

The Impact of Using Different Reference Populations on Measurement of Breast Cancer-Related Cognitive Impairment Rates

Jonathan D. Clapp^{1,2,*}, George Luta¹, Brent J. Small², Tim A. Ahles³, James C. Root³, Deena Graham⁴, Arti Hurria⁵, Paul B. Jacobsen⁶, Heather Jim⁷, Brenna C. McDonald⁸, Robert A. Stern⁹, Andrew J. Saykin⁸, Jeanne S. Mandelblatt¹, for the Thinking and Living with Cancer (TLC) Study

¹Georgetown-Lombardi Comprehensive Cancer Center, Washington, DC, USA

²University of South Florida, Tampa, FL, USA

³Memorial Sloan Kettering Cancer Center, New York, NY, USA

⁴Hackensack University Medical Center, Hackensack, NJ, USA

⁵City of Hope, Los Angeles, CA, USA

⁶National Cancer Institute, Division of Cancer Control & Population Sciences (this work was performed while Dr. Jacobsen was at the Moffitt Cancer Center and does not reflect the views of the National Cancer Institute)

⁷Moffitt Cancer Center, Tampa, FL, USA

⁸Indiana University, Indianapolis, IN, USA

⁹Alzheimer's Disease and CTE Center, Boston University School of Medicine, Boston, MA, USA

*Corresponding author at: Jonathan Clapp Georgetown-Lombardi Comprehensive Cancer Center, 3300 Whitehaven Street NW, Suite 4100, Washington, DC 20007, USA. Tel.: +1-202-687-2819; fax: +1-202-687-0305.

E-mail address: Jonathan.Clapp@georgetown.edu (J.D. Clapp).

Editorial Decision 14 December 2017; Accepted 7 February 2018

Abstract

Objective: To evaluate how use of different reference populations affects estimates of breast cancer-related cognitive impairment rates.

Methods: Patients aged ≥ 60 years with stage 0–3 breast cancer ($n = 371$) and matched non-cancer controls ($n = 370$) completed 13 neuropsychological tests prior to systemic therapy or at enrollment (controls). The tests captured three domains: attention, processing speed and executive function; learning and memory; and visual-spatial function. Domain-specific impairment was defined as having one test score $2SD$ below or two or more test scores $1.5SD$ below the reference population means. Different reference populations were used to define impairment: published normative data, study-specific controls, age and education-stratified controls, and age and education-adjusted controls. The associations between the resultant impairment rates and breast cancer (vs. control) were evaluated using chi-square tests and logistic regression models. Cohen's kappa coefficients were used to evaluate agreement of impairment rates between study-specific control performance and the other reference population groups.

Results: The patients and controls were aged 68.0 (SD 6.0) and 67.9 (SD 7.0) years, respectively. The association of breast cancer-control status with impairment did not differ based on reference group. Cognitive impairment based on published normative data yielded less agreement on impairment rates ($\kappa = 0.22$ – 0.89) than study-specific age and education-stratified control performance ($\kappa = 0.62$ – 1.00).

Conclusion: The choice of reference populations did not affect conclusions about the association of cognition with breast cancer. However, while study-specific reference populations provided greater internal consistency in defining cognitive impairment, benchmarking against published normative data will enhance the ability to compare results across studies.

Keywords: Cancer; Oncology; Cognition; Methods; Measurement

Introduction

Cancer-related cognitive impairment is increasingly recognized as a potential adverse effect of cancer and/or its systemic treatments, with the largest number of reports from breast cancer patients. (Mandelblatt, Jacobsen, & Ahles, 2014). Consensus

on measurement of cognitive impairment among breast and other cancer patients is evolving, but it is currently recommended that assessment should include self-report and neuropsychological testing of multiple domains of cognitive function, and comparison of results to a population without cancer (Wefel, Vardy, Ahles, & Schagen, 2011).

The choice of specific neuropsychological tests may vary across studies based on the domains of cognitive function posited to be most affected by the type of cancer and its treatments, as well as by individual experience of the investigators. The selected test results are often combined to create domain and global functioning scores and define impairment (Cordell et al., 2013; Wefel et al., 2011). However, there is not one standard approach for estimating the cancer study sample impairment relative to reference populations. Moreover, rates of impairment may be biased if the characteristics of the reference populations for each test are not comparable to each other and/or to the cancer study sample (Heaton et al., 2001; Miller & Rohling, 2001; Russell, Russell, & Hill, 2005). To address this possible source of bias, researchers can try to match their study sample to published normative test scores based on demographic factors like gender, age, and educational level (Joly, Rigal, Noal, & Giffard, 2011; Wefel et al., 2004). However, bias could remain if study testing was conducted in a different manner from that employed when testing the published normative population. To avoid some of these difficulties, it has been suggested that study-specific controls be recruited to match the cancer cases under investigation.

Variations in approach to these pragmatic design and measurement issues can lead to differences across studies in rates of impairment, limiting comparisons, and synthesis of results. There is also the potential to come to varying conclusions about how cancer and its therapies affect cognition when cases from different studies are compared to varying types of non-cancer groups.

Estimating rates of cancer-related cognitive impairment may be especially important in older breast cancer patients who have high rates of disease and may already be at high risk for cognitive decline based on aging and comorbidities (Mandelblatt et al., 2014). In a prior methodological study, Schilder et al. used data from a sample of young breast cancer cases and age-matched controls and reported that cognitive impairment rates estimated based on demographically matched published normative data differed from those estimated using the matched control group (Schilder et al., 2010). That study showed that regardless of the classification threshold, the magnitude of difference in impairment rates between breast cancer cases and controls remained similar across the reference groups. However, that study focused on published normative data for Dutch women, so that results may vary in studies with different methods, populations, and older age groups.

In a previous report, we used overall study-specific control means to evaluate factors related to pre-treatment impairment among patients ages 60 and older (“older”) with breast cancer enrolled in the *Thinking and Living with Cancer* (TLC) study (Mandelblatt et al., 2014). In this methodological paper, we compare the rates of cognitive impairment prior to systemic therapy using four different reference groups and a single, standard definition of impairment. We also estimate the concordance in rates as defined by the reference groups and the magnitude of association between cancer and cognitive impairment by choice of reference group. The results are intended to provide data to inform discussion about guidelines for future studies of cancer-related cognitive impairment and enhance comparison of results across studies. Ultimately, advancements in measurement of impairment could facilitate routine assessment of cognitive problems in clinical practice with cancer patients.

Methods

Setting and Population

The *Thinking and Living with Cancer* (TLC) study is a multi-site, longitudinal study of risk factors for cognitive decline among older patients with breast cancer and age, race, education, and site-frequency-matched controls. Prior reports included 164 patient cases and 182 controls recruited between August 1, 2010 and June 30, 2013 (Mandelblatt et al., 2014). This report includes participants enrolled through August 8, 2016; enrollment and follow-up remain ongoing. Baseline pre-treatment (or enrollment for controls) data were used for this methodological study.

The study sites included Georgetown University in Washington, DC; Memorial Sloan-Kettering Cancer Center in New York, New York; Moffitt Cancer Center in Tampa, Florida; City of Hope Comprehensive Cancer Center in Duarte, California; Hackensack University Medical Center in Hackensack, New Jersey; Indiana University in Indianapolis, Indiana. Neuropsychological test assessment training and certification was conducted at Boston University School of Medicine and Memorial Sloan-Kettering Cancer Center (<https://tlcstudy.mskcc.org>). The research protocol met HIPAA standards and was approved by all Institutional Review Boards.

Eligible cases were 60 years or older, newly diagnosed with primary non-metastatic breast cancer (AJCC six stage 0–3), and English-speaking; those with stroke, head injury, major mental illness, and neurodegenerative disorders were ineligible. Participants with a history of other cancers were excluded if active treatment was <5 years prior or they ever had

chemotherapy or hormonal therapy. Controls met the same eligibility criteria as cases except for cancer diagnosis. Eligible controls were recruited from friends of participating cases; if no friend was provided, community controls were frequency-matched to the cases with respect to age, race, education, and recruitment site.

Among eligible participants, 388 cases (38%) and 377 controls (95%) consented in the study period for the current analysis (Fig. 1). Consenting participants were screened for final eligibility using the Mini-Mental State Examination (MMSE) and the Wide Range Achievement Test, fourth edition (WRAT-4) to ensure ability to complete the study; those with MMSE scores of <24 or WRAT-4 <third grade were ineligible (11 cases; four controls). There were 371 cases and 370 controls in the final analytic sample.

Data Collection

Baseline neuropsychological testing and interviews were administered after post-operative visits, but before radiation or systemic therapy for cases or at enrollment for controls. Thirteen tests were selected to capture three domains previously reported as associated with cancer and its treatments: attention, processing speed and executive function; learning and memory; and visual-spatial function (Wefel et al., 2011). The specific tests were chosen based on past use in older populations, inclusion in the National Alzheimer's Coordinating Center's Unified Data Set (Weintraub et al., 2009), and recommendations from the International Cancer and Cognition Task Force (Wefel et al., 2011). The tests were grouped into domains based on a priori expectations and the results of principal components factor analysis (see Supplementary material online, Table 1) (Mandelblatt et al., 2014). The interview ascertained age and education.

Outcomes

The outcomes were cognitive impairment rates for the three domains. Sensitivity analyses examined the effects of reference group on case–control differences in individual tests within the domains. Domain-level impairment for a participant was defined as having one test within the domain that was $\geq 2 SD$ or two or more tests $\geq 1.5 SD$ lower compared to the reference group mean and standard deviation (Wefel et al., 2011). Individual test-level impairment for the participant was defined as having a score $\geq 1.5 SD$ lower compared to the reference group mean and standard deviation.

The four reference groups selected for this analysis were based upon published normative data and three different ways to consider study-specific control performance. The published normative data was used because the results had been developed and previously validated in samples of non-impaired older participants that mimicked the demographics of the current sample. The published normative data include adjustments for age and education and are shown in Supplementary material online, Table 1 (Mitrushina, Boone, & Razani, 2005; Shirk et al., 2011; Stern & White, 2003; Weintraub et al., 2009). The study-specific control performance approaches were selected since the TLC study focused solely on older women who were very well educated and may not have matched the participants included in developing the population normative data. Hence, TLC recruited matched non-breast cancer controls to allow for an internal comparison of participants receiving the same neuropsychological tests in the same intervals as the cases. The TLC study-specific controls provide three of the reference groups:

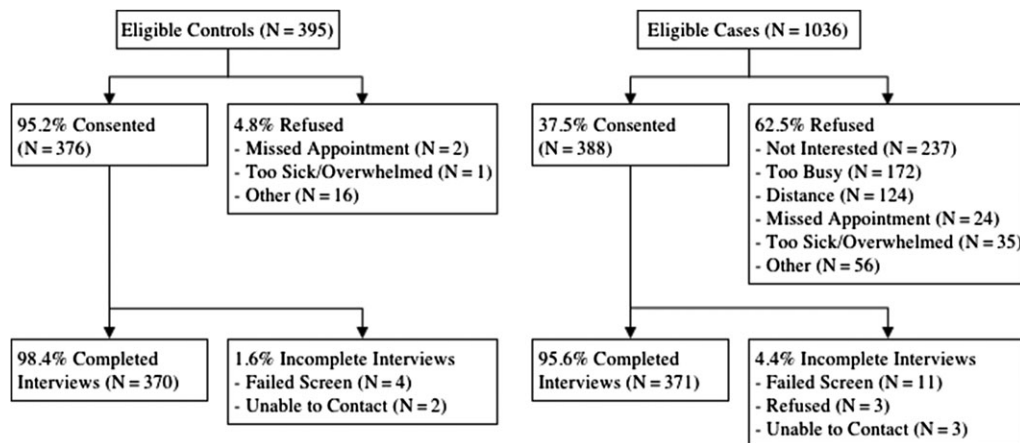


Fig 1. Selection of the study sample of older breast cancer patients and matched controls.

overall control, age and education-stratified control, and age and education-adjusted control. The first study-specific control approach defined impairment based on the difference between the participants' score and the overall control group mean score and SD. This approach served as the comparator for subsequent analyses because of its simplicity and its prior use with TLC results (Mandelblatt et al., 2014). The second study-specific control approach used the difference between the participant's score and the corresponding age- and education-stratified mean and SD for the study-specific control group. The third study-specific control approach obtained age and education coefficients from regression models performed on the control group test scores, which provided predicted scores when applied to all of the participants; the residual score, or the difference from the participant's predicted and actual scores, was the adjusted score analyzed by this approach. The study-specific age- and education-stratified and adjusted approaches were chosen to reflect the structure of the age and education adjustments that are present in the published normative data.

Covariates

Covariates included those used in the application of the reference group data, including age (continuous in years or grouped as 60–64, 65–69, and 70+ for stratification) and education (continuous in years or grouped as <high school vs. some college+ for stratification). Race and ethnicity were not used for this analysis since not all published normative data included race- and ethnicity-specific reference data.

Statistical Methods

Two approaches were utilized in order to explore the impact of the choice of reference group on conclusions about the association between impairment and case–control status. First, chi-square tests and logistic regression models were performed to examine the association of domain-level impairment with breast cancer (vs. control). Since the domains included some tests developed using different reference populations, we also conducted a sensitivity analysis to examine if conclusions about the association of impairment and case–control status varied at the test level.

Second, Cohen's kappa coefficient was used in order to assess agreement for domain-level impairment between the definitions based on different reference groups. We used the study-specific overall control reference values for comparison with the three other referent groups for these analyses. The level of agreement by the Cohen's kappa coefficient was interpreted as <0 indicating no agreement, 0–0.20 slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, and 0.81–1.00 excellent agreement (Landis & Koch, 1977). All analyses were completed using SAS software (version 9.4; SAS Institute, Cary, NC).

Results

The older breast cancer patients included were comparable to matched controls, with regard to age, education, WRAT-4, race, and ethnicity (Table 1). Most were White, and well educated. When age and years of education are categorized, the control group tends to be younger ($p = .04$) and better educated than patients ($p = .09$). At baseline (enrollment), there were non-significant variations in domain-level impairment rates across reference groups (Table 2). The rates of test-based impairment were consistent with the domain-level results, with only three of the 13 individual tests showing any significant differences by reference group (i.e., 5% of results, and potentially due to chance given the large number of comparisons) (Table 2). When comparing the unadjusted odds of impairment for breast cancer cases with the odds for non-cancer controls, the study-specific overall control method showed higher odds ratios for each of the domains when compared to the other methods (see Supplementary material online, Table 2), but there were no significant differences in odds of impairment between cases and controls for any method.

Using the study-specific overall control group as the comparator, the highest overall agreement of domain-level cognitive impairment rates among breast cancer cases was seen when using the study-specific age and education-stratified reference group, while the least agreement was seen when using the published normative data reference group (Table 3). There was slightly greater agreement with the comparator for the study-specific age and education-stratified method than for the study-specific age and education-adjusted method. These same patterns were seen among the controls (data not shown).

Table 1. Study sample of older breast cancer patients and matched controls

		All participants	Control	Breast cancer patients (Case)	<i>p</i> -Value ¹
Number of participants		741 (100%)	370 (50%)	371 (50%)	
Age group ²	60–64	262 (35.4)	147 (39.7)	115 (31.0)	.04
	65–69	234 (31.6)	112 (30.3)	122 (32.9)	
	70+	245 (33.1)	111 (30.0)	134 (36.1)	
Age ²	Mean (<i>SD</i>)	68.0 (6.5)	67.9 (7.0)	68.0 (6.0)	.72
	Range	(60–98)	(60–91)	(60–98)	
Education (years) ²	Mean (<i>SD</i>)	15.7 (2.4)	15.7 (2.4)	15.6 (2.4)	.59
	Range	(6–20)	(6–20)	(9–20)	
Education ²	HS grad or Less	102 (13.8)	43 (11.6)	59 (15.9)	.09
	Some college+	639 (86.2)	327 (88.4)	312 (84.1)	
WRAT 4 Standardized Score	Mean (<i>SD</i>)	111.3 (15.8)	111.7 (16.1)	111.0 (15.4)	.56
	Range	(64–145)	(74–145)	(64–145)	
MMSE ³	Mean (<i>SD</i>)	29.3 (1.1)	29.4 (1.0)	29.2 (1.2)	N/A
	Range	(24–30)	(24–30)	(24–30)	
Race (white vs. non-white)	Other	144 (19.6)	74 (20.1)	70 (19.0)	.71
	White, non-Hispanic	592 (80.4)	294 (79.9)	298 (81.0)	
Hispanic or Latino	No	684 (92.6)	339 (92.1)	345 (93.0)	.65
	Yes	55 (7.4)	29 (7.9)	26 (7.0)	
Surgery type	BCS			197 (55.2)	
	Mastectomy			156 (43.7)	
	No surgery			4 (1.1)	
Stage	0 or 1			240 (67.4)	
	2 or 3			116 (32.6)	
Treatment type	Chemo (\pm Hormonal)			97 (28.0)	
	Hormonal only			249 (72.0)	

¹*p*-Values are from chi-square or *T*-tests.

²Mean age and education are comparable. Age and education groups are presented since these were the strata used for age- and education-stratified reference values. Cases and controls vary somewhat when data are grouped into these categories.

³The MMSE was used only as an eligibility screening tool. Participants with scores <24 were not eligible for the study, hence the score range is truncated.

Table 2. Cognitive impairment rates among older breast cancer patients and matched controls by estimation method

	Study-specific overall control		Published normative data		Age & education-stratified control		Age & education-adjusted control	
	Case	Control	Case	Control	Case	Control	Case	Control
APE Domain	17.8%	15.1%	19.1%	16.5%	19.9%	20.0%	21.3%	19.5%
NAB Digit Span Forward	3.6%	3.9%	7.9%	9.4%	7.7%	6.1%	6.6%	6.6%
NAB Digit Span Backward	6.3%	4.2%	11.2%	11.4%	6.3%	5.3%	7.2%	5.8%
COWAT	10.1%	5.9%	6.9%	4.2%	10.1%	7.2%	9.9%	6.6%
Trail Making A	11.2%	8.3%	11.3%	9.8%	9.0%	9.1%	10.5%	9.1%
Trail Making B	11.5%	7.4%	7.7%	6.8%	10.7%	12.2%	10.2%	11.9%
Digit Symbol Test-Coding	11.8%	7.1%	1.7%	0.3%	11.2%	9.4%	12.4%	8.3%
Learning & Memory Domain	17.3%	12.4%	14.6%	15.9%	15.9%	16.2%	16.4%	16.2%
Logical Memory IA	7.1%	7.7%	4.4%	3.9%	10.4%	9.4%	8.5%	8.3%
Logical Memory IIA	11.0%	8.6%	3.9%	3.3%	9.6%	10.0%	9.4%	9.4%
NAB List Learning Immediate Recall	11.2%	8.6%	9.6%	9.4%	8.8%	8.6%	12.1%	10.5%
NAB List Learning Short Delay	14.5%	10.4%	12.9%	13.0%	7.9%	8.6%	10.2%	8.6%
NAB List Learning Long Delay	6.0%	6.5%	7.9%	11.1%	7.9%	8.6%	8.0%	8.9%
Visuospatial Domain	8.6%	7.8%	11.6%	14.1%	9.7%	9.7%	7.0%	8.1%
NAB Figure Drawing Copy	7.1%	8.0%	13.5%	19.7%	6.3%	8.3%	5.8%	7.5%
NAB Figure Drawing Organization	10.4%	10.7%	9.1%	11.9%	9.1%	10.5%	8.6%	11.4%

Impairment rates for the individual tests are based upon scores ≤ -1.5 *SD* from expected mean of the reference population group. Impairment rates for the domains are based upon having any one test ≤ -2 *SD* or two or more tests ≤ -1.5 *SD* from the expected mean of the reference group for the respective test.

Impairment rates with case–control differences at the 0.05 significance level are bolded. Note that with the multiple comparisons, 5% of tests (i.e., three tests of 13 compared in four manners) are expected to be significant by chance alone.

Table 3. Agreement of study-specific overall control cognitive impairment rates with alternative methods to estimate impairment rates for older breast cancer cases

Domain or test	Published normative data	Age & education-stratified control	Age & education-adjusted control
APE Domain	0.74 (0.65–0.83)	0.77 (0.69–0.86)	0.70 (0.61–0.79)
NAB Digit Span Forward	0.55 (0.37–0.73)	0.62 (0.44–0.79)	0.46 (0.26–0.66)
NAB Digit Span Backward	0.69 (0.56–0.83)	1.00 (1.00–1.00)	0.72 (0.57–0.86)
COWAT	0.79 (0.67–0.90)	0.91 (0.84–0.98)	0.65 (0.52–0.78)
Trail Making A	0.82 (0.72–0.91)	0.73 (0.61–0.85)	0.89 (0.81–0.96)
Trail Making B	0.79 (0.68–0.90)	0.71 (0.59–0.83)	0.86 (0.77–0.94)
Digit Symbol Test-Coding	0.22 (0.07–0.37)	0.76 (0.65–0.86)	0.69 (0.57–0.81)
Learning & Memory Domain	0.62 (0.51–0.73)	0.78 (0.69–0.86)	0.76 (0.67–0.85)
Logical Memory IA	0.75 (0.60–0.90)	0.69 (0.56–0.83)	0.64 (0.49–0.79)
Logical Memory IIA	0.49 (0.33–0.65)	0.93 (0.86–0.99)	0.79 (0.68–0.90)
NAB List Learning Immediate Recall	0.62 (0.48–0.75)	0.80 (0.70–0.91)	0.56 (0.43–0.69)
NAB List Learning Short Delay	0.68 (0.56–0.79)	0.67 (0.55–0.79)	0.70 (0.58–0.81)
NAB List Learning Long Delay	0.68 (0.53–0.83)	0.64 (0.48–0.80)	0.73 (0.58–0.87)
Visuospatial Domain	0.78 (0.67–0.89)	0.81 (0.70–0.91)	0.70 (0.56–0.84)
NAB Figure Drawing Copy	0.60 (0.47–0.74)	0.89 (0.80–0.99)	0.75 (0.61–0.89)
NAB Figure Drawing Organization	0.89 (0.81–0.97)	0.83 (0.73–0.93)	0.82 (0.72–0.93)

No Agreement: <0; Slight Agreement: 0–0.20; Fair Agreement: 0.21–0.40; Moderate Agreement: 0.41–0.60; Substantial Agreement: 0.61–0.80 (underlined); Almost Perfect Agreement: 0.81–1.00 (bold).

Discussion

Understanding the impact of breast cancer and its therapies on the cognitive function of older breast cancer patients is important for treatment decision-making and survivorship care planning. This study examined how the choice of the reference population for a battery of neuropsychological tests might affect conclusions about cancer-related cognitive impairment rates. It was reassuring that impairment rates and the associations between case–control status and impairment were similar regardless of whether published normative data or study-specific control reference values were used to define impairment. However, there was variability in agreement between methods, with study-specific control age- and education-stratified reference values having the highest, and published population normative data the lowest concordance with overall study-specific control values.

Most studies define impairment based on a threshold level of deviation from the mean of a normal, non-cancer referent population. The rates of pre-treatment impairment observed in this were based on usual thresholds of having results 1.5 or 2 *SD* from the mean. Using this method, impairment rates were similar to or somewhat higher in this older population than reported in past studies with younger breast cancer populations using either study-specific (Ahles et al., 2008; Schagen, Muller, Boogerd, Mellenbergh, & van Dam, 2006) or population-based reference groups (Wefel et al., 2004). The differences in study populations, reference groups, and definition of impairment are likely to account for some of the variation in rates of cancer-related cognitive impairment between the prior reports and current study (Schilder et al., 2010).

This study illustrates that if the same reference population is used to define impairment in both cases and controls, it does not appear to affect associations of case–control status and impairment. However, since the data used in this study were from a pre-treatment sample, where low rates of fewer differences in impairment are expected, it will be important to repeat this evaluation using post-treatment data, where greater impairment is expected.

This study was based on data from a prospective study of older women, where controls were selected to minimize socio-demographic or clinical differences with breast cancer cases that could affect test scores. Therefore, it is logical that using the study-specific control group as the reference group produced better agreement than the published normative reference group, which may not have had a sufficiently large sample of older individuals or may have systematically differed from the study sample in other characteristics related to impairment (e.g., comorbidity). Additionally, while the cases and study-specific controls had similar mean ages and educational levels, there were differences when these data were categorized. In situations like this, the study-specific control stratified reference values would be recommended.

Study-specific non-cancer control referent data have several advantages over external population reference values, including the ability to create comparable groups who have undergone the same tests in the same manner, evaluate the impact of specific covariates on cognitive impairment, and combine tests into domains even when the tests were derived in varying populations. Since the purpose of the TLC study is to examine the prospective effects of cancer and its treatments over and above aging, the study design included a study-specific matched control group. However, using a study-specific non-cancer

control group is not always feasible, and requires resources to enroll a sufficiently large sample to detect small group differences in impairment.

Key advantages of using published normative data are that it increases study feasibility and can increase comparability across studies. Thus, researchers might consider reporting impairment data from a comparable internal, study-specific control group and a published population reference group. However, it is important to consider that published normative-based test reference groups can bias conclusions about rates of impairment if the original populations used to develop the normative scores are not similar to the study population.

There are some caveats that should be considered in evaluating our results. First, we did not consider the effects of the number of tests within a battery on sensitivity of results with regards to the choice of reference group (Schilder et al., 2010). Next, the age and education strata that were used to select study-specific control reference values were not identical to those published for the individual neuropsychological tests. Rather, they were selected to have a sufficiently large sample within each stratum to provide stable reference values for the mean and standard deviation. Additionally, regardless of method, the threshold used to define impairment based on deviation from mean scores is well below the cut-points for a clinical diagnosis of mild cognitive impairment (Albert et al., 2011). Finally, this analysis was conducted using pre-treatment baseline data. It will be important to extend the results to evaluate the impact of reference groups on estimating prospective change over time that exceeds normative expectations.

This study was novel in evaluating breast cancer-related impairment rates using multiple referent populations and isolating their effects by utilizing a single definition of impairment. The results illustrate that study-specific reference populations provide greater internal consistency in defining cognitive impairment, but that benchmarking results against published normative data may enhance the ability to compare results across studies. The choice of reference populations will ultimately depend on the specific research questions and pragmatic considerations.

Supplementary Material

Supplementary material is available at *Archives of Clinical Neuropsychology* online.

Funding

This research was supported by the National Cancer Institute at the National Institutes of Health grant # R01CA129769. The research was also supported, in part, by NCI Grants #R35CA197289 to J.S.M., the Biostatistics and Bioinformatics Shared Resources at Georgetown-Lombardi Comprehensive Cancer Center funded by the National Cancer Institute at the National Institutes of Health under grant #P30CA051008, and Ahles and Root under the MSK Core Grant #P30CA008748. Earlier versions of this research were presented at the International Cancer and Cognition Task Force Annual Meeting, March 2016.

Conflict of interest

None declared.

References

- Ahles, T. A., Saykin, A. J., McDonald, B. C., Furstenberg, C. T., Cole, B. F., Hanscom, B. S., et al. (2008). Cognitive function in breast cancer patients prior to adjuvant treatment. *Breast Cancer Research and Treatment*, *110*, 143. doi:10.1007/s10549-007-9686-5.
- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *7*, 270–279. doi:10.1016/j.jalz.2011.03.008.
- Cordell, C. B., Borson, S., Boustani, M., Chodosh, J., Reuben, D., & Verghese, J.,... Medicare Detection of Cognitive Impairment, W. (2013). Alzheimer's Association recommendations for operationalizing the detection of cognitive impairment during the Medicare Annual Wellness Visit in a primary care setting. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *9*, 141. doi:10.1016/j.jalz.2012.09.011.
- Heaton, R. K., Gladsjo, J. A., Palmer, B. W., Kuck, J., Marcotte, T. D., & Jeste, D. V. (2001). Stability and course of neuropsychological deficits in schizophrenia. *Archives of General Psychiatry*, *58*, 24. doi:10.1093/ajph.101.11.1903.
- Joly, F., Rigal, O., Noal, S., & Giffard, B. (2011). Cognitive dysfunction and cancer: Which consequences in terms of disease management? *Psycho-oncology*, *20*, 1251–1258. doi:10.1002/pon.1903.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159.

- Mandelblatt, J. S., Jacobsen, P. B., & Ahles, T. (2014a). Cognitive effects of cancer systemic therapy: Implications for the care of older patients and survivors. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 32, 2617–2626. doi:10.1200/jco.2014.55.1259.
- Mandelblatt, J. S., Stern, R. A., Luta, G., McGuckin, M., Clapp, J. D., Hurria, A., et al. (2014b). Cognitive impairment in older patients with breast cancer before systemic therapy: Is there an interaction between cancer and comorbidity? *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 32, 1909. doi:10.1200/JCO.2013.54.2050.
- Miller, L. S., & Rohling, M. L. (2001). A statistical interpretive method for neuropsychological test data. *Neuropsychology Review*, 11, 143.
- Mitrushina, M., Boone, K. B., & Razani, J. (2005). *Handbook of normative data for neuropsychological assessment (2)*. New York: Oxford University Press.
- Russell, E. W., Russell, S. L., & Hill, B. D. (2005). The fundamental psychometric status of neuropsychological batteries. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 20, 785. doi:S0887-6177(05)00070-3[pii].
- Schagen, S. B., Muller, M. J., Boogerd, W., Mellenbergh, G. J., & van Dam, F. S. (2006). Change in cognitive function after chemotherapy: A prospective longitudinal study in breast cancer patients. *Journal of the National Cancer Institute*, 98, 1742. doi:98/23/1742[pii].
- Schilder, C. M., Seynaeve, C., Linn, S. C., Boogerd, W., Gundy, C. M., Beex, L. V., et al. (2010). The impact of different definitions and reference groups on the prevalence of cognitive impairment: A study in postmenopausal breast cancer patients before the start of adjuvant systemic therapy. *Psycho-oncology*, 19, 415. doi:10.1002/pon.1595.
- Shirk, S. D., Mitchell, M. B., Shaughnessy, L. W., Sherman, J. C., Locascio, J. J., Weintraub, S., et al. (2011). A web-based normative calculator for the uniform data set (UDS) neuropsychological test battery. *Alzheimer's Research & Therapy*, 3, 32. doi:10.1186/alzrt94.
- Stern, R. A., & White, T. (2003). Neuropsychological Assessment Battery[®] (NAB[®]). Retrieved from <http://www4.parinc.com/Products/Product.aspx?ProductID=NAB>
- Wefel, J. S., Lenzi, R., Theriault, R., Buzdar, A. U., Cruickshank, S., & Meyers, C. A. (2004). 'Chemobrain' in breast carcinoma? A prologue. *Cancer*, 101, 466. doi:10.1002/cncr.20393.
- Wefel, J. S., Vardy, J., Ahles, T., & Schagen, S. B. (2011). International Cognition and Cancer Task Force recommendations to harmonise studies of cognitive function in patients with cancer. *The Lancet. Oncology*, 12, 703. doi:10.1016/S1470-2045(10)70294-1.
- Weintraub, S., Salmon, D., Mercaldo, N., Ferris, S., Graff-Radford, N. R., Chui, H., et al. (2009). The Alzheimer's Disease Centers' Uniform Data Set (UDS): The neuropsychologic test battery. *Alzheimer Disease and Associated Disorders*, 23, 91. doi:10.1097/WAD.0b013e318191c7dd.