

mesial temporal sclerosis. *Epilepsy & Behavior*. 6, 337-341.

Lo Presti, A., Ruvolo, G., Gancitano, R. A., & Cittadini, E. (2004). Ovarian function following radiation and chemotherapy for cancer. *European Journal of Obstetrics & Gynecology and Reproduction Biology*, 113(Suppl.), 33-40.

Lucas, J., & Addeo, R.R. (2005). Traumatic brain injury and post-concussive syndrome. In P.J. Snyder & P.D. Nussbaum (Eds) *Clinical Neuropsychology: A Pocket Handbook for Assessment (2nd Ed)* pp. 351-380. Washington, DC: American Psychological Association.

Lucas, J. A., Ivnik, R. J., Willis, F. B., Ferman, T. J., Smith, G. E., Parfitt, F. C., et al. (2005). Mayo's older African American normative studies: normative data for commonly used clinical neuropsychological measures. *Clinical Neuropsychologist*, 19, 162-183.

Mar Fan, H. G., Houede-Tchen, N., Yi, Q., Chemerynsky, I., Downie, F. P., Sabate, K., et al. (2005). Fatigue, menopausal symptoms, and cognitive function in women after adjuvant chemotherapy for breast cancer: 1 and 2 year follow-up of a prospective controlled study. *Journal of Clinical Oncology*, 23, 8025-8032.

Martin, R., Griffith, H. R., Sawrie, S., Knowlton, R., & Faught, E. (2006). Determining empirically based self-reported cognitive change: development of reliable change indices and standardized regression-based change norms for the multiple abilities self-report questionnaire in an epilepsy sample. *Epilepsy & Behavior*. 8, 239-245.

Martin, R., Sawrie, S., Gilliam, F., Mackey, M., Faught, E., Knowlton, R., et al. (2002). Determining reliable cognitive change after epilepsy surgery: development of reliable

- change indices and standardized regression-based change norms for the WMS-III and WAIS-III. *Epilepsia*, 43, 1551-1555.
- McCaffrey, R. J., Cousins, J. P., Westervelt, H. J., Martnowicz, M., Remick, S. C., Szebenyi, S., et al. (1995). Practice effect with the NIMH AIDS abbreviated neuropsychological battery. *Archives of Clinical Neuropsychology*, 10, 241-250.
- McCracken, L. M. & Franzen, M. D. (1992). Principal-components analysis of the equivalence of alternate forms of the trail making test. *Psychological Assessment*, 4, 235-238.
- McCrea, M., Barr, W. B., Guskiewica, K., Randolph, C., Marshall, S. W., Cantu, R., et al. (2005). Standard regression-based methods for measuring recovery after sport-related concussion. *Journal of the International Neuropsychological Society*, 11, 58-69.
- McGeer, P., & McGeer, E. G. (2001). Polymorphisms in inflammatory genes and risk of Alzheimer disease. *Archives of Neurology*, 58, 1790-1792.
- McNair, D. M., Lorr, M., & Droppleman, L. F. (1992). *Manual for the Profile of Mood States*. San Diego, CA, Educational and Industrial Service.
- McSweeney, A. J., Naugle, R. I., Chelune, G. J., & Lunders, H. (1993). "T scores for change": an illustration of a regression approach to depicting change in clinical neuropsychology. *Clinical Neuropsychology*, 7, 300-312.
- Mehnert, A., Scherwath, A., Schirmer, L., Schleimer, B., Peterson, C., Schulz-Kindermann, et al. (2007). The association between neuropsychological impairment, self-perceived cognitive deficits, fatigue and health related quality of life in breast cancer survivors following standard adjuvant versus high-dose chemotherapy. *Patient Education and Counseling*, 66, 108-118.

- Mertens, V. B., Gagnon, M., Coulombe, D., & Messier, C. (2006). Exploratory factor analysis of neuropsychological tests and their relationship to the Brown-Peterson task. *Archives of Clinical Neuropsychology, 21*, 733-739.
- Mierow, D. (1999). Ovarian injury and modern options to preserve fertility in female cancer patients treated with high dose radio-chemotherapy for hemato-oncological neoplasias and other cancers. *Leukemia and Lymphoma, 33*, 65-76.
- Mierow, D. (2000). Reproduction post-chemotherapy in young cancer patients. *Molecular and Cellular Endocrinology, 169*, 123-131.
- Mitrushina, M. N., Boone, B. B., & D'Elia, L. F. (1999). *Handbook of Normative Data for Neuropsychological Assessment*. New York: Oxford University Press.
- Mitrushina, M. N., Boone, B. B., Razani, J., & D'Elia, L. F. (2005) *Handbook of Normative Data for Neuropsychological Assessment*. New York : Oxford University Press.
- Monk, T. G., Weldon, B. C., Gavan, C. W., Dede, D. E., van der Aa, M. T., Heilman, K. M., et al. (2008). Predictors of cognitive dysfunction after major noncardiac surgery. *Anesthesiology, 108*, 18-30.
- Montgomery, G. H., & Bovbjerg, D. H. (2004). Presurgery distress and specific response expectancies predict postsurgery outcomes in surgery patients confronting breast cancer. *Health Psychology, 23*, 381-387.
- Mulrooney, T. (2008). Cognitive impairment after breast cancer treatment. *Clinical Journal of Oncology Nursing, 12*, 678-680.
- Nail, L. M. (2006). Cognitive changes in cancer survivors. *American Journal of Nursing, 106*(Suppl. 3), 48-54.
- National Cancer Institute. (2007). Finding cancer statistics: cancer stats fact sheets – cancer

of the breast. Retrieved November 19, 2007, from:

<http://seer.cancer.gov/statfacts/html/breast.html>

- Nelson, H. E., & McKenna, P. (1975). The use of current reading ability in the assessment of dementia. *British Journal of Social and Clinical Psychology, 14*, 259-267.
- Neubauer, R. A., & Golden, C. (2005). Can postoperative cognitive dysfunction be prevented? *Journal of American Physicians and Surgeons, 10*, 22.
- New, P. Z. (2005). Neurological complications of chemotherapeutic and biological agents. *Neuro-Oncology, 11*, 116-152.
- Newby, R. F., Hallenbeck, C. E., & Embretson, S. (1983). Confirmatory factor analysis of four general neuropsychological models with a modified halstead-reitan battery. *Journal of Clinical Neuropsychology, 5*, 115-133.
- O'Donnell, B. F., Drachman, D. A., Lew, R. A., & Swearer, J. M. (1988). Measuring dementia: assessment of multiple deficit domains. *Journal of Clinical Psychology, 44*, 916-923.
- Olin, J. J. (2001). Cognitive function after systemic therapy for breast cancer. *Oncology, 15*, 613-624.
- O'Shaughnessy, J. A. (2003). Chemotherapy-induced cognitive dysfunction: a clearer picture. *Clinical Breast Cancer, 4*(Suppl. 2), 89-94.
- Ouimet, L. A., Stewart, A., Collins, B., Schindler, D., & Bielajew, C. (2009). Measuring neuropsychological change following breast cancer: an analysis of statistical models. *Journal of Clinical and Experimental Neuropsychology, 31*, 73-89.
- Ouimet, L. A., Stewart, A., Collins, B., Schindler, D., & Bielajew, C. (2010). Investigation of a neuropsychological screen for chemo-fog. Manuscript submitted for publication.

- Packer, R. J., & Mehta, M. Neurocognitive sequelae of cancer treatment. *American Academy of Neurology*, 59, 8-10.
- Paganini-Hill, A., & Clark, L. J. (2000). Preliminary assessment of cognitive function in breast cancer patients treated with tamoxifen. *Breast Cancer Research and Treatment*, 64, 165-176.
- Peace, K. A., Orme, S. M., Padayatty, S. J., Godfrey, H. P. D., & Belchetz, P. E. (1998). Cognitive dysfunction in patients with pituitary tumour who have been treated with transfrontal or transsphenoidal surgery or medication. *Clinical Endocrinology*, 49, 391-396.
- Pedraza, O., Lucas, J. A., Smith, G. E., Willis, F. B., Graff-Radford, N. R., Ferman, T. J., et al. (2005). Mayo's older African American normative studies: confirmatory factor analysis of a core battery. *Journal of the International Neuropsychological Society*, 11, 184-191.
- Phillips-Bute, B., Mathew, J. P., Blumenthal, J. A., Grocott, H. P., Laskowitz, D. T., Jones, R. H., et al. (2006). Association of neurocognitive function and quality of life 1 year after coronary artery bypass graft (CABG) surgery. *Psychosomatic Medicine*, 68, 369-375.
- Predictive Analytic Software. Version 18.0, SPSS Inc., Chicago, Illinois.
- Public Health Agency of Canada. (2007). *Breast Cancer*. Retrieved November 19, 2007, from: http://www.phac-aspc.gc.ca/ccdpc-cpcmc/bc-cds/index_e.html
- Raffa, R. B., & Tallarida, R. J. (2010). Effects on the visual system might contribute to some of the cognitive deficits of cancer chemotherapy-induced 'chemo-fog'. *Journal of Clinical Pharmacy and Therapeutics*, 35, 249-255.

- Raymond, P. D., Hinton-Bayre, A. D., Radel, M., Ray, M. J., & Marsh, N. A. (2006). Assessment of statistical change criteria used to define significant change in neuropsychological test performance following cardiac surgery. *European Journal of Cardio-thoracic Surgery*, 29, 82-88.
- Reid-Arndt, S. A., Yee, A., Perry, M. C., & Hsieh, C. (2009). Cognitive and psychological factors associated with early posttreatment functional outcomes in breast cancer survivors. *Journal of Psychosocial Oncology*, 27, 415-434.
- Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan Neuropsychological Test Battery*. Tucson, AZ: Neuropsychology Press.
- Resampling Stats for Excel. Version 4.0.
- Rey, A. (1968). *Epreuves mnesiques et d'apprentissage*. Delachaux & Nestle.
- Ruff, R. M., & Parker, S. B. (1993). Gender- and -age-specific changes in motor speed and eye-hand coordination in adults: normative values for the Finger Tapping and Grooved Pegboard Tests. *Perceptual and Motor Skills*, 76, 1219-1230.
- Sawrie, S. M. (2002). Analysis of cognitive change: a commentary on Keith et al. (2002). *Neuropsychology*, 16, 429-431.
- Sawrie, S. M., Chelune, G. J., Naugle, R. I., & Luders, H. O. (1996). Empirical methods for assessing meaningful change following epilepsy surgery. *Journal of the International Neuropsychological Association*, 2, 556-564.
- Sawrie, S. M., Marson, D. C., Boothe, A. L., & Harrell, L. E. (1999). A method for assessing clinically relevant individual cognitive change in older adult populations. *Journal of Gerontology*, 54B, 116-124.
- Saykin, A. J., Ahles, T., A., & McDonald, B. C. (2003). Mechanisms of chemotherapy

- induced cognitive disorders: neuropsychological, pathophysiological, and neuroimaging perspectives. *Seminars in Clinical Neuropsychiatry*. 8, 201-216.
- Schagen, S. B., Hamburger, H. L., Muller, M. J., Boogerd, W. & van Dam, F. S. (2001). Neurophysiological evaluation of late effects of adjuvant high-dose chemotherapy on cognitive function. *Journal of Neuro-Oncology*. 51, 159-165.
- Schagen, S. B., Muller, M. J., Boogerd, W., Rosenbrand, R. M., van Rhijn, D., Rodenhuis, S. et al. (2002a). Late effects of adjuvant chemotherapy on cognitive function: a follow-up study in breast cancer patients. *Annals of Oncology*. 13, 1387-1397.
- Schagen, S. B., Muller, M. J., Boogerd, W., & van Dam, F. S. (2002b). Cognitive dysfunction and chemotherapy: neuropsychological findings in perspective. *Clinical Breast Cancer*. 3(Suppl. 3), 100-108.
- Schagen, S. B., van Dam, F. S., Muller, M. J., Boogerd, W. Lindeboom, J., & Bruning, P. F. (1999). Cognitive deficits after postoperative adjuvant chemotherapy for breast carcinoma. *Cancer*. 85, 640-650.
- Scherwath, A., Mehnert, A., Schleimer, B., Schirmer, L., Fehlaue, F., Kreienberg, R., et al. (2006). Neuropsychological function in high-risk breast cancer survivors after stem-cell supported high-dose therapy versus standard-dose chemotherapy: evaluation of long-term treatment effects. *Annals of Oncology*. 17, 415-423.
- Schilder, C. M., Eggens, P. C., Seynaeve, C., Linn, S. C., Boogerd, W., Gundy, C. M., et al. (2009). Neuropsychological functioning in postmenopausal breast cancer patients treated with tamoxifen or exemestane after AC-chemotherapy: cross-sectional findings from the neuropsychological TEAM-side study. *Acta Oncologica*, 48, 76-85.
- Schilder, C., & Schagen, S. (2007). Effects of hormonal therapy on cognitive functioning in

- Breast cancer patients: a review of the literature. *Minerva ginecologica*, 59, 387-401.
- Selnes, O. A., Grega, M. A., Borowicz, L. M., Barry, S., Zeger, S., & McKhann, G. M. (2004). Self-reported memory symptoms with coronary artery disease: a prospective study of CABG patients and nonsurgical controls. *Cognitive and Behavioral Neurology*, 17, 148-156.
- Servaes, P., Verhagen, C. A., & Bleijenberg, G. (2002). Relations between fatigue, neuropsychological functioning, and physical activity after treatment for breast carcinoma. *Cancer*, 95, 2017-2026.
- Shilling, V., & Jenkins, V. (2007). Self-reported cognitive problems in women receiving adjuvant therapy for breast cancer. *European Journal of Oncology Nursing*, 11, 6-15.
- Shilling, V., Jenkins, V., Fallowfield, L., & Howell, T. (2003). The effects of hormone therapy on cognition in breast cancer. *Journal of Steroid Biochemistry & Molecular Biology*, 86, 405-412.
- Shilling, V., Jenkins, V., Morris, R., Deutsch, G., & Bloomfield, D. (2005). The effects of adjuvant chemotherapy on cognition in women with breast cancer – preliminary results of an observational longitudinal study. *The Breast*, 14, 142-150.
- Shilling, V., Jenkins, V., & Trapala, I. S. (2006). The (mis)classification of chemo-fog – methodological inconsistencies in the investigation of cognitive impairment after chemotherapy. *Breast Cancer Research and Treatment*, 95, 125-129.
- Silverstein, M. L., & Nelson, L. D. (2000). Clinical and research implications of revising psychological tests. *Psychological Assessment*, 12, 298-303.
- Spreen, O., & Benton, A. L. (1969). *Neurosensory Centre Comprehensive Examination for Aphasia (NCCEA)*. Victoria: University of Victoria, Neuropsychology Laboratory.

- Spreeen, O. & Strauss, E. (1991). *A compendium of neuropsychological tests*. New York, NY: Oxford University Press.
- Spreeen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary (2nd ed)*. New York, NY: Oxford University Press.
- Statistical Package for the Social Sciences. Version 15.0, Chicago, Illinois.
- Stewart, A., Bielajew, C., Collins, B., Parkinson, M., & Tomiak, E. (2006). A meta-analysis of the neuropsychological effects of adjuvant chemotherapy treatment in women treated for breast cancer. *Clinical Neuropsychology*, *20*, 76-89
- Stewart, A., Collins, B., Mackenzie, J., Tomiak, E., Verma, S., & Bielajew, C. (2008). The cognitive effects of adjuvant chemotherapy in early stage breast cancer: a prospective study. *Psycho-Oncology*, *17*, 122-130.
- Stewart, R., Dufouil, C., Godin, O., Ritchie, K., Maillard, P., Delcroix, N., et al., (2008). Neuroimaging correlates of subjective memory deficits in a community population. *Neurology*, *70*, 1601-1607.
- Strauss, M. E. & Fritsch, T. (2004). Factor structure of the CERAD neuropsychological battery. *Journal of the International Neuropsychological Society*, *10*, 559-565.
- Strauss, E., Sherman, E., & Spreeen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary (3rd. ed)*. New York, NY, US: Oxford University Press; US.
- Strauss, E., Spreeen, O., & Hunter, M. (2000). Implications of test revisions for research. *Psychological Assessment*, *12*, 237-244.

- Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics (5th edition)*. Allyn and Bacon: Boston.
- Tager, F. A., McKinley, P. S., Schnabel, F. R., El-Tamer, M., Cheung, Y. K. K., Fang, Y. et al. (2010). The cognitive effects of chemotherapy in post-menopausal breast cancer patients: a controlled longitudinal study. *Breast Cancer Research and Treatment*, *123*, 25-34.
- Taillibert, S., Voillery, D., & Bernard-Marty, C. (2007). Chemobrain: is systemic chemotherapy neurotoxic? *Current Opinion in Oncology*, *19*, 623-627.
- Tannock, I. F., Ahles, T. A., Ganz, P. A., & van Dam, F. S. (2004). Cognitive impairment associated with chemotherapy for cancer: report of a workshop. *Journal of Clinical Oncology*, *22*, 2233-2239.
- Tchen, N., Juffs, H. G., Downie, F. P., Yi, Q., Hu, H., Chemerynsky, I. et al. (2003). Cognitive function, fatigue, and menopausal symptoms in women receiving adjuvant chemotherapy for breast cancer. *Journal of Clinical Oncology*. *22*, 4175-4183.
- Temkin, N. R., Heaton, R. K., Grant, I., & Dikmen, S. S. (1999). Detecting significant change in neuropsychological test performance: a comparison of four models. *Journal of the International Neuropsychological Society*. *5*, 357-369.
- Tiersky, L. A., Johnson, S. K., Lange, G., Natelson, B. H., & DeLuca, J. (1997). Neuropsychology of chronic fatigue syndrome: a critical review, *Journal of Clinical and Experimental Neuropsychology*, *19*, 560-586.
- Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology*, *14*, 167-177.

- Torrents, E., Boiso, I, Barri, P. N., & Veiga, A. (2003). Applications of ovarian tissue transplantation in experimental biology and medicine. *Human Reproduction Update*, 9, 471-481.
- Traub, G. S., & Spruill, J. (1982). Correlations between the Quick Test and Wechsler Adult Intelligence Scale – Revised. *Psychological Reports*, 51, 309-310.
- Troy, L., McFarland, K., Littman-Power, S., Kelly, B. J., Walpole, E. T., Wyld, D. et al. (2000). Cisplatin-based therapy: a neurological and neuropsychological review. *Psycho-oncology*, 9, 29-39.
- Tuxen, M. K. & Werner, H. S. (1994). Neurotoxicity secondary to antineoplastic drugs. *Cancer Treatment Review*, 20, 191-214.
- van Dam, F. S., Schagen, S. B., Muller, M. J., Boogerd, W., Wall, E., Droogleever Fortuyn, M. E., et al. (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, 210-218.
- Vanderploeg, R. D., & Schinka, J. A. (2004). Estimation of premorbid cognitive abilities: issues and approaches. In J. H. Ricker (Ed) *Differential Diagnosis in Adult Neuropsychological Assessment* pp. 27-65. New York, NY: Springer Publishing Company Inc.
- Vardy, J., Rourke, S., & Tannock, I. F. (2007). Evaluation of cognitive function associated with chemotherapy: a review of published studies and recommendations for future research. *Journal of Clinical Oncology*, 25, 2455-2463.
- Vardy, J., Wong, K., Yi, Q., Park, A., Maruff, P., Wagner, L., et al. (2006). Assessing cognitive function in cancer patients. *Supportive Care In Cancer*, 14, 1111-1118.

- Verstappen, C. C. P., Heimans, J. J., Hoekman, K., & Postma, T. J. (2003). Neurotoxic complications of chemotherapy in patients with cancer. *Drugs*, *63*, 1549-1563.
- Von Ah, D., Russell, K. M., Storniolo, A. M., & Carpenter, J. S. (2009). Cognitive dysfunction and its relationship to quality of life in breast cancer survivors. *Oncology Nursing Forum*, *36*, 326- 336.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale – Revised*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale and manual (3rd ed)*. San Antonio, TX: Harcourt, Brace, & Co.
- Wechsler, D. (1997b). *Wechsler Memory Scale and manual (3rd ed)*. San Antonio, TX: Harcourt, Brace, & Co.
- Wefel, J. S., Lenzi, R., Theriault, R. Buzdar, A. U., Cruickshank, S., & Meyers, C. A. (2004a). ‘Chemobrain’ in breast carcinoma? A prologue. *Cancer*, *101*, 466-475.
- Wefel, J. S., Lenzi, R., Theriault, R., Davis, R. N., & Meyers, C. A., (2004b). The cognitive sequelae of standard-dose adjuvant chemotherapy in women with breast carcinoma. *Cancer*, *100*, 2292-2299.
- Weis, J., Poppelreuter, M., & Bartsch, H. H. (2009). Cognitive deficits as long-term side-effects of adjuvant therapy in breast cancer patients: ‘subjective’ complaints and ‘objective’ neuropsychological test results. *Psycho-Oncology*, *18*, 775-782.
- Wieneke, M. H., & Dienst, E. R. (1995). Neuropsychological assessment of cognitive functioning following chemotherapy for breast cancer. *Psycho-Oncology*, *4*, 61-66.
- Wong, T. M. (2006). Ethical controversies in neuropsychological test selection, administration, and interpretation. *Applied Neuropsychology*, *13*, 68-76.

Yellen, S. B., Cella, D. F., Webster, K., Blendowski, C., & Kaplan, E. (1997). Measuring fatigue and other anemia-related symptoms with the functional assessment of cancer therapy (FACT) measurement system. *Journal of Pain and Symptom Management*, *13*, 63-74.

APPENDIX A

CONTRIBUTIONS OF THE COLLABORATORS

The formulation of the three studies in this thesis was a collaborative effort between me and Dr. Bielajew. Data were collected by Dr. Stewart as part of her dissertation project, and was overseen by Dr. Collins. For all studies, I was responsible for data analysis, interpretation, and preparation of the manuscripts. Consultations regarding statistical questions were handled by Dr. Schindler and manuscript editing was done in conjunction with Dr. Bielajew.

APPENDIX B

DESCRIPTION OF NEUROPSYCHOLOGICAL AND PSYCHOLOGICAL MEASURES

The following text describes the psychometric properties for neuropsychological tests included in these studies. All scoring was completed using the standard instructions from the test manuals. Neuropsychological tests are presented by domain, followed by the psychological tests in alphabetical order.

Neuropsychological Tests

Executive function

Paced Auditory Serial Addition Test

The Paced Auditory Serial Addition Test (Gronwall, 1977) is a serial-addition task used to evaluate divided attention, working memory and information processing. An audiotape is used to present a string of 61 random numbers from one through nine. The participant is required to add each number to the one preceding it. There are four trials, each one using a faster presentation time than the previous one (2.4, 2.0, 1.6 and 1.2 seconds, respectively), with a 60 second break between each trial. The correct number of responses is calculated for each trial, with a maximum score of 60.

The measure has demonstrated very good internal reliability in adults (Cronbach's alpha $r = .90$; Crawford, Obansawin, & Allan, 1988) and excellent long-term test-retest reliability ($r \geq .90$; McCaffrey et al., 1995) for intervals between 7 to 10 days. Practice effects have been found on re-testing (Strauss et al., 2006). Normative data are available from Mitrushina, Boone, & D'Elia (1999).

Trail Making Test, Part A & B

The Trail Making Test, Part A and B (Army Individual Test Battery, 1944) evaluates attention, speed, and set-shifting. Part A requires the participant to connect a series of 25 encircled numbers, spaced randomly on a page. In Part B, the participant must shift between numbers and letters, in alternating order. Practice exercises are included for both parts and scores are based on completion times.

Test-retest reliability varies with age ranges. However, Dikmen, Heaton, Grant & Temkin (1999) reported coefficients of .79 for Part A and .89 for Part B in a sample of healthy adults aged 15-83, after an eleven month interval. Normative data are available from Spreen and Strauss (1998).

Wisconsin Card Sort Test

The Wisconsin Card Sort Test (Heaton, 1981) is a measure used to assess set-shifting, strategic planning and response modulation. The test includes four stimulus cards (one red triangle, two green stars, three yellow crosses, four blue circles) and two response decks of 64 cards with designs that are similar to the stimulus cards, but vary in colour, number and geometric form. The participant is required to match each card in the response decks to one of the stimulus cards. They are then given feedback as to whether the match is right or wrong, and using this feedback they continue sorting the remaining response cards. The sorting rule changes without warning to the participant after every ten correct trials.

Scores consist of the number of correct responses and number of perseverative responses.

Reliability measures are variable in normal samples, but are higher in clinical populations (Strauss et al., 2006). Normative data are available from Heaton, Chelune, Talley, Kay, & Curtiss (1993).

Language

Boston Naming Test (BNT)

The Boston Naming Test (Goodglass, Kaplan, & Weintraub, 1983) is a measure of visual naming and consists of sixty black and white drawings of objects ranging from common (e.g. comb) to rare (e.g. abacus). Participants are required to name the object in each picture. If they cannot spontaneously answer within 20 seconds, cues are given, beginning with a description of the item (stimulus cue) and followed by the beginning sound of the object (phonemic cue). The score is comprised of the number of correct spontaneous and stimulus cued responses.

The alpha coefficient for internal consistency is high (.95; Franzen, Haut, Rankin, & Keefover, 1995) and test- retest reliability was found to be .95 over an eight month period (Sawrie, Chelune, Naugle, & Luders, 1996). Normative data are available from Heaton, Avitable, Grant, & Matthews (1999).

Controlled Oral Word Association Test (COWA)

The Controlled Oral Word Association Test (Spreen & Benton, 1969) measures the ability to spontaneously produce words in phonemic categories (F, A, S). Participants are given 60 seconds to generate as many words beginning with each letter as possible

(excluding proper nouns, word variations and repetitions). The score is the total number of words generated for all letters.

Tombough, Kozak, & Rees (1999) found high internal consistency (coefficient alpha $r = .83$; and good test-retest reliability (.74; Tombaugh et al., 1999). Normative data are available from Spreen & Strauss (1998).

Motor

Grooved Pegboard

The Grooved Pegboard test (Reitan & Wolfson, 1985) measures eye-hand coordination and motor speed. It consists of a metal board with 25 slotted holes. Participants are asked to place pegs into the holes as quickly as possible, using their dominant, then non-dominant hand. The score is the latency to insert the pegs into all holes.

Ruff and Parker (1993) found test-retest reliability of .76 and .78 over six months for dominant and non-dominant hands, respectively. Normative data can be found in Heaton, Grant, & Matthews (1991).

Processing speed

Digit Symbol Coding and Symbol Search

These two tests comprise the Processing Speed Index of the Wechsler Adult Intelligence Scale, 3rd Edition (Wechsler, 1997a). They measure the participant's ability to quickly process visual information. Digit Symbol requires participants to copy symbols as quickly as possible, whereas Symbol Search involves identifying target symbols. Each task

has a 120 second time limit. Scores are calculated by summing the number of correct responses.

The Processing Speed Index has a high internal reliability index with a split-half coefficient of .88 (Strauss et al., 2006). Reliability coefficients for test-retest are also high (.80 - .89; Strauss et al., 2006). Canadian and American age-stratified normative data are available in the test manual (Wechsler, 1997a).

Trail Making Test Part A

See Trail Making Test Part A & B under *Executive function* (pg. 132).

Verbal learning and memory

California Verbal Learning Test – II

The California Verbal Learning Test – II (Delis, Kramer, Kaplan, & Ober, 2000) measures verbal learning and memory. The test consists of a series of five trials where participants are asked to recall a word list containing 16 items in four categories. This is followed by the presentation of a second 16 item list that evaluates interference. The first list is elicited again to provide a measure of short-delay recall, and again after a 20 minute delay to assess long-delay recall. Cued and recognition trials are also included. Various scores are derived based on the total of the spontaneously recalled items across trials, cued recall, recognition recall, and retention.

Delis et al. (2000) found split-half reliability to be very high in both a healthy sample ($r = .94$) and a mixed clinical sample ($r = .96$). Test-retest reliability coefficients ranged from

high (.80 - .89) to adequate (.70 - .79) for the scores interpreted in our studies (Strauss et al., 2006). Normative data are available in the test manual (Delis et al., 2000).

Logical Memory

Logical Memory I and II are subtests of the Wechsler Memory Scale, 3rd Edition (Wechsler, 1997b) that measure immediate and delayed memory for verbal information. Two brief stories are read aloud to the participant, who is then asked to recall as many details as possible, both immediately and following a 20-30 minute delay.

Internal consistency coefficients were high for Logical Memory I (.80 - .89) and adequate for Logical Memory II (.70 - .79) and test-retest coefficients were adequate (.70 - .79) for both subtests (Strauss et al., 2006). Normative data are available in the test manual (Wechsler, 1997b).

Visual learning and memory

Rey Visual Design Learning Test

The Rey Visual Design Learning Test (Rey, 1968) is a measure of visual memory and learning. The participant is shown a series of 15 cards with geometric shapes for two seconds each, and is then asked to draw as many of them as possible, from memory. There are five trials, followed by a recognition task which requires the participant to identify the 15 designs from among 30. Recall and recognition are assessed after a 20-minute delay.

The test-retest coefficient for a sample of healthy adults was .45 over a one month interval (Spreeen & Strauss, 1991). Normative data are available from Spreeen and Strauss (1991).

Family Pictures

Family Pictures I and II are subtests of the Wechsler Memory Scale, 3rd Edition (Wechsler, 1997b) and are used to evaluate immediate and delayed memory for visual information. A series of four family scenes are presented one after another for ten seconds each. Participants are then asked to immediately recall details pertaining to the specific characters in each scene, their positions and their actions. This is repeated following a 20-30 minute delay.

Strauss et al. (2006) reported high internal consistency coefficients (.80 - .89) for both Family Pictures I and II, and marginal (.60 - .69) test-retest coefficients. As previously noted for Logical Memory I and II, normative data are available in the test manual (Wechsler, 1997b).

Visuospatial function

Block Design

Block Design is part of the Perceptual Organization Index of the Wechsler Adult Intelligence Scale, 3rd Edition (Wechsler, 1997a). This subtest was designed to measure visuo-construction abilities and non-verbal reasoning. The participant is given four to nine red and white blocks and then asked to construct patterns that match a series of pictures. Performance is in minutes and scores are derived from correct responses, with bonuses for speed.

Internal consistency and test-retest coefficients for Block Design were both high (.80 - .89; Strauss, 2006). As mentioned previously, normative data are available in the test manual (Wechsler, 1997a).

Working memory

Arithmetic, Digit Span, Letter-Number Sequencing

These subtests are part of the Working Memory Index of the Wechsler Adult Intelligence Scale, 3rd Edition (Wechsler, 1997a). This index is designed to evaluate the ability to hold verbal information in mind while manipulating it, then repeating it back to the examiner. The arithmetic subtest consists of a series of word problems that must be solved mentally within a time limit. For the Digit Span subtest, the participant is read a sequence of numbers and is asked to repeat them. A second series of trials requires the participant to repeat the sequence in reverse order. In Letter-Number Sequencing subtest, the participant is read a series of numbers and letters, and must repeat them back with the numbers first in ascending order, followed by the letters in alphabetical order.

Internal consistency for the Working Memory Index is very high (.90+) and the test-retest reliability coefficient is high (.80 - .89). Consistency coefficients for individual subtests are as follows: internal consistency for Digit Span (.90+) and Arithmetic and Letter-Number Sequencing (.80 - .89); test-retest for Digit Span and Arithmetic (.80 - .89) and Letter-Number Sequencing (.70 - .79; Strauss et al., 2006). Normative data are available in the test manual (Wechsler, 1997a).

Consonant Trigrams

Consonant Trigrams (or the Brown-Peterson Task; Brown, 1958) is a measure that evaluates divided attention and working memory. The participant is asked to recall a series of three consonants following an interference task (mental subtractions) performed for various intervals (0, 9, 18, and 36 seconds).

Internal consistency was found to be high for a Turkish version of the test (.85; Strauss et al., 2006). Normative data are available from Strauss et al. (2006).

Spatial Span

The Spatial Span subtest of the Wechsler Memory Scale, 3rd Edition (Wechsler, 1997b) was designed to measure the ability to hold visual information in mind while manipulating it. A board with three dimensional cubes is used to administer visual-spatial sequences which the participant is asked to duplicate. Sequences get progressively longer over trials, and are administered both forward and backward.

Internal consistency and test-retest coefficients were found to be adequate (.70 - .79; Strauss et al., 2006). Normative data are available in the test manual (Wechsler, 1997b).

Estimated premorbid IQ

Quick Test

The Quick Test (Ammons & Ammons, 1962) is a measure of receptive vocabulary that provides an IQ equivalency score. It was included in the battery in order to provide an estimate of pre-morbid cognitive function. Participants are presented with a word orally, and are asked to point to one of four pictures that most accurately represents the word. The score is obtained by summing the correct responses, and is then converted to a standard IQ score. The Quick Test has been found to correlate with the Full Scale IQ on the Wechsler Adult Intelligence Scale – Revised ($r = .64$; Wechsler, 1981; Traub & Spruill, 1982).

Psychological tests

Profile of Mood States

The Profile of Mood States (McNair et al., 1992) was designed to assess psychological distress and changes in mood over time. It contains six subscales including Anger-Hostility, Confusion-Bewilderment, Depression-Dejection, Fatigue-Inertia, Tension-Anxiety, and Vigor-Activity as well as a Total Mood Disturbance score. Participants are asked to use a 5-point Likert scale to rate their mood over the past 7 days.

Internal consistency was found to be very high, ranging from .87 to .95, and test-retest reliability over a median of 20 days ranged from .65 to .74 (McNair et al., 1992). Normative data are available in the test manual (McNair et al., 1992).

Beck Depression Inventory-II

The Beck Depression Inventory-II (Beck, Steer, & Brown, 1996) is designed to measure the presence and severity of depressive symptoms. It contains a series of 21 self-report items that reflect the criteria for depressive disorders in the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (American Psychiatric Association, 1994).

As reported by Strauss et al. (2006), internal consistency ranged from .84 - .93 and in a university sample, and $\geq .88$ in a psychiatric sample, and test-retest coefficients were found to be adequate (.74 - .75) to high (.93 - .96).

APPENDIX C

RCI SYNTAX

Syntax for RCI analyses using healthy control as comparison group.

USE ALL.

COMPUTE filter_\$=(int12 > 0 & group = 1).

VARIABLE LABEL filter_\$ 'int12 > 0 & group = 1 (FILTER)'.
 VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.

FORMAT filter_\$ (f1.0).

FILTER BY filter_\$.

EXECUTE .

DESCRIPTIVES

 VARIABLES=sspanr1

 /STATISTICS=STDDEV .

FILTER OFF.

USE ALL.

COMPUTE filter_\$=(int12 > 0).

VARIABLE LABEL filter_\$ 'Int12 > 0'.

VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.

FORMAT filter_\$ (f1.0).

FILTER BY filter_\$.

EXECUTE .

CORRELATIONS

 /VARIABLES=sspanr1 sspanr2

 /PRINT=TWOTAIL NOSIG

 /MISSING=PAIRWISE .

COMPUTE SEM1 = 2.28*(sqrt((1-0.631))).

VARIABLE LABEL SEM1 'Std error of measurement'.

EXECUTE .

COMPUTE SEDiff1 = sqrt((2*((SEM1)**2))).

VARIABLE LABELS SEDiff1 'Std error of difference'.

EXECUTE .

COMPUTE sspanr_d = sspanr2-sspanr1 .

VARIABLE LABELS sspanr_d 'sspanr_ difference between T2-T1' .

EXECUTE .

FILTER OFF.

USE ALL.

```

COMPUTE filter_$(int12 > 0 & group = 1).
VARIABLE LABEL filter_$(int12 > 0 & group = 1 (FILTER)'.
VALUE LABELS filter_$(0 'Not Selected' 1 'Selected').
FORMAT filter_$(f1.0).
FILTER BY filter_$.
EXECUTE .

```

```

DESCRIPTIVES
  VARIABLES=sspanr_d
  /STATISTICS=MEAN .

```

```

FILTER OFF.
USE ALL.
COMPUTE filter_$(int12 > 0).
VARIABLE LABEL filter_$(Int12 > 0)'.
VALUE LABELS filter_$(0 'Not Selected' 1 'Selected').
FORMAT filter_$(f1.0).
FILTER BY filter_$.
EXECUTE .

```

```

COMPUTE RCIsspanr_U = (SEdiff1)*(+1.64)+(-0.0357) .
VARIABLE LABELS RCIsspanr_U 'RCI sspanr_ Upper limit' .
EXECUTE .

```

```

COMPUTE RCIsspanr_L = (SEdiff1)*(-1.64)+(-0.0357) .
VARIABLE LABELS RCIsspanr_L 'RCI sspanr_ Lower limit' .
EXECUTE .

```

```

DO IF sspanr_d GE RCIsspanr_U.
COMPUTE RCIsspanr_ss = 1.
ELSE IF sspanr_d LE RCIsspanr_L.
COMPUTE RCIsspanr_ss = -1.
ELSE.
COMPUTE RCIsspanr_ss = 0.
END IF.
VARIABLE LABELS RCIsspanr_ss 'sspanr-1=under lower limit, 0=inside, 1=over upper
limit' .
EXECUTE .

```

```

COUNT
  sum_improved = RCIsspanr_ss to RCIGroond_ss (1) .
COUNT
  sum_expected = RCIsspanr_ss to RCIGroond_ss (0) .
COUNT
  sum_declined = RCIsspanr_ss to RCIGroond_ss (-1) .
EXECUTE .

```

FILTER OFF.

USE ALL.

compute final_improved1 = 1.

if (sum_improved eq 0 or sum_improved eq 1) final_improved1 = 0.

value labels final_improved1 0 "0 or 1 test" 1 "2 or more tests".

execute.

CROSSTABS

/TABLES=group BY final_improved1

/FORMAT= AVALUE TABLES

/STATISTIC=CHISQ

/CELLS= COUNT EXPECTED ASRESID

/COUNT ROUND CELL .

FILTER OFF.

USE ALL.

compute final_declined1 = 1.

if (sum_declined eq 0 or sum_declined eq 1) final_declined1 = 0.

value labels final_declined1 0 "0 or 1 test" 1 "2 or more tests".

execute.

CROSSTABS

/TABLES=group BY final_declined1

/FORMAT= AVALUE TABLES

/STATISTIC=CHISQ

/CELLS= COUNT EXPECTED ASRESID

/COUNT ROUND CELL .

FILTER OFF.

USE ALL.

compute final_improved2 = 0.

if (sum_improved eq 1) final_improved2 = 1.

if (sum_improved ge 2) final_improved2 = 2.

value labels final_improved2 0 "0 tests" 1 "1 test" 2 "2 or more tests".

execute.

CROSSTABS

/TABLES=group BY final_improved2

/FORMAT= AVALUE TABLES

/STATISTIC=CHISQ

/CELLS= COUNT EXPECTED ASRESID

/COUNT ROUND CELL .

FILTER OFF.

USE ALL.

```
compute final_declined2 = 0.  
if (sum_declined eq 1) final_declined2 = 1.  
if (sum_declined ge 2) final_declined2 = 2.  
value labels final_declined2 0 "0 tests" 1 "1 test" 2 "2 or more tests".  
execute.
```

CROSSTABS

```
/TABLES=group BY final_declined2  
/FORMAT= AVALUE TABLES  
/STATISTIC=CHISQ  
/CELLS= COUNT EXPECTED ASRESID  
/COUNT ROUND CELL .
```

APPENDIX D

SRB SYNTAX

Syntax for SRB analyses using healthy controls as comparison group.

USE ALL.

```
COMPUTE filter_$=(group = 1 & int12 > 0).
VARIABLE LABEL filter_$ 'group = 1 & int12 > 0 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
```

REGRESSION

```
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA CHANGE
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT sspanr2
/METHOD=ENTER age educat qt_raw1 /METHOD=ENTER sspanr1.
```

REGRESSION

```
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA CHANGE
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT sspanr2
/METHOD=ENTER sspanr1 /METHOD=ENTER age educat qt_raw1.
```

USE ALL.

```
COMPUTE filter_$=(int12 > 0).
VARIABLE LABEL filter_$ 'int12 > 0 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
```

```
COMPUTE sspanrPredSc = (.628*sspanr1) + (-.039*age) + (-.216*educat) + (.345*qt_raw1)
+ (-4.373) .
VARIABLE LABELS sspanrPredSc 'sspanr Predictor Score' .
EXECUTE .
```

```
COMPUTE sspanrSRBcs = (sspanr2-sspanrPredSc)/2.018 .
```

```
VARIABLE LABELS sspanrSRBcs 'sspanr SRB Change Score'.
EXECUTE .
```

```
DO IF (sspanrSRBcs GE 1.64).
COMPUTE SRBsspanr_ss = 1.
ELSE IF (sspanrSRBcs LE -1.64).
COMPUTE SRBsspanr_ss = -1.
ELSE .
COMPUTE SRBsspanr_ss = 0.
END IF.
```

```
VARIABLE LABELS SRBsspanr_ss 'sspanr-1=under lower limit, 0=inside, 1=over upper
limit'.
EXECUTE .
```

```
COUNT
  sum_improved = SRBsspanr_ss to SRBground_ss (1) .
COUNT
  sum_expected = SRBsspanr_ss to SRBground_ss (0) .
COUNT
  sum_declined = SRBsspanr_ss to SRBground_ss (-1) .
EXECUTE .
```

```
FILTER OFF.
USE ALL.
compute final_improved1 = 1.
if (sum_improved eq 0 or sum_improved eq 1) final_improved1 = 0.
value labels final_improved1 0 "0 or 1 test" 1 "2 or more tests".
execute.
```

```
CROSSTABS
  /TABLES=group BY final_improved1
  /FORMAT= AVALUE TABLES
  /STATISTIC=CHISQ
  /CELLS= COUNT EXPECTED ASRESID
  /COUNT ROUND CELL .
```

```
FILTER OFF.
USE ALL.
compute final_declined1 = 1.
if (sum_declined eq 0 or sum_declined eq 1) final_declined1 = 0.
value labels final_declined1 0 "0 or 1 test" 1 "2 or more tests".
execute.
```

```
CROSSTABS
  /TABLES=group BY final_declined1
  /FORMAT= AVALUE TABLES
```

```
/STATISTIC=CHISQ  
/CELLS= COUNT EXPECTED ASRESID  
/COUNT ROUND CELL .
```

FILTER OFF.

USE ALL.

```
compute final_improved2 = 0.
```

```
if (sum_improved eq 1) final_improved2 = 1.
```

```
if (sum_improved ge 2) final_improved2 = 2.
```

```
value labels final_improved2 0 "0 tests" 1 "1 test" 2 "2 or more tests".
```

execute.

CROSSTABS

```
/TABLES=group BY final_improved2
```

```
/FORMAT= AVALUE TABLES
```

```
/STATISTIC=CHISQ
```

```
/CELLS= COUNT EXPECTED ASRESID
```

```
/COUNT ROUND CELL .
```

FILTER OFF.

USE ALL.

```
compute final_declined2 = 0.
```

```
if (sum_declined eq 1) final_declined2 = 1.
```

```
if (sum_declined ge 2) final_declined2 = 2.
```

```
value labels final_declined2 0 "0 tests" 1 "1 test" 2 "2 or more tests".
```

execute.

CROSSTABS

```
/TABLES=group BY final_declined2
```

```
/FORMAT= AVALUE TABLES
```

```
/STATISTIC=CHISQ
```

```
/CELLS= COUNT EXPECTED ASRESID
```

```
/COUNT ROUND CELL .
```