Neuropsychology
2014, Vol. 28, No. 5, 726–740

0894-4105/14/$12.00 DOI: 10.1037/neu0000064

A Meta-Analysis of Cognitive Impairment Following Adult Cancer Chemotherapy

Oana C. Lindner
University of Manchester

Bob Phillips
University of York

Martin G. McCabe, Andrew Mayes, Alison Wearden, Filippo Varese, and Deborah Talmi
University of Manchester

Objective: Chemotherapy-induced cognitive impairments are reported by many cancer survivors. Research to date has not provided a clear description of their nature, extent, mechanisms, and duration. To investigate the impairments and factors that could influence their identification and severity, the present meta-analysis brings together research on this topic in adult cancer patients. ***Method:*** Our random-model meta-analysis includes 44 studies investigating the cognitive performance of adults treated with chemotherapy for non-central nervous system malignancies, primarily breast and testicular cancer. We conducted several subgroup analyses to identify the level of cognitive impairments in longitudinal and cross-sectional studies. We also pursued several multilevel model regressions to investigate the impact of methodological (study quality) and clinical moderators (diagnosis, age, time since treatment) on the observed effect sizes. ***Results:*** Cognitive impairments were found in cross-sectional studies in immediate free recall, delayed memory, verbal memory, delayed recognition memory, selective attention, and attention capacity. Surprisingly, prior to chemotherapy, patients performed better than matched controls. In longitudinal studies, patients' performance increased from baseline to follow-up, an effect that was stronger in patients than controls. None of the chosen moderators influenced the magnitude of estimated summary effect sizes. ***Conclusions:*** The likelihood to identify impairments rests on the type of design employed, as memory and attention impairments are only detected in cross-sectional studies. We discuss the lack of significant impact of moderators on the effect sizes despite the heterogeneity of results, while providing recommendations toward decreasing the heterogeneity in future studies.

Keywords: cancer, chemotherapy, impairment, neuropsychology, chemo-brain

Supplemental materials: http://dx.doi.org/10.1037/neu0000064.supp

Cognitive impairments may contribute to a lower quality of life following cancer diagnosis and treatment (Short, Vasey, & Tunceli, 2005). Despite this, health care systems in many countries do not have the appropriate resources to help people cope with chemotherapy-induced cognitive impairments (Ferguson et al., 2007). Furthermore, the continuous care needs of survivors, in terms of their cognitive deficits, and how these might relate to potentially higher distress levels and lower quality of life are yet to be identified. This is a problem given the increasing number of people living with and beyond cancer (Maddams et al., 2009). The limited knowledge regarding this phenomenon may be a result of several factors.

First, chemotherapy-induced cognitive changes in adult patients do not have a long research history. Consequently, guidelines for conducting neuropsychological research with former adult patients were proposed only recently by the International Cancer and Cognition Taskforce (ICCTF) (Vardy, Wefel, Ahles, Tannock, & Schagen, 2008). Second, research in this area may have possibly been hampered by inconsistencies in previous findings regarding cognitive impairments. On the one hand, there are differences between the degrees of objective impairment reported by different studies, being found in 12% to 68% of cancer survivors (Ahles & Saykin, 2007; Shilling, Jenkins, & Trapala, 2006). On the other hand, subjective impairments are reported by up to 80% of these patients (Kohli et al., 2006). Inconsistencies in the percentage and types of impairments reported by the literature yielded some

This article was published Online First March 17, 2014.

Oana C. Lindner, School of Psychological Sciences, University of Manchester, Manchester, United Kingdom; Bob Phillips, Centre for Reviews and Disseminations, University of York, York, United Kingdom; Martin G. McCabe, Institute of Cancer Sciences, University of Manchester; Andrew Mayes, Alison Wearden, Filippo Varese, and Deborah Talmi, School of Psychological Sciences, University of Manchester.

This work is part of a project funded by the Medical Research Council. We would like to acknowledge Dr. Linda McGowan and Dr. Dimitris Tsivilis for the support and advice in conducting this meta-analysis, Diana Foltean for her help in the quality assessment of the studies, the authors who have supplied data for this meta-analysis, and the Medical Research Council of the United Kingdom for funding this work.

Correspondence concerning this article should be addressed to Oana C. Lindner, Zochonis Building, Brunswick Street, School of Psychological Sciences, University of Manchester, Manchester, M13 9PL, UK. E-mail: oana.mereuta@postgrad.manchester.ac.uk

uncertainty about which functions are impaired, and may have reduced the emphasis on evidence-based intervention strategies to help patients overcome these problems (Ferguson et al., 2007).

Previous reviews have claimed that the lack of cohesion within the literature might stem from variability in several factors, including participant demographics such as age and gender (Ahles et al., 2003), treatment protocols (Freeman & Broshek, 2002; Hurria, Somlo, & Ahles, 2007; Jansen, Miaskowski, Dodd, Dowling, & Kramer, 2005), and variability in the neuropsychological tests used in assessments (Hutchinson, Hosking, Kichenadasse, Mattiske, & Wilson, 2012; Jansen, Miaskowski, Dodd, & Dowling, 2007). Additional confounding may be due to limited consideration of practice effects (Vardy et al., 2008), whether patients are compared to test norms or matched control participants, and whether the matching is solely on age and gender or also on intelligence and education, (Anderson-Hanley, Sherman, Riggs, Agocha, & Compas, 2003; Jansen et al., 2005), and differences in the statistical cut-offs for defining the impairments (Hurria et al., 2007; Wefel, Vardy, Ahles, & Schagen, 2011). However, other than the impact of comparing patient results to norms or controls, most of these potential moderators have not been systematically investigated in relation to the degree of observed impairments. We will further give a brief account of the results obtained by previous literature that sought to identify the nature and extent of impairments.

Conclusions From Previous Meta-Analyses

Four previous meta-analyses have summarized the cognitive outcomes of earlier studies (Anderson-Hanley et al., 2003; Falleti, Sanfilippo, Maruff, Weih, & Phillips, 2005; Jansen et al., 2005; Stewart, Bielajew, Collins, Parkinson, & Tomiak, 2006). Relative to controls, survivors exhibited a broad range of mild to moderate cognitive deficits in attention, information processing, verbal and visual, long-term and working memory, spatial skills, language, executive and motor functioning, summarized in Figure 1. While the four analyses agreed on the direction of the effects, there was less agreement on their magnitude, despite analyzing approximately the same literature; for example, the effect size of speed of

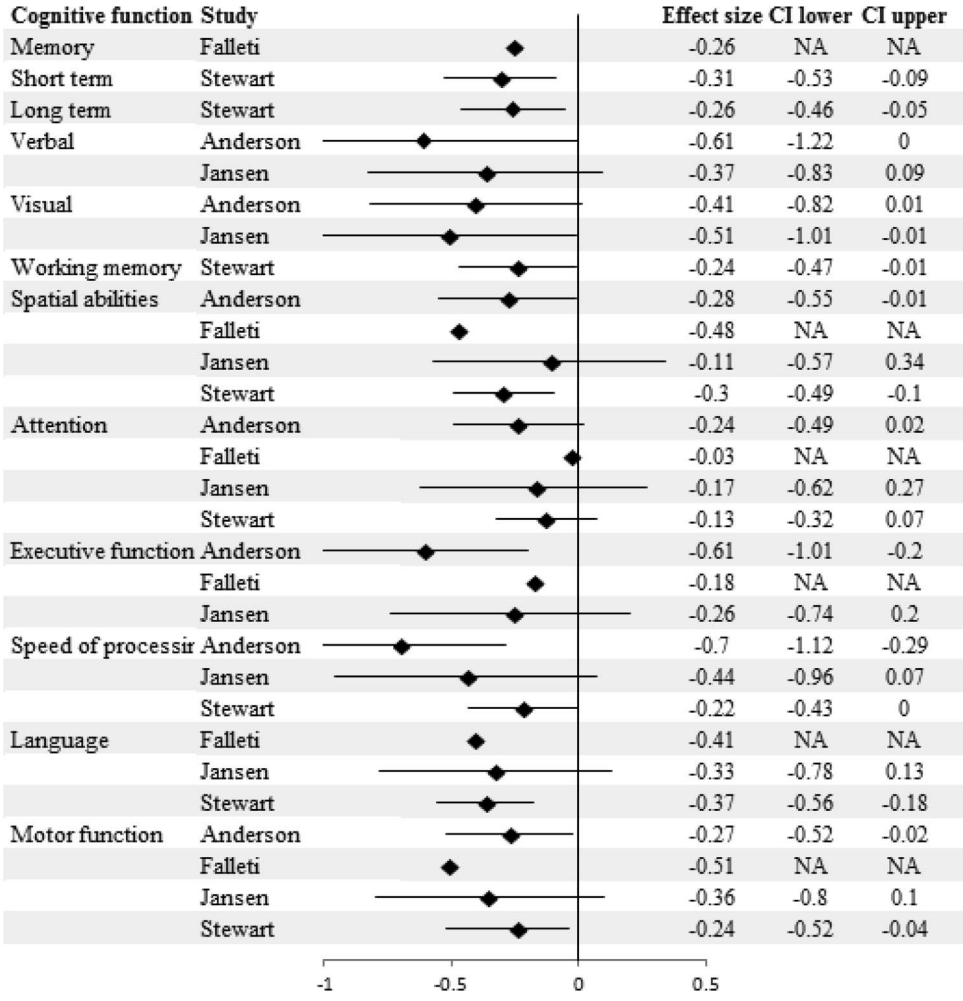


Figure 1. Forest plot of summary effect sizes obtained by four previous meta-analyses of studies with adults. *Note*. We report the type of cognitive function reported by each study, as well as the reported effect size, and its 95% CI. Falleti et al. (2005) did not report confidence intervals.

processing impairment ranges from low (Stewart et al., 2006), through medium (Jansen et al., 2005), to large (Anderson-Hanley et al., 2003).

In addition to variability in designs and reported outcomes in primary studies, there was also variation between the meta-analyses in the reporting of key methodological factors. As an example, these meta-analyses reported 11, 12, 18, and 20 items from the 27-item checklist of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Kyoko, Yoshitoku, & Toyonori, 2011). The items missed most frequently were those pertaining to the process of study screening, the reporting of confidence intervals, consistency, and risk of bias analyses (Anderson-Hanley et al., 2003; Falleti et al., 2005; Jansen et al., 2005; Stewart et al., 2006). Detailed literature search techniques were not always reported and the meta-analyses were based on small numbers of studies (n = 6, 7, 16, 29, respectively). When subgroup analyses were reported, they included even fewer studies (Stewart et al., 2006), although it is not generally recommended to run summary effect size analyses with a very small number of studies (Borenstein, Hedges, Higgins, & Rothstein, 2009).

There was also significant variability in the way neuropsychological test scores were pooled into cognitive functions and reported as single, general scores (i.e., verbal, short-term memory (STM), attention). Yet, current understanding of memory and attention is far more detailed than this (Strauss, Sherman, & Spreen, 2006). Such simplification of results, through the combination of very different measures, might impact the observed effect sizes, and the heterogeneity of results.

The Present Meta-Analysis

The present meta-analysis is a synthesis of current literature (K = 44) reporting cognitive functioning in adult cancer survivors treated with chemotherapy (described in online supplementary materials). We examined the extent of cognitive impairment but, compared to previous meta-analyses, we also explored potential sources of methodological and clinical heterogeneity, which might have influenced the results obtained within the literature.

We pursued two types of subgroup analyses. First, we grouped test scores into constructs based on well-established guidelines (Strauss et al., 2006) with a clear distinction between different memory and attention types. Table 1 summarizes the tests grouped within each neuropsychological function, as well as the number of studies (K) and effect sizes (N) included in the analyses.

Second, we divided studies into subgroups based on their designs, as cross-sectional studies and longitudinal studies might have different sources of bias. For example, in cross-sectional studies, the effective matching between patients and controls is crucial to the identification of real impairments, while longitudinal designs are particularly vulnerable to practice effects, both when tests have alternative formats and especially when they do not.

We further investigated the influence of two specific sources of bias, or moderators (Rosenthal & DiMatteo, 2001), which have not been considered by previous literature and could have influenced the observed effect sizes. The first is a methodological moderator, the quality of studies. The methods of conducting and reporting the results of primary studies, the quality of participant matching, the type of cognitive tests used, as well as availability of tests with alternative forms, are integral parts of their quality having a potential influence on the results of subsequent meta-analyses. Thus, we ran quality assessments of the studies and included the scores as potential methodological sources of heterogeneity.

The second sources of bias are clinical, related to the participant characteristics reported by primary studies and suggested by previous literature to have an impact on cognitive test results. These are the type of diagnosis, age of participants, and time since treatment. There could be several additional factors that might have had a significant impact, but those were either not reported (i.e., test results on treatment types or genders), reported inconsistently (i.e., types of treatments, time since diagnosis) or reported through different test scores (i.e., premorbid intelligence level of matched groups). The diagnosis was chosen as a proxy for the types of treatment and genders of the participants. The age of participants was chosen due to evidence from pediatric cancer studies that a younger age may be a vulnerability factor for cognitive impairments (von der Weid et al., 2003). Finally, some behavioural studies fail to show differences in functioning in breast cancer patients tested at several time points after treatment (Hermelink et al., 2007). Given that some impairments may fade with time, the time lapsed from treatment to assessment might be another significant factor influencing cognitive test scores.

Our meta-analysis followed the robust and comprehensive guidelines of the Cochrane Collaboration (Higgins & Green, 2008) for conducting systematic reviews and meta-analyses. Additionally, due to the problems with methods such as the fail-safe N, which assumes that the effect sizes of missing studies would be zero (Borenstein et al., 2009), we performed a regression-based publication bias analysis using Egger's method to account for potentially unreported data (Egger, Davey Smith, Schneider, & Minder, 1997; Sterne, Gavaghan, & Egger, 2000).

Methodology

Search Strategies

The relevant literature was examined by one person (OCL) through a search of the electronic databases (PubMed, Ebsco, Web of Science, PsychInfo, PRISMA, Cochrane) using the following search terms: (cancer OR chemotherapy) AND (cognition OR neuropsychology) AND (adults). We also conducted the search by replacing the words (cancer OR chemotherapy) with the names of chemotherapy drugs (i.e., doxorubicin, cyclophosphamide, etc.) and by replacing the words (cognition OR neuropsychology) with names of specific cognitive functions (i.e., attention, memory, verbal memory, executive functions, etc.). The reference lists of reviews were visually scanned and key journals of the International Psycho-Oncology Society, and conference proceedings were hand searched for additional articles not detected by the literature search (list included in online supplementary materials).

Inclusion and Exclusion Criteria

The study eligibility criteria are described in Table 2. They were driven by the participant/intervention/comparison/outcomes/study design elements (Higgins & Green, 2008), while accommodating the ICCTF guidelines for studies in the field (Vardy et al., 2008).

Our search included all studies from 1980 to January 2011. We did not include unpublished data, or articles written in languages

Table 1
List of Cognitive Tests Included in Each Function

Aggregate constructs	Specific constructs	Tests used	*K*	*N*
	Full-scale IQ	Groeninger Intelligence Scale, MMSE, WAIS.	5	5
Memory		All memory tests regardless of retention interval (immediate or delayed), test format (free recall or recognition), and modality (visual or verbal).	37	189
Verbal memory		Verbal memory tests regardless of retention interval and test format (free recall or recognition).	31	122
Visual memory		Visual memory tests regardless of retention interval and test format (free recall or recognition).	20	57
Immediate free recall		Immediate memory tests regardless of modality.	26	47
Delayed memory		Delayed memory tests, regardless of modality and test format.	26	69
Delayed recognition		All recognition memory tests regardless of modality.	12	19
	Verbal immediate free recall	Logical memory I, CVLT/RAVLT/HVLT/Rey 15, WMS Verbal memory immediate, RBANS Immediate memory, VLMT 1–5, VSRT Short term, Encoding/recall correct.	24	48
	Verbal delayed free recall	Logical memory II, CVLT/RAVLT/15 Rey Delayed, RBANS Delayed memory, WMS Delayed recall, VSRT Delayed.	24	48
	Verbal delayed recognition	RAVTL/CVLT/Rey 15 recognition, HVLT Discrimination, Paired associates recognition.	9	21
	Visual immediate free recall	Logical memory I, variants of CVLT/RAVLT/HVLT/VLMT, RBANS Immediate memory, Encoding/Recall correct, WMS Visual memory immediate.	17	27
	Visual delayed free recall	Visual reproduction II, Family pictures II, ROCFT delayed, NVSRT delayed, WMS Visual memory delayed.	15	22
	Visual delayed recognition	Visual reproduction recognition, ROCFT recognition, Visual association test.	8	7
Attention		Includes all attention tests.	11	17
	Focused attention	Trails A, Stroop, Digit symbol, Symbol search, Symbol modalities, D2, Continuous performance tests, Visual search tests.	28	74
	Selective attention	D2, Fepsy binary, Go/No go selective attention, TEA Auditory/visual elevator, Ruff 2&7.	10	24
	Attention capacity	Letter-number cancellation/sequencing, PASAT, Digit span, Visual span Forward, Sentence repetition.	19	59
	Executive functions	Stroop, Trail B, WCST, Tower of London, Consonant Trigrams, COWA or variants.	33	147
	Verbal abilities	Lexical/Semantic search, Boston naming test, WAIS/WRAT Reading, RBANS Language.	8	15
	Spatial abilities	Block design, ROCFT-Copy, RBANS Visuospatial (Figure copy and Line orientation).	15	18
	Arithmetic	WAIS, WISC, WPPSI and any other mathematical achievement tests.	10	17
Motor functions		Pegboard, Fingertapping, Grip strength dominant and non-dominant.	16	45

Note. The number of studies (*K*) and effect size estimates (*N*) within all the analyses, before the multiple outcomes transformations. (MMSE = Mini Mental State Examination; WAIS = Wechsler Adult Intelligence Scale; WMS = Wechsler Memory Scale; WTAR = Wechsler Test of Adult Reading; CVLT = California Verbal Learning Test; RAVLT = Rey Auditory Verbal Learning test; HVLT = Hopkins Verbal Learning Test; RBANS = Repeatable Battery for the Assessment of Neuropsychological Status; VLMT = Verbal Learning and Memory Test; VSRT = Verbal Selective Reminding Test; NVSRT = Non-Verbal Selective Reminding Test; ROCFT = Rey-Osterrieth Complex Figure Test; TEA = Test of Everyday Attention; PASAT = Paced Auditory Serial Addition Test; WCST = Wisconsin Card Sorting Test; COWA = Controlled Oral Word Associations).

other than English. When studies did not report means and standard deviations, they were requested from the authors. If the data were not provided, we did not include the articles because of the nature of the software used for the initial analyses. Figure 2 depicts the search process that led to the inclusion of 44 studies.

Coding Procedures

Means and standard deviations for all cognitive tests were recorded for each individual study, alongside the study quality and participant level moderators (cancer type, mean age, and mean number of years since treatment). The scores reported for each cognitive test were extracted from each article (i.e., Rey-Osterrieth Complex Figure Test-Copy as a measure for visuospatial ability). These scores were then grouped within functions based on the guidelines suggested by Strauss et al. (2006).

The strengths and weaknesses of the studies included in the analyses were graded according to a quality assessment tool recommended by the Cochrane Collaboration. The Downs and Black scale (Downs & Black, 1998) contains 27 questions pertaining to both randomized and nonrandomized studies and looks at all aspects of data reporting and analysis. Three questions, referring to the blinding of participants and experimenters, were removed due to the lack of suitability in the context of this research. Six additional questions were added to accommodate the ICCTF guidelines: whether any patients were exposed to local or cranial radiotherapy, had CNS malignancies, if a control group was present, and the inclusion of ICCTF recommended neuropsychological and self-assessment tools. Thus, each study received a score between 0 and 31.

The quality assessment was performed blindly by one master coder (OCL) and an independent researcher (DF). Disagreements about the

Table 2
Criteria for Including Studies in the Meta-analysis

	INCLUSION	EXCLUSION
Population	Patients exposed to chemotherapy.	Patients with central nervous system tumors.
Intervention	Patients exposed to chemotherapy.	Studies that test the effect of drugs other than chemotherapy (i.e. hormonal treatments). Patients exposed to CNS-directed radiotherapy.
Comparison/control group	Studies comparing patients with norms, healthy controls, or cancer patients who were not treated with chemotherapy.	Studies that did not report results of any control group.
Outcomes	Articles reporting means and standard deviations of at least one neuropsychological test. See Table 1 for a description of the type of tests.	Duplicate results (i.e. articles based on dissertations) and studies only reporting changes in psychosocial functioning such as quality of life. Studies not reporting means and standard deviations on the tests.
Study design	Longitudinal and cross-sectional studies.	Case-studies were excluded because the design is rarely used to examine intended effects of a treatment.

scores were resolved by consensus. Following consensus, an interclass correlation was performed because the scores were measured on an interval scale; it yielded an interrater correlation of 0.97 ($p < .001$), with scores ranging from 11 to 30.

Subgroup Analyses

Most studies reported multiple outcomes and multiple time-points scores for each cognitive function. This facilitated the separation of the data into subgroups based on the types of designs employed by individual studies. The resulting subgroup analyses were:

- Postchemotherapy cross-sectional studies: patients versus controls, after treatment.
- Patient longitudinal studies: patients at follow-up versus baseline.
- Baseline cross-sectional studies: patients versus controls, before treatment.
- Control longitudinal studies: control participants at follow-up versus baseline.

The results of the first two analyses triggered the analyses pertaining to the performance of patients versus controls before chemotherapy and the performance of controls in longitudinal assessments. Their aim was to determine whether patients were impaired before chemotherapy and whether practice effects were present in both patients and controls evaluated multiple times. Each analysis was based on at least four individual study estimates, which is sufficient to perform a meta-analysis (Patel, 1989).

Effect Size Estimations

We used the random effects model for each level of the analyses, in order to account for the high variability of the data (Overton, 1998). We have computed Hedge's g standardized mean difference between groups as it provides a tighter estimate of the true effect size (Borenstein et al., 2009). Each study reported several scores for each cognitive test, for each comparison group, and for different time-points. Initially, we ran individual random model meta-analyses for all the data, and then in the four additional subgroups for each of the 22 cognitive functions. These initial analyses resulted in specific standard errors, and weights assigned to each outcome within each study. All of these calculations were run using Meta-Analyst (Wallace, Schmid, Lau, & Trikalinos, 2009).

In order to account for the dependency of data due to multiple outcomes and time points, we further calculated study-level composite effect sizes and variances for each set of outcomes reported by each study, per cognitive function. These were computed based on the formulas suggested by Borenstein et al. (2009). Finally, each study appeared once in the final summary effect size analyses. Studies reporting the same outcomes on different patient groups (i.e., lymphoma and breast cancer, Ahles et al., 2002, 2003), were analyzed as two separate studies because of the importance of the differences between diagnoses.

Effect Size Integration

In order to create the random effects model summary effect sizes we used an adapted version of the meta-analysis macros developed by Field and Gillett (2010). This enabled us to use the composite effect sizes, and variances previously computed. Moreover, it helped us correct for unequal sample sizes by using the minimum weight received by each study in the initial analyses, and to compute the I^2 heterogeneity value.

The subgroup summary effect sizes are not directly comparable, thus they were interpreted as low if they were 0.2 or below, moderate between 0.2 and 0.5, and high above 0.8 (Cohen, 1992). However, as this is a general rule of thumb, effect sizes were also interpreted in the light of the specific literature on the topic of chemotherapy-induced cognitive impairments (Lipsey & Wilson, 2001), in which the effects are usually considered mild. For each analysis, we report the summary effect size, confidence intervals (95% CI), I^2, and the overall significance level representing the null hypothesis that the treatment effect is zero (Borenstein et al., 2009). The negative or positive valence of effect sizes denotes the direction of the effect of chemotherapy on cognitive function: negative if suggestive of impairments, and positive if suggestive of performance increases in one group relative to the other.

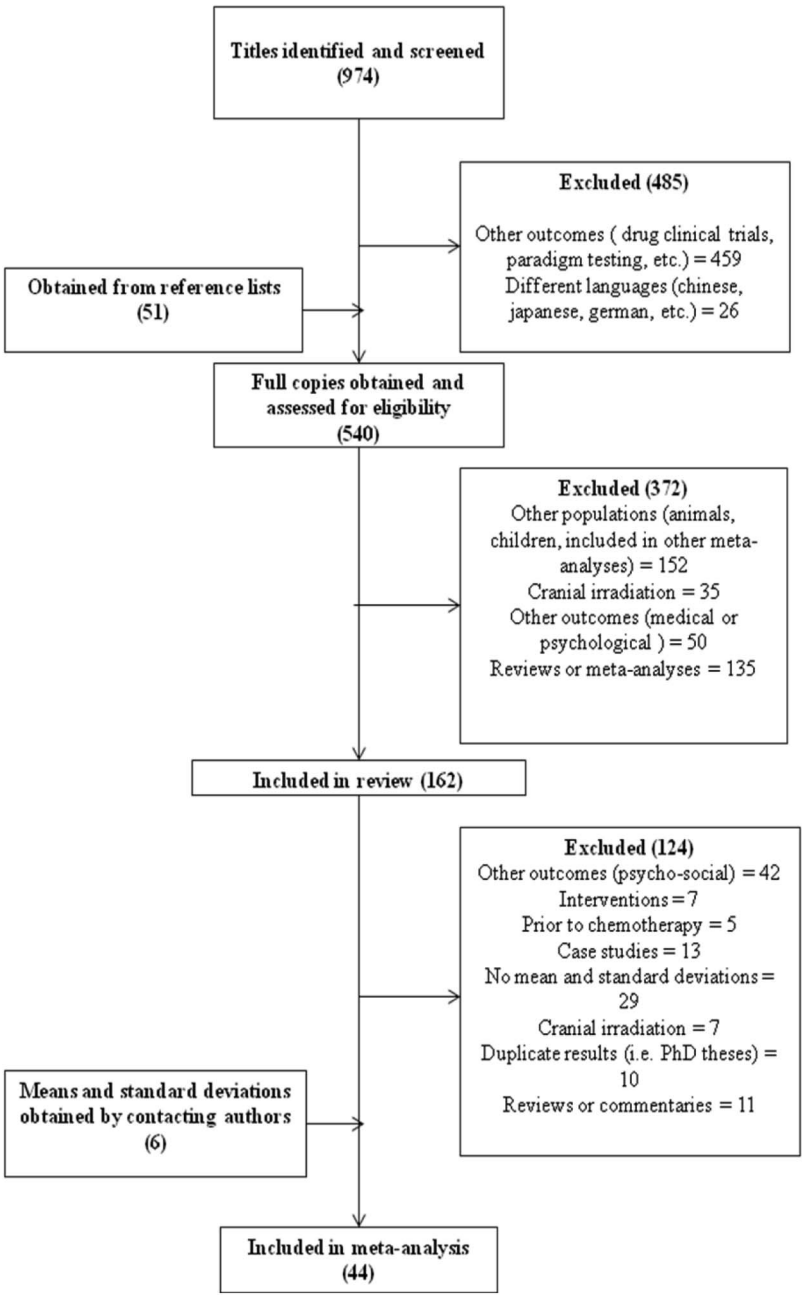


Figure 2. Study selection flowchart.

Publication and Selection Bias

The publication bias for each main cognitive function was assessed through Egger's regression test (Egger et al., 1997). It estimates the asymmetry of the funnel plot due to underreporting of data through a linear regression comprised of a normalized effect size (divided by its standard error) and precision (inverse of standard error) We present the significance level of the intercept, which was considered to suggest publication bias if significant for $p < .10$.

Multilevel Moderator Analyses

Presently, there are two options available for conducting moderator analyses within a meta-analysis. The first and most highly used method was to run metaregressions for each function, each design, and each moderator separately with the restricted maximum likelihood macros developed by Lipsey and Wilson (2001). The advantage of this method was that we could run specific analyses for each subgroup. However, this method can lead to a decrease in variance and a high likelihood of an increased Type I

error when comparing multivariate effect sizes. Due to this reason and the presence of a categorical moderator (diagnosis) which requires dummy coding, we also pursued a multilevel model analysis for all effect sizes and moderators (Hox, 2010). We will briefly report the results of the classical metaregressions, while focusing more on the results of the multilevel model approach.

The multilevel analyses were conducted with MLwin 2.1 (Rabash, Charlton, Browne, Healy, & Cameron, 2009) with the restricted maximum likelihood procedure. We used a 3-level model with the study outcomes (summary effect sizes) as the first level, cognitive functions as the second level, and the studies as the third level. The moderators were the ones described above: study quality, diagnosis (coded as a dummy variable), age of participants, and time since treatment.

We will first report the intercept-only model, when no predictors are included. This is described by the equation:

$$ES_{ij} = \beta_{0j} + u_{0j} + e_{ij}$$

ES_{ij} refers to the effect size for outcome i from study j, β_{0j} is the value of the intercept (average effect size for an average outcome), u_{0j} is the random error at level 2, and e_{ij} is the random error residual at level 1. The variance of u_{0j} suggests the variability in effect sizes.

In the moderator analyses the equations take the form of:

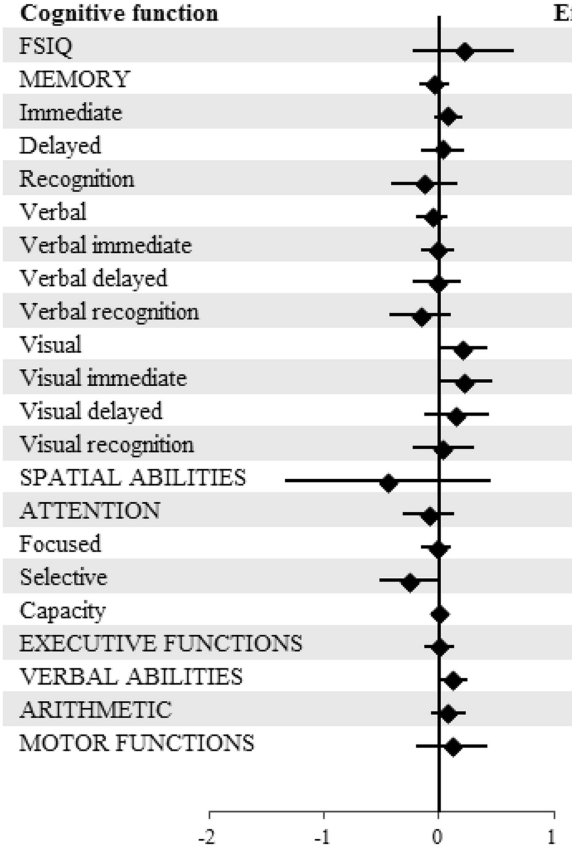
$$ES_{ij} = \beta_{0j} + u_{0j} + \beta_1 Moderator_{ij} + e_{ij}$$

All the parameters represent the same values as in the empty intercept model, while the β_1 value represents the slope of the regression, suggesting the strength and direction of the change in effect size for a one-unit change of the moderator. All the analyses were run with predictors centered on their grand mean, to reduce the possibility of correlations between the intercept and predictors, as well as between levels (Kreft & de Leeuw, 1998).

Results

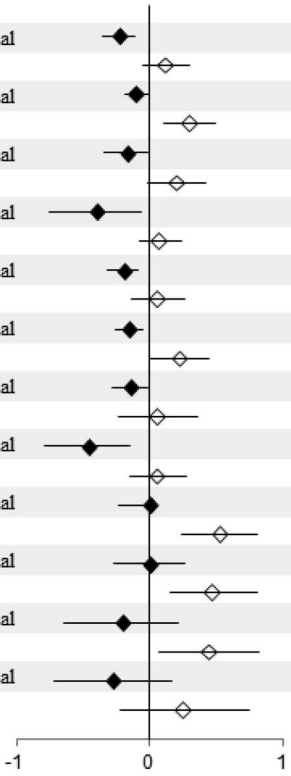
Study Characteristics

We analyzed data pertaining to the neuropsychological evaluations of 1,940 adult patients and 2,000 controls. In total, 30 out of 44 studies (70%) included only breast cancer patients. The remainder included patients with testicular cancer, or lymphoma, or other hematological malignancies. Mean participant age was 51.57 (SD = 6.29), and 75% of the studies included only female participants. All studies evaluated patients at an average of 2 years posttreatment (SD = 2.52). In online supplementary materials we



Cognitive function	Effect size	CI lower	CI upper	I^2	p	K	Egger's
FSIQ	0.21	-0.21	0.64	0	0.32	5	0.27
MEMORY	-0.04	-0.16	0.07	19.05	0.45	36	0.35
Immediate	0.08	-0.03	0.19	37.33	0.15	27	0.27
Delayed	0.03	-0.14	0.2	20.66	0.72	26	0.4
Recognition	-0.13	-0.41	0.14	0	0.33	12	0.7
Verbal	-0.06	-0.19	0.06	0	0.32	31	0.88
Verbal immediate	-0.01	-0.14	0.12	12.23	0.86	24	0.12
Verbal delayed	-0.02	-0.22	0.17	0	0.8	24	0.13
Verbal recognition	-0.16	-0.42	0.09	0	0.22	9	0.58
Visual	0.2	0	0.4	55.8	**0.04***	20	0.21
Visual immediate	0.22	0	0.45	39.31	**0.05***	17	0.84
Visual delayed	0.15	-0.11	0.42	47.54	0.25	15	0.27
Visual recognition	0.03	-0.22	0.29	0	0.81	6	0.74
SPATIAL ABILITIES	-0.44	-1.33	0.44	0	0.32	14	0.11
ATTENTION	-0.09	-0.3	0.11	2.12	0.36	11	0.17
Focused	-0.02	-0.14	0.09	10.05	0.71	27	0.43
Selective	-0.26	-0.51	0	0.15	**0.04***	10	0.87
Capacity	0	-0.06	0.06	0	0.94	20	**0.09**
EXECUTIVE FUNCTIONS	0	-0.11	0.11	0	0.95	34	0.4
VERBAL ABILITIES	0.12	0	0.23	0	0.03	10	0.2
ARITHMETIC	0.07	-0.06	0.22	0	0.27	10	0.69
MOTOR FUNCTIONS	0.11	-0.19	0.41	0	0.46	16	0.16

Figure 3. Forest plot of summary effect sizes in adult patients compared to any control group. *Note*. We report the Hedge's g effect size, the 95% CI, I^2, p as the significance level of the analysis, k as the number of studies in the analysis, and the significance level of the intercept in Egger's test. * $p < .05$.



Cognitive function	Design	Effect size	CI lower	CI upper	I^2	p	K	Egger's
MEMORY	Cross-sectional	-0.23	-0.35	-0.11	44.29	**<.001*** **	22	0.68
	Longitudinal	0.12	-0.05	0.3	41.51	0.16	18	0.64
Immediate	Cross-sectional	-0.1	-0.19	0	0	**0.03***	15	0.41
	Longitudinal	0.3	0.1	0.49	36.09	**<.01****	16	**0.02**
Delayed	Cross-sectional	-0.18	-0.34	-0.02	0	**0.02***	14	**0.07**
	Longitudinal	0.2	-0.02	0.42	31.26	0.08	16	0.44
Recognition	Cross-sectional	-0.4	-0.75	-0.06	0	**0.02***	8	0.27
	Longitudinal	0.07	-0.08	0.24	0	0.33	6	0.78
VERBAL	Cross-sectional	-0.2	-0.32	-0.09	0	**<.001*** **	17	0.17
	Longitudinal	0.06	-0.14	0.26	0	0.56	18	0.55
Immediate	Cross-sectional	-0.16	-0.26	-0.05	0	**<.01****	15	**0.03**
	Longitudinal	0.22	0	0.44	2.57	**0.05***	13	**0.03**
Delayed	Cross-sectional	-0.15	-0.28	-0.02	0	**0.02***	12	0.46
	Longitudinal	0.06	-0.23	0.36	3.99	0.65	15	0.23
Recognition	Cross-sectional	-0.47	-0.79	-0.15	0	**<.01****	5	0.13
	Longitudinal	0.05	-0.15	0.27	0	0.59	4	**0.09**
VISUAL	Cross-sectional	-0.1	-0.24	0.03	0	0.15	12	0.2
	Longitudinal	0.52	0.24	0.8	60.7	**<.001*** **	12	0.13
Immediate	Cross-sectional	0	-0.27	0.26	9.74	0.98	11	0.18
	Longitudinal	0.47	0.15	0.8	36.21	**<.01****	10	0.2
Delayed	Cross-sectional	-0.21	-0.64	0.21	0	0.33	9	0.12
	Longitudinal	0.44	0.07	0.81	28.07	**0.01****	10	0.12
Recognition	Cross-sectional	-0.28	-0.72	0.16	0	0.21	5	0.22
	Longitudinal	0.25	-0.22	0.74	0	0.29	3	0.36

Figure 4. Forest plot of longitudinal and cross-sectional study effect sizes in adult patients for memory functioning only. *Note*. k = number studies. Full diamonds = cross-sectional effect sizes; empty diamonds = longitudinal effect sizes. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

described the characteristics of each study included in our meta-analysis and the associated moderators.

Analysis of Effect Sizes Across All Subgroups

First, we compared the performance of all patient participants after chemotherapy to the performance of controls. Analyses were undertaken for all patients, irrespective of the comparison group—healthy controls or their own baseline performance—to allow comparison with previous meta-analyses. Effect sizes were small, had broad confidence intervals, and high heterogeneity (see Figure 3). Patients had statistically significant performance increases (positive summary effect sizes) for visual memory and visual immediate free recall. Patients' performance was significantly reduced (negative summary effect sizes) only for selective attention.

Analyses by Study Design

Cross-sectional designs. We performed separate analyses for cross-sectional studies, at posttreatment and baseline. Following chemotherapy, patients exhibited significant low to moderate impairments relative to controls (Figures 4 and 5). These were observed in memory, immediate free recall, delayed memory, delayed recognition, verbal memory, verbal immediate free recall, verbal delayed free recall, verbal delayed recognition, selective attention, and capacity of attention. Summary effect sizes of other cognitive functions were not statistically significant.

At baseline, before chemotherapy, patients performed better than controls, across most cognitive functions (see Figure 6). Effect sizes were moderate to high, and statistically significant, but had high heterogeneity values. Superior patient performance was observed in memory, attention, executive functions, spatial abilities, and verbal abilities. Due to an absence of reported outcome data we could not compute cognitive function effect sizes for 12 functions. Results were not significant for the remaining four functions, two of which (verbal memory and attention capacity) had negative values, suggestive of potential impairments.

To summarize the findings from cross-sectional designs, as expected, patients performed worse than controls after treatment. Contrary to expectations, patients performed better than controls at baseline, before treatment began. This pattern was observed across most cognitive functions, even in cases where it was not statistically significant. Figure 7 visually depicts the difference between our findings and what we predicted on the basis of available literature.

Longitudinal designs. When longitudinal studies were analyzed separately, there was a clear improvement in patients following chemotherapy, compared to baseline. This finding was true across most cognitive functions we were able to analyze (Figures 4 and 5). The improvements in performance were statistically

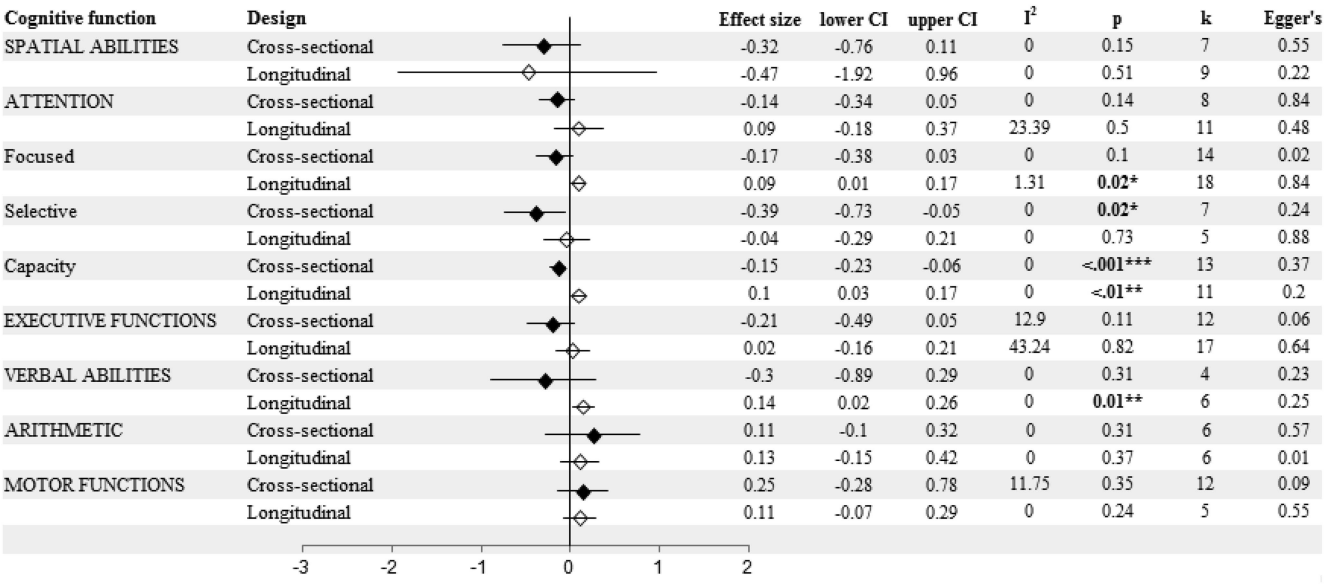


Figure 5. Forest plot of longitudinal and cross-sectional study effect sizes in adult patients for cognitive functions other than memory. *Note*. k = number studies. Full diamonds = cross-sectional effect sizes; empty diamonds = longitudinal effect sizes. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

significant for immediate free recall, verbal immediate free recall, visual immediate free recall, visual delayed memory, focused attention, capacity of attention, and verbal abilities. Effect sizes of other cognitive functions did not reach statistical significance. Notably, heterogeneity was higher in studies measuring longitudinal changes than in cross-sectional studies.

Longitudinal data from healthy control participants were infrequently reported (see Figure 8). As a result, effect sizes could only be computed for five of the 22 cognitive functions, each analysis including up to a maximum of seven studies. The analyses were restricted to those reporting patient and control comparison at different time points. Controls performed significantly better at follow-up than at baseline on memory and visual memory. These analyses were less heterogeneous, despite the smaller number of studies reporting longitudinal outcomes for control participants.

To summarize, contrary to expectations, patients performed better after chemotherapy than before. The only exceptions were spatial abilities and selective attention, which had negative values, but were not statistically significant. Controls also improved, but the effect sizes were not as large as the ones estimated in patients. Because the two sets of effect sizes within the two subgroups are not directly comparable, we depict the expected versus observed values for controls and patients in longitudinal studies (see Figure 9). For all other cognitive functions, which were not significant, patients' effect sizes were only negligibly reduced compared to those of controls (e.g., executive functions, patients g = .02 and control g = .08).

Moderator Analyses

The classical moderator analysis, which was run on Lipsey and Wilson's (2001) macros, had the advantage of analyzing the impact of each moderator on each function within the two main designs. However, it had the disadvantage of comparing multiple effect sizes characterized by group dependencies, thus the variances of the results may have been underestimated and the signif-

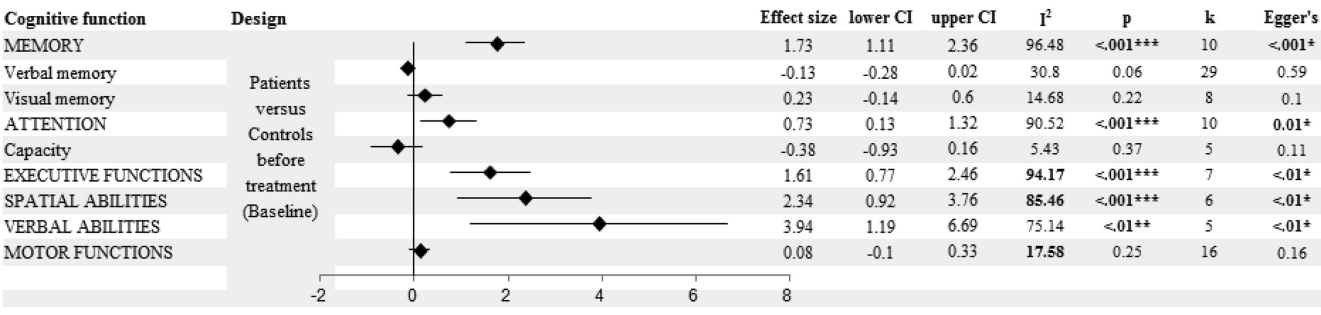


Figure 6. Forest plot effect sizes of patients versus controls at baseline for all cognitive functions. *Note*. We report the Hedge's g effect size, the 95% CI, I^2, p as the significance level of the analysis, k as the number of studies in the analysis, and the significance level of the intercept in Egger's test. $^{**}p < .01$, $^{***}p < .001$.

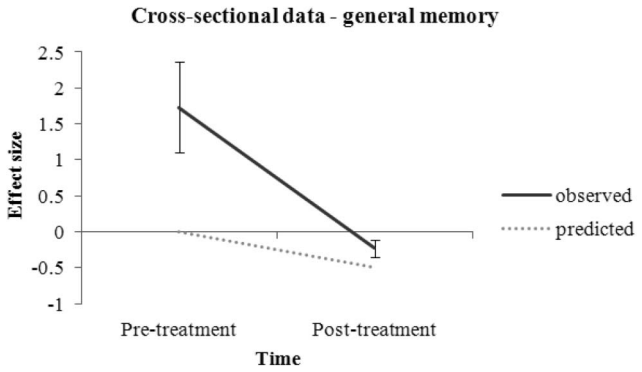


Figure 7. Pattern of significant effect sizes results for memory in cross-sectional studies. *Note*. Straight line represents the relationship between effect sizes obtained by patients versus control at pretreatment and post-treatment.

icance values overestimated. Still, in R^2 values, the quality of the studies, age, and time since treatment significantly explained between 38% and 69% of the variance of effect sizes in cross-sectional studies, and between 16% and 66% of the one in longitudinal studies. The most noteworthy results were those of the quality of studies influencing up to 34% of motor function effect size variance in cross-sectional studies, and time since treatment explaining 56% of the visual immediate free recall variance in longitudinal studies (data not presented in paper, available upon request).

Our multilevel model did not have the advantage of analyzing the results in subgroups, but for all available effect sizes. We will first report the result of the intercept-only model (see Tables 3 and 4 for all the coefficients). The first level was described by the summary effect sizes, the second level was represented by the cognitive functions, while the third level was the study itself. The intercept only model, independent from variances at the third level was estimated at .18 (standard error = .07). Compared to the *z*-critical value for $p <$ .05, the residual variance was significant, 3.10 (.17). In other terms, the overall mean effect size, irrespective of the type of design, functions, or variances at the third level, was a low positive summary effect size, but there is a great amount of unexplained variance.

When the intercept was set to vary at the study level, the average effect size decreased to .13 (.11). The between study variance of .24 (.10) was significant for $p <$ .05 and the residual variance decreased but was still significant with 2.79 (.16).

Given the highly significant study level and residual variances, we carried on by including each of our moderators in the multilevel model. Compared to the results obtained in the classical metaregression, none of the continuous moderators had a significant impact on the summary average effect size estimation. The slope for the quality of the studies was .05 (.03), for age .01 (.01), and for time since treatment −.04 (.04). The average effect sizes for the types of diagnoses were .18 (.12) for breast cancer, −.07 (.30) for testicular cancer, and .05 (.27) for other diagnoses. The slopes had different orientations depending on the reference dummy-coded category. For all these results both between study and residual variances continued to be significant and of roughly similar values (see Tables 3 and 4 for details).

Publication Bias Analysis

We performed publication bias analyses for each cognitive function examined, and some analyses need to be treated with caution. When analyzing the data irrespective of the type of control group, the intercept had a $p <$ .10 only for capacity of attention. In cross-sectional studies, verbal immediate free recall, delayed memory, focused attention, executive and motor functions, were influenced by publication bias. In longitudinal studies, the bias was present for immediate free recall, arithmetic, verbal free recall immediate, and visual free recall immediate. In the baseline subgroup, memory, executive functions, spatial, and verbal abilities were lower than .10 significance values, while the analyses of controls in longitudinal studies showed no influence of publication bias. Thus, these cognitive functions may have been reported more often in the literature if they showed impairments, while accounts of increases or lack of change might have been underreported.

Discussion

Our meta-analysis summarizes the findings from 44 studies examining an array of cognitive functions in adult cancer patients. Our primary objective was to identify which functions are impaired in each type of design. To that end, we divided studies into subgroups based on their design and calculated summary effect sizes for each type of cognitive function. The secondary objective was to identify potential factors that might explain the variability of results in previous literature. Thus, we analyzed the impact of four moderating factors on all effect sizes.

When analyzing all data, regardless of design and type of control group, only selective attention was impaired. Compared to previous meta-analyses, our effect sizes either did not reach sta-

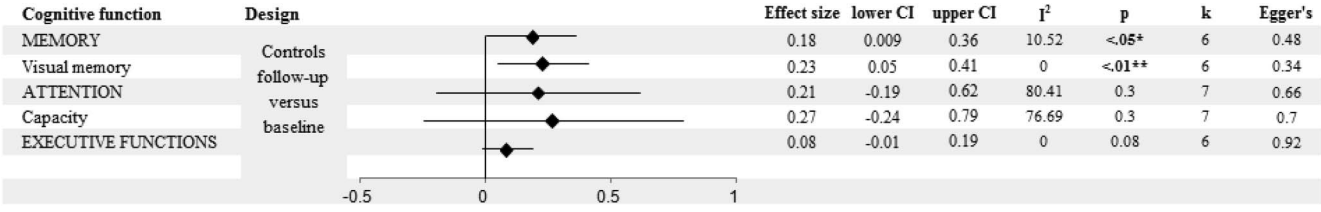


Figure 8. Forest plot effect sizes of controls at follow-up versus baseline for all cognitive functions. *Note*. We report the Hedge's *g* effect size, the 95% CI, I^2, *p* as the significance level of the analysis, *k* as the number of studies in the analysis, and the significance level of the intercept in Egger's test. * $p < .05$, ** $p < .01$.

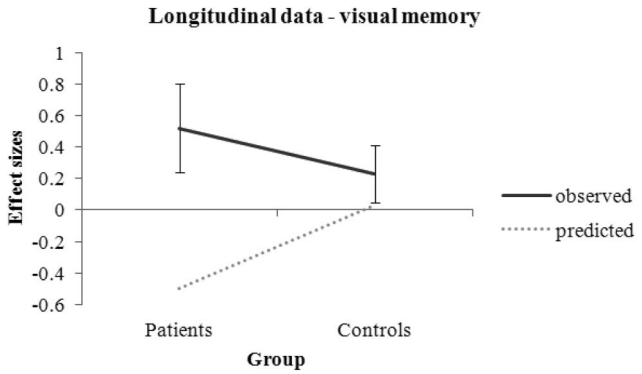


Figure 9. Pattern of significant effect sizes for visual memory in longitudinal studies *Note*. Straight line represents the relationship between effect sizes obtained by patients at pretreatment versus posttreatment, and separately for healthy controls.

tistical significance, or were very close to zero. Thus, pooling together a higher number of primary studies with inconsistent results, ultimately summed up to heterogeneous summary effect sizes, which did not distinguish between types of impairments (Rosenthal & DiMatteo, 2001).

Despite the influence of several confounding variables on the effect sizes, when patients were compared to controls at posttreatment, we observed small to moderate effect sizes, suggesting impairments in all aspects of verbal memory (immediate free recall, delayed free recall, and delayed recognition), in selective attention, and attention capacity. Furthermore, results in the cross-sectional subgroup analyses had lower heterogeneity values than those in other subgroups. This may suggest that cross-sectional studies are influenced by less and more similar sources of heterogeneity, while the latter may be more diverse in longitudinal studies. Irrespective of which hypothesis is true, they both suggest the need for better controlled studies to reduce subsequent heterogeneity (i.e., through proper participant matching).

Our cross-sectional analysis examining patients and controls at baseline investigated patients' performance before exposure to chemotherapy. Patients performed strikingly better than matched controls, for instance in their verbal ($g = 3.94$) and spatial abilities ($g = 2.34$).

The difference between the baseline and posttreatment results in patients and controls is depicted in Figure 7. This pattern contradicted our expectations based on previous literature. First, at baseline in cross-sectional studies, we would have expected patients to either perform worse than controls if the cancer itself would have had a deleterious impact (Ahles et al., 2008), or at the same level as controls if participants had been matched accordingly on their educational and intelligence levels. This effect is not visible in our results. On the one hand, the high effect sizes at baseline clearly suggest that patients are not impaired before chemotherapy. They also question whether there were other factors that differentiated patients from controls from the start of the study, such as different education, socioeconomic status, different intelligence levels, and even different motivation to perform well in testing. We note, however, that the two sets of analyses are not directly comparable, thus this is a relative comparison with the results one would expect based on previous literature reviews (Vardy et al., 2008; Zachariae & Mehlsen, 2011). Posttreatment analyses are drawn mostly from cross-sectional studies, while the baseline analyses are only drawn from longitudinal studies in which patients were compared to control participants at baseline, and then at several time points postchemotherapy. Thus, the potential poorer participant matching in the baseline assessments might stem from variability in longitudinal studies.

In longitudinal studies, patients performed better in follow-up evaluations than at baseline, with small to moderate effect sizes across multiple functions. Improvements between the first and second tests were also observed in control participants, for the limited set of cognitive functions we were able to analyze. These were lower and based on a smaller number of studies than the effects computed on patients, but were less heterogeneous than the patient analyses. Just as in the two cross-sectional subgroup analyses, the lower heterogeneity may be explained by the fact that the two sets of data were sourced from different articles, which may have been affected by confounders differentially. Despite this, the effect sizes of the patient group are still high, specifically for immediate free recall, and verbal immediate free recall.

The pattern of impairments in cross-sectional studies appears congruent with previous imaging studies. Compared to matched controls, breast cancer patients had significant left lateralized white matter decreases in the parahippocampal gyrus (de Ruiter et al., 2011; Inagaki et al., 2007; McDonald, Conroy, Ahles, West, & Saykin, 2010), and reduced activations in the left lateral posterior parietal regions and left dorsolateral prefrontal cortex (de Ruiter et al., 2012, 2011; Deprez et al., 2011). The moderate impairments in selective attention may be associated with the decreased white matter in the superior fronto-occipital fasciculus and superior and

Table 3
Results of Multilevel Regression Analyses for Each Continuous Moderator

Moderator	β_0 (SE)	β_1(SE)	σ^2_u (SE)	σ^2_e (SE)	-2^*loglikelihood
Empty model (independent from study)	.18 (.07)	NA	NA	3.106 (.17)	2377.67
Empty model (dependent on study)	.13 (.11)	NA	.24 (.103)	2.793 (.16)	2344.65
Quality	.16 (.11)	.05 (.03)	.22 (.09)	2.791 (.16)	2341.55
Age	.13 (.11)	.01 (.01)	.24 (.10)	2.793 (.16)	2343.56
Time	.14 (.11)	−.04 (.04)	.23 (.10)	2.796 (.16)	2343.69
					$K = 44$ $N = 599$

Note. We report K as the number of studies in the analysis, N as the number of effect sizes included, β_0 as the value of the intercept, β_1 as the value of the slope associated with a certain moderator, the associated standard errors, σ^2_u as the variance associated with the study level, σ^2_e the variance of the random error, and the -2^*loglikelihood values.

Table 4
Results of Multilevel Regression Analyses for Each Type of Cancer

	Breast cancer	Testicular cancer	Mixed diagnoses
β_0 (SE)	(β_1/SE)	(β_1/SE)	(β_1/SE)
.18 (.12)	Reference	−.25 (.32)	−.13 (.30)
−.07 (.30)	.25 (.32)	Reference	.12 (.39)
.05 (.27)	.13 (.09)	−.12 (.39)	Reference
		σ^2_u = 2.798 (.16)	
		σ^2_e = .27 (.02)	
K = 44	N = 599		−2*loglikelihood = 2343.96

Note. We report K as the number of studies in the analysis, N as the number of effect sizes included, β_0 as the value of the intercept, β_1 as the value of the slope associated with a certain moderator, the associated standard errors, σ^2_u as the variance associated with the study level, σ^2_e the variance of the random error, and the −2*log-likelihood values.

medial frontal gyri observed in other imagining studies (de Ruiter et al., 2012, 2011; Inagaki et al., 2007; McDonald et al., 2010; Silverman et al., 2007). Despite the legitimacy of these hypotheses, more studies are needed in order to determine if these are the actual structural and functional changes. Furthermore, the pattern of deficits in attention and memory makes it difficult to conclude whether the memory problems exhibited by cancer survivors are dependent on damage to the medial temporal lobes, or secondary to damage in frontal or parietal neocortical regions associated with attention performance. We hypothesize that some aspects of the memory impairments may be primary and others dependent on attention deficits, but further research is needed to address this issue.

The cognitive performance increases in patients in longitudinal studies are surprising, but may be linked to either additional sources of bias or genuine long-term improvements. Patients may be prone to a relatively stronger influence of practice effects due to certain characteristics that may increase their motivation to take part in such studies, compared to control participants. Reasons could include the desire to perform well in the test, preexisting knowledge that chemotherapy may be associated with cognitive impairment, and differential setting and framing of the tests for patients and control participants (Schagen, Das, & Vermeulen, 2012). However, some deficits may only be short-lasting effects (McDonald et al., 2010; Silverman et al., 2007), or some participants may have higher cognitive reserves allowing the development of compensatory strategies in specific cognitive tasks (Ahles et al., 2010). Both these hypotheses warrant further investigation.

The multilevel moderator analyses showed that neither of our chosen moderators explained the variance in effect sizes. This is a surprising result, given that it is unlikely that a different age, or a different time since treatment would not affect cognitive functioning. Some of our hypotheses regarding this nonsignificant result are that the multilevel model collapsed all summary effect sizes for all cognitive functions and all types of designs. If cross-sectional designs have negative summary effects, while longitudinal ones have positive summary effects, the resulting average effect size may approach zero.

A second potential explanation is the description of moderators. The quality of the studies was assessed with one of the best assessment tools recommended by the Cochrane Collaboration. However, this may have been unsuitable for the types of studies conducted in this specific field, and for the time being there are no such tools available to evaluate the quality of neuropsychological studies. It may have been necessary to differentiate between specific aspects of quality, such as the presence of tests with alternative formats, the reporting of certain test scores (i.e., full-scale IQ), and the type of matching available in primary studies. The age of participants varied between 38 and 71, with an average mean age of 51.57 (SD = 6.29). The type of diagnosis is restricted to breast cancer in 70% of the studies, and the time since treatment in years varies between zero and 10 (M = 2.09, SD = 2.52). The nonhomogenous samples in primary studies reflects the higher variance within the moderators, a reason why the nonsignificant results are expected: there are not many studies available including participants of roughly the same age, or the same time since treatment, which may have led to nonsignificant results when all effect sizes are collapsed together. We consider this aspect to warrant further examination in future meta-analyses.

Due to the limitations of our meta-analysis (majorly due to the unequal distribution of primary studies) we caution the interpretation of some of our results. First, our analyses show that the heterogeneity of our estimated effect sizes was not explained by our chosen moderators. Both between-study and residual variances remained high despite the inclusion of predictors. However, effect sizes in longitudinal studies are more heterogeneous than in cross-sectional studies, possibly due to additional factors which have not been measured, or have not been reported (e.g., impact of hormonal treatment, psychological comorbidities, etc.). While many studies report matching on age and gender, intelligence is often a factor reported through various scores, ranging from the full scale IQ, to verbal or performance IQ. If the groups were not properly matched in primary studies, that would deem any meta-analytical results less trustworthy. To reduce such a possibility in future studies, intelligence should be measured and reported through the same measures (FSIQ). Throughout our analyses we assumed the control participants and patients were matched on premorbid IQ, while this may not have been the case in all primary studies. For example, if patients' premorbid FSIQ were higher than controls', the observed deficits might not actually be mild, but severe for a highly functioning person. Due to this reason, the reporting of these scores in all future primary studies is warranted.

Second, the cancer diagnosis moderator was actually a proxy of treatment protocols and gender. While the type of treatment would be a valuable factor to analyze in relation to cognitive functioning, there is a high variability in the treatments reported, depending on

the staging of the illness and patient-level medical characteristics. Importantly, the use of other classes of drugs such as corticosteroids, hormone antagonists or antiemetics, may also have influenced the results (Lupien et al., 2002), but these details are not always reported within the literature. Primary studies do not usually report the results of cognitive tests separately on the types of treatments.

As most studies with adults focus on patients with breast cancer, the cancer diagnosis is also a proxy of the gender. Previous studies with pediatric cancer patients have shown that female gender may be a vulnerability factor to chemotherapy-induced changes (von der Weid et al., 2003). However, the results of cognitive tests were not reported separately in any of the studies including both males and females; thus gender could be a moderator to be accounted for in future studies, especially when extending to the assessment of treatments for malignancies other than breast cancer.

Third, due to technical limitations, our review only included studies reporting means and standard deviations. Although we attempted to obtain this data from corresponding authors, this was not always possible, and resulted in 23 studies being excluded. However, the Egger's values are only significant for four cognitive functions, thus most our analyses were not influenced by publication or selection bias. This relates to our fourth limitation—not adjusting for the publication bias found in these analyses. The distribution of our data within multiple subgroups, as well as the unaccounted between-study heterogeneity (Terrin, Schmid, Lau, & Olkin, 2003) would have resulted in inaccurate trim-and-fill results. However, there are only four specific results that should be treated with caution; specifically, in cross-sectional studies, delayed and verbal immediate memory, and in longitudinal studies, immediate memory and verbal immediate free recall.

Finally, while heterogeneity was lower in cross-sectional subgroup analyses, it remained high in longitudinal studies. Our work should assist in reducing measurement noise in future empirical work. This should help minimize heterogeneity in future meta-analyses, as well as reducing the number of confounding variables influencing results of primary studies.

Conclusions

The present paper summarizes research in the field of chemotherapy-induced cognitive impairments, while highlighting the nature, extent of impairments, and factors influencing their identification. Despite the considerable heterogeneity of the data, the results obtained from cross-sectional studies could be considered the most reliable. With potentially less influence from additional variables, patients in cross-sectional studies performed worse than controls on tests of capacity of attention, selective attention, verbal memory, immediate, and delayed, in both free recall and recognition tasks. Although the interplay between attention and memory impairments remains a matter for future research, our results suggest that the impairments might be linked to both frontal and medial temporal lobe dysfunction.

We have shown that cognitive performance prior to chemotherapy was higher in patients than in controls. That suggests that malignancy itself was not responsible for neuropsychological late effects, but it also casts doubt on the quality of participant matching and unreported sources of bias in longitudinal studies.

Our moderator analyses were not significant, which is surprising given the plethora of factors that could influence cognitive data. This is the reason why, for the aid of future analyses on this topic, we suggest a number of guidelines that could be followed in future studies:

1. The use of shorter neuropsychological batteries by focusing specifically on certain cognitive functions. This strategy would shorten the testing time and maintain participants' interest active throughout the sessions. This option could decrease the differences in participants' motivational levels during testing.

2. Longitudinal studies should only use cognitive tests with alternative formats to avoid practice effects. Alternatively, when this standard cannot be achieved, it would be preferable to pursue cross-sectional studies.

3. Avoiding the use of tests that lack equivalent alternative formats and lacking sensitivity to very mild cognitive impairments (i.e., MMSE or RBANS).

4. When using neuropsychological tests, striving to use very similar versions of the same cognitive tests between research groups, and reporting the same scores. This would promote a more consistent impression of the impairments across studies. Examples of tests to be used would be the HVLT (and other similar versions, such as RAVLT or CVLT), the ROCFT (or other similar versions), any sections of the DKEFS (or similarly, the Stroop, Trail Making Test, and Controlled Oral Word Associations), D2 (or Ruff 2&7), and Digit Span for working memory.

5. Consistently grouping test scores into cognitive functions, as the high number of neuropsychological tests makes it difficult to understand whether two different results refer to the same function. The present meta-analysis groups the tests within cognitive functions as suggested by Strauss et al. (2006).

6. Memory and attention have been consistently found impaired in many primary studies and meta-analyses, including our own. However, the links and mechanisms of these impairments are not yet explained. They could be investigated further through the development and administration of newly designed tests inspired by the neural mechanisms of these processes.

7. Reporting premorbid intelligence levels in a unitary fashion, as this is a key factor in matching controls and patients. This can be achieved by reporting the full-scale IQ score as measured with the WTAR or NART. These tests should correlate with most cognitive measures, unless the patients have very specific functional or structural brain changes due to treatment.

8. All cross-sectional studies should match participants closely on age, education, gender, and IQ, as all four

variables could potentially change the whole interpretation of a cognitive dataset.

9. Results should be reported separately on the basis of moderators that could induce additional bias: gender, age of participants (if they vary between younger and older adults), relapses (as a potential factor relevant for the severity of impairments), types of diagnoses, and treatments.

10. Studies including control groups at baseline and follow-up could also report this set of data in the same table as the results of the patients.

While incorporating these conclusions and suggestions, future research should focus on the stability of these side effects, the link between memory and attention impairments, and the treatments and clinical vulnerability factors that may predispose some patients more to impairments, rather than others. These future findings would inform the cognitive intervention strategies required to help present and former patients cope with chemotherapy-induced cognitive impairments.

References

Ahles, T. A., & Saykin, A. J. (2007). Candidate mechanisms for chemotherapy-induced cognitive changes. *Nature Reviews Cancer, 7,* 192–201. doi:10.1038/nrc2073

Ahles, T. A., Saykin, A. J., Furstenberg, C. T., Cole, B., Mott, L. A., Skalla, K., . . . Silberfarb, P. M. (2002). Neuropsychologic impact of standard-dose systemic chemotherapy in long-term survivors of breast cancer and lymphoma. *Journal of Clinical Oncology, 20,* 485–493. doi:10.1200/JCO.20.2.485

Ahles, T. A., Saykin, A. J., McDonald, B. C., Furstenberg, C. T., Cole, B. F., Hanscom, B. S., . . . Kaufman, P. A. (2008). Cognitive function in breast cancer patients prior to adjuvant treatment. *Breast Cancer Research and Treatment, 110,* 143–152. doi:10.1007/s10549-007-9686-5

Ahles, T. A., Saykin, A. J., McDonald, B. C., Li, Y., Furstenberg, C. T., Hanscom, B. S., . . . Kaufman, P. A. (2010). Longitudinal assessment of cognitive changes associated with adjuvant treatment for breast cancer: Impact of age and cognitive reserve. *Journal of Clinical Oncology, 28,* 4434–4440. doi:10.1200/JCO.2009.27.0827

Ahles, T. A., Saykin, A. J., Noll, W. W., Furstenberg, C. T., Guerin, S., Cole, B., & Mott, L. A. (2003). The relationship of APOE genotype to neuropsychological performance in long-term cancer survivors treated with standard dose chemotherapy. *Psycho-Oncology, 12,* 612–619. doi:10.1002/pon.742

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. doi:10.1017/S1355617703970019

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. New York, NY: Wiley. doi:10.1002/9780470743386

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. doi:10.1037/0033-2909.112.1.155

Deprez, S., Amant, F., Yigit, R., Porke, K., Verhoeven, J., Van den Stock, J., . . . Sunaert, S. (2011). Chemotherapy-induced structural changes in cerebral white matter and its correlation with impaired cognitive functioning in breast cancer patients. *Human Brain Mapping, 32,* 480–493. doi:10.1002/hbm.21033

de Ruiter, M. B., Reneman, L., Boogerd, W., Veltman, D. J., Caan, M., Douaud, G., . . . Schagen, S. B. (2012). Late effects of high-dose adjuvant chemotherapy on white and gray matter in breast cancer survivors: Converging results from multimodal magnetic resonance imaging. *Human Brain Mapping, 33,* 2971–2983. doi:10.1002/hbm.21422

de Ruiter, M. B., Reneman, L., Boogerd, W., Veltman, D. J., van Dam, F. S., Nederveen, A. J., . . . Schagen, S. B. (2011). Cerebral hyporesponsiveness and cognitive impairment 10 years after chemotherapy for breast cancer. *Human Brain Mapping, 32,* 1206–1219. doi:10.1002/hbm.21102

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology Community Health, 52,* 377–384. doi:10.1136/jech.52.6.377

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629–634. doi:10.1136/bmj.315.7109.629

Falleti, M. G., Sanfilippo, A., Maruff, P., Weih, L. A., & Phillips, K. A. (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. doi:10.1016/j.bandc.2005.05.001

Ferguson, R. J., Ahles, T. A., Saykin, A. J., McDonald, B. C., Furstenberg, C. T., Cole, B. F., & Mott, L. A. (2007). Cognitive-behavioral management of chemotherapy-related cognitive change. *Psycho-Oncology, 16,* 772–777. doi:10.1002/pon.1133

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63,* 665–694. doi:10.1348/000711010X502733

Freeman, J. R., & Broshek, D. K. (2002). Assessing cognitive dysfunction in breast cancer: What are the tools? *Clinical Breast Cancer, 3,* Suppl 3, S91–S99. doi:10.3816/CBC.2002.s.019

Hermelink, K., Untch, M., Lux, M. P., Kreienberg, R., Beck, T., Bauerfeind, I., Munzel, K. (2007). Cognitive function during neoadjuvant chemotherapy for breast cancer. *Cancer, 109,* 1905–1913. doi:10.1002/cncr.22610

Higgins, J. P. T., & Green, S. (Eds). (2008). *Cochrane handbook for systematic reviews of interventions*. Chichester, England: Wiley-Blackwell. doi:10.1002/9780470712184

Hox, J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Hurria, A., Somlo, G., & Ahles, T. (2007). Renaming "chemobrain". *Cancer Investigation, 25,* 373–377. doi:10.1080/07357900701506672

Hutchinson, A. D., Hosking, J. R., Kichenadasse, G., Mattiske, J. K., & Wilson, C. (2012). Objective and subjective cognitive impairment following chemotherapy for cancer: A systematic review. *Cancer Treatment Reviews, 38,* 926–934. doi:10.1016/j.ctrv.2012.05.002

Inagaki, M., Yoshikawa, E., Matsuoka, Y., Sugawara, Y., Nakano, T., Akechi, T., . . . The Breast Cancer Survivors' Brain MRI Database Group. (2007). Smaller regional volumes of brain gray and white matter demonstrated in breast cancer survivors exposed to adjuvant chemotherapy. *Cancer, 109,* 146–156. doi:10.1002/cncr.22368

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 impairment in patients with breast cancer. *Oncology Nursing Forum, 34,* 997–1005. doi:10.1188/07.ONF.997-1005

Jansen, C. E., Miaskowski, C., Dodd, M., Dowling, G., & Kramer, J. (2005). A meta-analysis of studies of the effects of cancer chemotherapy on various domains of cognitive function. *Cancer, 104,* 2222–2233. doi:10.1002/cncr.21469

Kohli, S., Griggs, J. J., Jean-Pierre, P., Mustian, K. M., Roscoe, J. A., Wade, J. L., . . . Morrow, G. R. (2006). Self-reported cognitive impair-

ment in cancer patients: A URCC CCOP longitudinal study of 595 patients. *Journal of Clinical Oncology, 24*(18S), 8502.

Kreft, I., de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousands Oaks, CA: Sage.

Kyoko, T., Yoshitoku, Y., & Toyonori, O. (2011). Practice guideline of evidence-based medicine: Preferred Reporting Items for Systematic Reviews and Meta-analyses (the PRISMA statement). *Journal of Information Processing & Management/Joho Kanri, 54,* 254–266.

Lipsey, M. K., Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Lupien, S. J., Wilkinson, C. W., Briere, S., Menard, C., Ng Ying Kin, N. M., & Nair, N. P. (2002). The modulatory effects of corticosteroids on cognition: Studies in young human populations. *Psychoneuroendocrinology, 27,* 401–416. doi:10.1016/S0306-4530(01)00061-0

Maddams, J., Brewster, D., Gavin, A., Steward, J., Elliott, J., Utley, M., & Moller, H. (2009). Cancer prevalence in the United Kingdom: Estimates for 2008. *British Journal of Cancer, 101,* 541–547. doi:10.1038/sj.bjc.6605148

McDonald, B. C., Conroy, S. K., Ahles, T. A., West, J. D., & Saykin, A. J. (2010). Gray matter reduction associated with systemic chemotherapy for breast cancer: A prospective MRI study. *Breast Cancer Research and Treatment, 123,* 819–828. doi:10.1007/s10549-010-1088-4

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods, 3,* 354–379. doi:10.1037/1082-989X.3.3.354

Patel, M. S. (1989). An introduction to meta-analysis. *Health Policy, 11,* 79–85. doi:10.1016/0168-8510(89)90058-4

Rabash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2009). *MLwiN Version 2.1* [software]. Bristol, UK: Centre for Multilevel Modelling, University of Bristol.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52,* 59–82. doi:10.1146/annurev.psych.52.1.59

Schagen, S. B., Das, E., & Vermeulen, I. (2012). Information about chemotherapy-associated cognitive problems contributes to cognitive problems in cancer patients. *Psycho-Oncology, 21,* 1132–1135. doi:10.1002/pon.2011

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. doi:10.1007/s10549-005-9055-1

Short, P. F., Vasey, J. J., & Tunceli, K. (2005). Employment pathways in a large cohort of adult cancer survivors. *Cancer, 103,* 1292–1301. doi:10.1002/cncr.20912

Silverman, D. H. S., Dy, C. J., Castellon, S. A., Lai, J., Pio, B. S., Abraham, L., . . . Ganz, P. A. (2007). Altered frontocortical, cerebellar, and basal ganglia activity in adjuvant-treated breast cancer survivors 5–10 years after chemotherapy. *Breast Cancer Research and Treatment, 103,* 303–311. doi:10.1007/s10549-006-9380-z

Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology, 53,* 1119–1129. doi:10.1016/S0895-4356(00)00242-0

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. *The Clinical Neuropsychologist, 20,* 76–89. doi:10.1080/138540491005875

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York, NY: Oxford University Press.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine, 22,* 2113–2126. doi:10.1002/sim.1461

Vardy, J., Wefel, J. S., Ahles, T., Tannock, I. F., & Schagen, S. B. (2008). Cancer and cancer-therapy related cognitive dysfunction: An international perspective from the Venice cognitive workshop. *Annals of Oncology, 19,* 623–629. doi:10.1093/annonc/mdm500

von der Weid, N., Mosimann, I., Hirt, A., Wacker, P., Nenadov Beck, M., Imbach, P., . . . Wagner, H. P. (2003). Intellectual outcome in children and adolescents with acute lymphoblastic leukaemia treated with chemotherapy alone: Age- and sex-related differences. *European Journal of Cancer, 39,* 359–365. doi:10.1016/S0959-8049(02)00260-5

Wallace, B. C., Schmid, C. H., Lau, J., & Trikalinos, T. A. (2009). Meta-Analyst: Software for meta-analysis of binary, continuous and diagnostic data. *BMC Medical Research Methodology, 9,* 80. doi:10.1186/1471-2288-9-80

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. *Lancet Oncology, 12,* 703–708. doi:10.1016/S1470-2045(10)70294-1

Zachariae, R., & Mehlsen, M. Y. (2011). Is chemotherapy associated with cognitive impairment? *Nature Reviews Urology, 8,* 182–183. doi:10.1038/nrurol.2011.29

Received March 12, 2013
Revision received December 17, 2013
Accepted December 22, 2013 ■