The relevance of online cognitive assessment in oncology

The Amsterdam Cognition Scan

Lee Meeuw Kjoe, P.R.

Publication date
2023

Document Version
Final published version

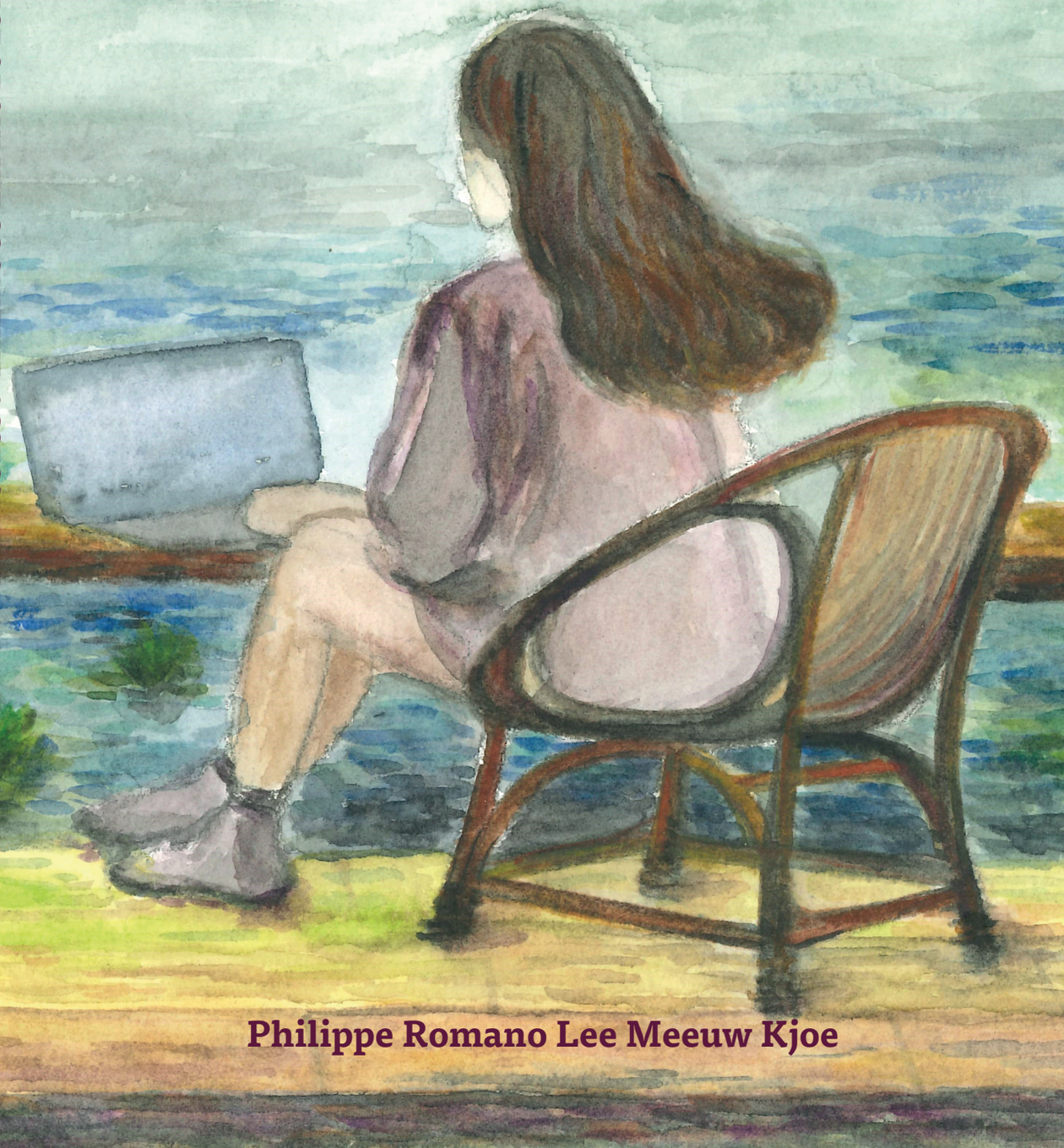
THE RELEVANCE OF ONLINE COGNITIVE ASSESSMENT IN ONCOLOGY: THE AMSTERDAM COGNITION SCAN

Philippe Romano Lee Meeuw Kjoe

9 789464 731262 >

THE RELEVANCE OF ONLINE COGNITIVE ASSESSMENT IN ONCOLOGY:

THE AMSTERDAM COGNITION SCAN



Philippe Romano Lee Meeuw Kjoe

THE RELEVANCE OF ONLINE COGNITIVE ASSESSMENT IN ONCOLOGY:

THE AMSTERDAM COGNITION SCAN

Philippe Romano Lee Meeuw Kjoe

MIX
Paper | Supporting responsible forestry
FSC® C128610
FSC www.fsc.org

The work presented in this thesis was performed at the Division of Psychosocial Research and Epidemiology of the Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital Amsterdam.

This work was supported by the Dutch Cancer Society, KWF Kankerbestrijding [grant number KWF 2015-7937]. The baseline and short-term follow-up measurement of the TEAM study described in chapter 6 were supported by an independent research grant from Pfizer.

Cover illustration by: Fiona Shan
Design and lay-out by: Ipskamp Printing, Enschede

The icons in Figure 4 and 8 in Chapter 3 are downloaded from the website www.flaticon.com. Icons are made by Freepik, juicy_fish, monkik, Darius Dan, Pixel perfect, max.icons, Roundicons Premium, ultimatearm, Adib Sulthon, Chanut-is-Industries, Smashicons, DinosoftLabs, berkahicon, Vectors Market, Good Ware, turkkub, photo3idea_studio from www.flaticon.com.

Printed by: Ipskamp Printing, Enschede

ISBN: 978-94-6473-126-2

THE RELEVANCE OF ONLINE COGNITIVE ASSESSMENT IN ONCOLOGY:

THE AMSTERDAM COGNITION SCAN

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit van Amsterdam op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek

ten overstaan van een door het College voor Promoties ingestelde commissie, in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 20 juni 2023, te 12.00 uur

door Philippe Romano Lee Meeuw Kjoe
geboren te Amsterdam

Promotiecommissie

Promotores:	prof. dr. S.B. Schagen	Universiteit van Amsterdam
	prof. dr. E. van der Wall	Universiteit Utrecht
Copromotores:	dr. I.E. Vermeulen	Vrije Universiteit Amsterdam
Overige leden:	prof. dr. J.C.M. van Weert	Universiteit van Amsterdam
	prof. dr. E.H.F. de Haan	Universiteit van Amsterdam
	prof. dr. G.S. Sonke	AvL-NKI
	prof. dr. Y. Wengström	Karolinska Institutet
	dr. A. Compter	AvL-NKI
	prof. dr. N.J. de Wit	Universiteit Utrecht

Faculteit der Maatschappij- en Gedragswetenschappen

Contents

Chapter 1:
General Introduction

Portrait: *A. is a 48-year-old woman who was diagnosed with breast cancer. After successful treatment with radiation and chemotherapy, she started treatment with tamoxifen for the next 10 years and was told she was ready to return to normal life. At first, A. was relieved to return to normal life including work, but soon she did not recognize her past self. She regularly forgets about appointments, she is having difficulty to read, and she is less able to express herself.*

What A. describes in the portrait above, is what neuropsychologists call problems with 'cognitive functions': mental processes such as thinking, concentrating and planning.[1] Cognitive functions are conventionally measured by neuropsychological tests which are standardized tests that tap into a certain cognitive function, such as memory or attention tests. During recent years, there is increasing interest in computerized/online testing, both in clinical and research settings,[2] as online tests are more time-efficient, user-friendly and cost-efficient than traditional paper-and-pencil tests.[3] However, most online tests lack adequate psychometric properties and norms. Therefore, the Amsterdam Cognition Scan (ACS) was developed several years ago —a new validated set of online neuropsychological tests.[4-5] The goal of this dissertation is to further optimize the ACS and implement cognitive tests in oncological research.

The ACS consists of seven online versions of well-established classic neuropsychological tests measuring attention, information processing speed, learning and memory, executive functioning, and psychomotor speed.[4] The test can be completed by patients at home without supervision and takes around one hour to complete. The minimum requirements to perform the ACS are a computer with internet connection —no downloads are needed. Normative data are collected.[5] Currently, a Dutch, American- and British-English and a Swedish version of the ACS is available.

The ACS is originally designed for research in oncology, as this field is in need of efficient large-scale cognitive assessment in patients with and survivors of cancer.[6] An increasing number of studies show that cancer and

cancer treatment is associated with cognitive impairment.[6] Studies often report decline in information processing speed, learning and memory, and lower executive function from pre- to post-treatment. The number of cancer patients is growing, as well as the community of cancer survivors, making cancer survivorship an important part of health care. Due to heterogeneity in individual, disease and treatment characteristics, large-scale data collection is needed, but this is difficult to achieve using traditional paper-and-pencil testing. Because of the online nature of the ACS, the ACS could play an important role in the field of cancer and cognition.

Prior to large-scale international implementation of the ACS, the ACS should be further developed. First, it should be examined how to correct for computer experience when analyzing performance on the ACS, as a previous study showed that more computer experience was associated with better ACS performance.[5] Secondly, before the ACS can be translated to other languages, cross-lingual criteria for its verbal learning test should be created, as direct translation of language-based tests is not recommended.[7-8] Lastly, modern data-driven methods should be used to reliably differentiate cognitively affected and unaffected patients, which is a point of concern in the field.[9]

Subsequently, the ACS could be implemented in studies on cognitive effects of cancer, cancer treatment and prevention. Up till now, research has focused mainly on the cognitive effects of chemotherapy,[10] but given the arrival of new (combinations) of treatments other than chemotherapy, and better survival rates of cancer patients, effects of other systemic therapies also become relevant. One of the most frequently used therapies in breast —and prostate— cancer is endocrine therapy (ET), in both early stage disease, so-called (neo-)adjuvant setting, as in metastatic disease. ET for breast cancer interacts with the function of estrogens, e.g., agents such as tamoxifen, or reduces estrogen levels, e.g., agents such as exemestane or surgical removal of the ovaries, so-called oophorectomy. Since studies have shown that estrogens play an important role in several biological systems

associated with normal cognitive function, it is likely that ET may have cognitive adverse effects.[11] Since more recent years, duration of ET-use can be as long as 10 years in early stage hormone-sensitive breast cancer.[12] In addition, ET-combination-therapies with other promising agents, such as cyclin-dependent kinases 4 and 6 (CDK4/6) inhibitors,[13] are increasingly implemented. Furthermore, the uptake of premenopausal risk-reducing salpingo-oophorectomy (RRSO) to reduce ovarian cancer risk is increasingly high in women at high familial risk for ovarian cancer.[14] In view of the earlier mentioned data of Zwart and colleagues (2015),[11] these developments warrant the examination of cognitive effects of ET to provide patients with the best possible care. If research with the ACS in patients with ET can provide us with data on cognitive effects of ET, this can help cancer specialists to properly inform patients, like patient A, about potential cognitive adverse effects at the start of ET.

Aims and outline of this thesis

The aims of the current thesis are: 1) further development of the ACS and 2) implementation of cognitive tests in studies on cognitive effects of endocrine therapy.

Part I: Further development of the Amsterdam Cognition Scan

In **Chapter 2,** a study is described in which it is examined how to correct for computer experience when analyzing online cognitive assessments by examining the influence of computer experience on both the ACS and its equivalent traditional paper-and-pencil cognitive tests.

In **Chapter 3,** to facilitate the development of new versions of the memory subtest of the ACS within and between countries, cross-lingual word criteria are presented for new Rey Auditory Verbal Learning Test(-based) word lists.

In **Chapter 4**, a study is described that examined the existence of cognitive subgroups in breast cancer patients who were treated with chemotherapy using a data-driven approach of ACS test scores.

Part II: Implementation of cognitive tests in studies on cognitive effects of endocrine therapy

In **Chapter 5,** an update of the literature is presented on the cognitive adverse effects of ET and CDK4/6 inhibitors in patients with hormone-receptor positive breast cancer (HR+BC).

In **Chapter 6,** short- and long-term cognitive adverse effects of tamoxifen and exemestane are investigated in patients with HR+ BC using traditional cognitive tests.

In **Chapter 7,** the long-term impact of premenopausal risk-reducing salpingo-oophorectomy (RRSO) is described on cognitive functioning by assessing tested cognition using the ACS and self-reported cognition in women at increased risk for ovarian cancer.

In **Chapter 8,** a summary of the findings of the aforementioned studies is presented including critical considerations. Future directions for research are given.

References

1. Lezak MD, Howieson DB, Loring DW, Fischer JS. *Neuropsychological assessment*. New York, NY: Oxford University Press; 2004.
2. Diaz-Orueta U, Blanco-Campal A, Lamar M, Libon DJ, Burke T. Marrying past and present neuropsychology: is the future of the process-based approach technology-based? *Front Psychol*. 2020;11:361.
3. Feenstra HEM, Vermeulen IE, Murre JMJ, Schagen SB. Online cognition: factors facilitating reliable online neuropsychological test results. *Clin Neuropsychol*. 2017;31(1):59-84.
4. Feenstra HEM, Murre JMJ, Vermeulen IE, Kieffer JM, Schagen SB. Reliability and validity of a self-administered tool for online neuropsychological testing: the Amsterdam Cognition Scan. *J Clin Exp Neuropsychol*. 2018;40(3):253-273.
5. Feenstra HEM, Vermeulen IE, Murre JMJ, Schagen SB. Online Self-Administered Cognitive Testing Using the Amsterdam Cognition Scan: establishing Psychometric Properties and Normative Data. *J Med Internet Res*. 2018;20(5):e192.
6. Schagen SB, Tsvetkov AS, Compter A, Wefel JS. Cognitive adverse effects of chemotherapy and immunotherapy: are interventions within reach? *Nat Rev Neurol*. 2022;18(3):173-185.
7. Madan CR. Exploring word memorability: how well do different word properties explain item free-recall probability? *Psychon Bull Rev*. 2021;28(2):583-595.
8. Djordjevic J, Jones-Gotman M. Neuropsychological assessment of memory in patients with epilepsy. In: Zeman A, Kapur N, Jones-Gotman M, eds. *Epilepsy and memory*. Oxford, UK: Oxford University Press; 2012:177-188.
9. Bernstein LJ, McCreath GA, Komeylian Z, Rich JB. Cognitive impairment in breast cancer survivors treated with chemotherapy depends on control group type and cognitive domains assessed: a multilevel meta-analysis. *Neurosci Biobehav Rev*. 2017;83:417-428.
10. Wefel JS, Kesler SR, Noll KR, Schagen SB. Clinical characteristics,

pathophysiology, and management of noncentral nervous system cancer-related cognitive impairment in adults. *CA Cancer J Clin*. 2015;65(2):123-138.
11. Zwart W, Terra H, Linn SC, Schagen SB. Cognitive effects of endocrine therapy for breast cancer: keep calm and carry on? *Nat Rev Clin Oncol*. 2015;12(10):597-606.
12. Burstein HJ, Temin S, Anderson H, et al. Adjuvant endocrine therapy for women with hormone receptor–positive breast cancer: American Society of Clinical Oncology clinical practice guideline focused update. *J Clin Oncol*. 2014;32(21):2255-2269.
13. Cristofanilli M, Turner NC, Bondarenko I, et al. Fulvestrant plus palbociclib versus fulvestrant plus placebo for treatment of hormone-receptor-positive, HER2-negative metastatic breast cancer that progressed on previous endocrine therapy (PALOMA-3): final analysis of the multicentre, double-blind, phase 3 randomised controlled trial. *Lancet Oncol*. 2016;17(4):425-439.
14. van Driel CM, de Bock GH, Arts HJ, et al. Stopping ovarian cancer screening in BRCA1/2 mutation carriers: effects on risk management decisions & outcome of risk-reducing salpingo-oophorectomy specimens. *Maturitas*. 2015;80(3):318-322.

Part I:

Further development of the Amsterdam Cognition Scan

Chapter 2:
How to correct for computer experience in online cognitive testing?

Philippe R. Lee Meeuw Kjoe, Joost A. Agelink van Rentergem,
Ivar E. Vermeulen, Sanne B. Schagen.

Assessment. 2021;28(5):1247-1255.

Abstract

Objective: Since computerized cognitive test performance may be influenced by computer experience, correction for this measure might be needed. This study examined how to correct for computer experience by examining its influence on online and traditional tests.

Methods: Two hundred forty-eight healthy adults completed an online neuropsychological test battery. Seventy adults completed traditional equivalents of the tests. Computer experience was assessed by a performance-based and a self-report measure. Regression analyses were applied to examine their influence on the online and traditional tests.

Results: After correction for demographics, the performance-based measure was associated with online and traditional, predominantly speed-based, tests. The self-report measure was also associated with speed-based online tests but not with most traditional tests.

Conclusions: Correcting computerized neuropsychological tests using a performance-based measure of computer experience would be unwise, because this measure also seems to tap into cognitive functions. A correction using a self-report measure might be better and is appropriate.

Background

As more and more people have access to and experience with computers, computerized cognitive testing, more specifically, online testing, is increasingly being used for the detection of cognitive impairment in both research and clinical practice.[1-4]

Computerized tests hold several advantages over traditional paper-and-pencil tests, such as more fine-grained response measurements,[5-7] flexibility in time and location, and cost-efficiency.[8-13] As such, computerized testing is becoming more and more important, not only for diagnosis, but also for evaluation of treatments, in many fields such as traumatic brain injury,[14] stroke,[15] dementia,[16] multiple sclerosis[17] and neuro(oncology).[18]

However, implementation of computerized/online cognitive testing remains quite limited,[19] amongst others due to differences in level of computer experience. Although computer use has increased vastly and still is growing, there are still parts of the population with limited access to and experience with computers. A 2015 US community survey showed that elderly people, persons with lower income, and those who are less educated are less likely to own a desktop or laptop,[20] resulting in less computer experience in these subpopulations.

A concern in computerized cognitive testing is that differences in computer experience lead to differences in performance on cognitive tests regardless of cognitive status. Especially for tests that require speed and coordination, more experienced computer users are potentially at an advantage. When differences in cognitive test performance also reflect differences in computer experience, the validity of computerized tests is threatened. Using computerized tests for clinical decision making could lead to overdiagnosis in people with limited computer experience and underdiagnosis in people with high computer experience.

When tests are used to detect cognitive impairment, it is good practice to use norms that are demographically corrected, e.g., for age, education level, and sex.[21] By using demographically corrected norms one obtains a more accurate estimate of premorbid functioning.[22] It might be that in computerized testing, normative comparisons would be more accurate if we would correct for computer experience as well.

Problematically, however, there is no clear consensus on an operational definition of computer experience, and as such no empirical consensus has been reached on its influence on computerized cognitive tests.[23-29] Studies often measure computer experience as a global construct and assess experience by self-report frequency or history of general computer use.[30] However, operationalizing computer experience as a global measure does not account for experience with specific actions, such as using a mouse (e.g., in computer gamers), or typing on a keyboard (e.g., in office workers). We recently suggested to apply a performance-based computer skills measure, assessing one's speed of keyboard typing, mouse clicking, and mouse dragging skills.[24] However, due to its reliance on psychomotor speed and motor coordination, this measure might also tap into cognitive domains. Correcting cognitive performance for this measure would in that case mean that one is essentially correcting a measure for itself, making the analysis unsound.

One way to determine whether and how computer experience influences computerized cognitive tests, is to also examine its influence on traditional paper-and-pencil test performance. If, for example, a computer experience measure correlates not only with computerized test performance but also with paper-and-pencil test performance, it can be argued that such a computer experience measure taps into cognitive domains and that using it to correct test performance would be unwise. If, in contrast, a computer experience measure correlates only with computerized test performance, it can be argued that one *should* use it to correct test performance and that it should be included in the norms. So far, hardly any research has been

done relating computer experience and paper-and-pencil test performance, but see Fazeli and colleagues (2012)[23] and Feenstra and colleagues (2018),[25] suggesting tentative positive relationships between computer experience and paper-and-pencil test scores.

Therefore, in sum, the aim of the current study is to investigate how to correct for computer experience by examining the influence of performance-based vs. self-report measures of computer experience on both online computerized neuropsychological test performance and traditional paper-and-pencil neuropsychological test performance.

Methods

We compared the results of two samples. One sample completed a self-administered online neuropsychological test battery: the Amsterdam Cognition Scan (ACS)[24] and the other sample completed a traditional neuropsychological test battery. In both samples, performance-based and self-report measures of computer experience were administered. See Table 1 for descriptive statistics from both samples.

Table 1

Demographics of study population

Demographics	Online data (N=248)	Traditional data (N=70)
Women, n (%)	157 (63.3%)	46 (65.7%)
Age (years), M (SD)	49.2 (13.0)	37.3 (16.1)
Education level[a]		
Low, n (%)	73 (29.4%)	19 (27.1%)
High, n (%)	175 (70.6%)	51 (72.9%)

Abbreviations: M = mean. SD = standard deviation.
[a]Education is based on Verhage's classification 1 to 7.[31] Low = Verhage 1 to 5; and High = Verhage 6 to 7.

Participants

For the online tests, we used existing data from the normative study of the ACS,[25] for which 248 Dutch-speaking adults were recruited via patients who participated in the validation study of the ACS.[24] Inclusion criteria were: 1) sufficient proficiency of the Dutch language, 2) basic computer skills (being able to use the computer mouse and send emails autonomously), and 3) access to a computer with a mouse, sound and internet connection. Exclusion criteria were: 1) history of cancer, and 2) self-reported medical conditions possibly influencing cognition (e.g., schizophrenia, psychosis, clinical depression, substance dependence, or brain pathology), and 3) completion of a neuropsychological assessment in the last five years.

For the traditional tests, new data were collected from 70 Dutch-speaking adults. Participants were recruited via flyers, social media and word of mouth referral among participants, researchers and volunteers at the Netherlands Cancer Institute. Inclusion criteria were: 1) age between 18 and 76, 2) sufficient proficiency of the Dutch language, and 3) basic computer skills (as defined earlier). Exclusion criteria were: 1) history of cancer, 2) self-reported medical conditions possibly influencing cognition, and 3) completion of a neuropsychological assessment in the last five years.

Ethical approval was given by the review board of the Netherlands Cancer Institute conform ethical guidelines for human experimentation stated in the Declaration of Helsinki (reference approval number: NL37964.031.11). All participants provided written, informed consent before participation.

Self-reported computer experience

Participants were asked to rate their level of computer experience by answering two questions: "How many years have you been using the computer?" and "How many hours per week do you use the computer?". This led to two variables: 1) years of computer experience, with scores ranging on

a continuous scale and 2) computer use per week, with scores indicated as 1 (0-5 hours), 2 (5-15 hours), 3 (15-35 hours) and 4 (>35 hours a week). Computer use per week was binned into four categories to deal with the fact that the amount of hours of computer use per week might vary greatly across weeks.

Performance-based computer skills

Computer skills were assessed via three computerized tests measuring 1) keyboard type skills (Type skills), 2) mouse click skills (Click Skills), and 3) mouse drag skills (Drag Skills) that were developed in the context of the ACS.

1) Type skills

This subtest measured speed of typing on a computer keyboard. Participants were asked to retype a target sentence that is presented on the computer screen as quickly as possible. Performance was measured in number of milliseconds (so, higher scores indicated worse performance).

2) Click Skills

This subtest measured speed of clicking via the chosen input device (e.g., mouse or touchpad). A number of circles forming a spiral were presented. The circles were placed of descending size from the outside to the inside of the spiral. Participants were asked to click the circles, beginning from the outside, moving to the inside as quickly as possible. Performance was measured in number of milliseconds. A time limit was set on three minutes.

3) Drag Skills

This subtest measured speed of dragging and dropping shapes on the screen with the mouse. For eight trials, two shapes were presented on the screen, one in black, the other in white, and slightly larger than the black shape. Participants were asked to drag the black shape into the white shape and

drop it in such a way that the black shape did not overlap with the white one. Performance was measured in number of milliseconds. A time limit was set on three minutes.

Online and traditional neuropsychological test battery

The ACS consisted of seven computerized neuropsychological tests, which are all based on well-established traditional neuropsychological tests.[24] The ACS was designed for the oncology setting, but is equally suitable for other settings, as it measures a variety of cognitive domains. Assessments took place in an unmonitored setting, either at home or at other private locations. Participants were asked to complete the online test in a single session and in a quiet room without distractions. The entire online test took about one hour (on average 56 minutes) to complete, including two fixed, standardized breaks and two questionnaires (Hospital Anxiety and Depression Scale[32] and Multidimensional Fatigue Inventory,[33] which were not used in the current study). The ACS was designed in such a way that it was suitable for all major Internet browsers and common operating systems. No software needed to be downloaded and cognitive measurements were independent of Internet speed. The ACS was shown in Dutch (English and Swedish versions are also available, while French, German and Spanish versions are currently under development).

The traditional neuropsychological test battery comprised seven widely used tests on which the online test battery was based. Six of those were paper-and-pencil tests that were administered face-to-face, one test was a computerized test (Visual Reaction Time – subtest FePsy). Assessments took place in a quiet test room at the hospital under supervision of a research assistant without any breaks in between. The assessment took about one hour and 15 minutes to complete.

See Table 2 for all online and traditional cognitive tests along with tested cognitive domains and main outcome measures.

Table 2

Online and traditional tests

Online tests	Traditional tests	Cognitive Domains	Main outcome measures
1) Connect the Dots I and II	Trail Making Test A and B	Visuomotor tracking, planning, cognitive flexibility, divided attention	Completion time in seconds (I and II)
2a) Wordlist Learning	Dutch Rey Auditory Verbal Learning Test	Verbal learning	Total number of correct words (trial 1 to 5)
3) Reaction Speed	Visual Reaction Time (subtest FePsy)	Information processing speed and attention	Mean reaction time in seconds
4) Place the Beads	Tower of London, Drexel University	Planning, response inhibition, visuospatial memory	Total number of extra moves
5) Box Tapping	Corsi Block-tapping Test	Visuospatial short-term memory	Total number of correctly repeated sequences
6) Fill the Grid	Grooved Pegboard	Fine motor skills	Completion time in seconds
2b) Wordlist Delayed Recall and Recognition	Dutch Rey Auditory Verbal Learning Test	Retention of information: free recall and recognition	Total number of correct words; free recall and recognition
7) Digit Sequences I and II	WAIS III Digit Span forward and backward	I: attention II: working memory	Total number of correctly repeated sequences forward and backward

NOTE. Half of the participants completed the Dutch RAVLT, while the other half completed another memory task. For analysis of memory, we used only data from the Dutch RAVLT.

Trail Making Test A and B[34]; Dutch Rey Auditory Verbal Learning Test: 15 Words Test[35]; Visual Reaction Time (subtest FePsy)[36]; Tower of London, Drexel University[37]; Corsi Block-tapping Test[38]; Grooved Pegboard[39]; WAIS III Digit Span forward and backward.[40]

Education

Education was based on Verhage's classification 1 to 7,[31] which was comparable to the International Standard Classification of Education[41] and corresponded with the following U.S. years of education: 1: 1-5 years; 2: 6 years; 3: 7-8 years; 4: 7-9 years; 5: 7-10 years; 6: 7-16 years; and 7: 17-20 years.

We transformed the score into a dichotomous high-low score: 0=low (Verhage 1 to 5), 1=high (Verhage 6 and 7). This was done to prevent assumption violations due to low numbers of cases in the first three education levels.

Data analysis

For all statistical analyses, SPSS Statistics for Windows, version 22.0 (IBM Corp., Armonk, NY) was used. Probabilities of $P < .05$ were considered statistically significant.

Variables were tested for normality with the Kolmogorov-Smirnov test. Depending on the distribution of the data, parametric or non-parametric tests were applied.

Outliers

Outliers on the tests measuring computer skills and the neuropsychological test scores were excluded. For reaction time test scores, the median absolute deviation (MAD) method was used.[42] This method makes use of the MAD, which takes the absolute difference between each observation and the median, and then calculates the median of these absolute differences. Observations are labeled as outlier when they are more than 3 MAD's times a constant scale factor (1.4826) above or below the median of the variable. Because of its reliance on medians, MAD is considered a superior method of outlier detection than methods relying on means.[42] To take age-related influences into account, MADs were calculated and applied separately in

the online sample (≤40 years, 41-59 years, and ≥60 years) and the traditional sample (≤40 years and >40 years). For test scores that measure number of correct responses and for which scores of zero are highly unlikely in a healthy group of participants, zero scores were excluded.

Influence of computer experience on online and traditional tests

A series of hierarchical multiple regression analyses (MRA) were performed, whereby we separately regressed the online and traditional cognitive test outcomes on the performance-based and self-report measures of computer experience, correcting for demographics. We used pairwise deletion to prevent losing data points from a case with an outlier. For all regression models, independent variables were entered blockwise. In the first block, we entered the variables sex (0=women, 1=men), age and education (0= low, 1= high). In the second block, we included the predictor computer experience. We first performed Spearman's Rank Order Correlations to examine which computer experience measure was significantly associated with online and traditional tests. Preliminary analyses were conducted to check assumptions of multicollinearity, homoscedasticity and independence of residuals by checking whether any of the analyses included variance inflation factors (VIFs) above 10, and visual inspection of residual histograms, p-p plots and residual-predicted values scatter plots, respectively.

Results

Outlier removal

For the online sample, six participants were excluded from analyses on the Wordlist Learning test and one participant was excluded from analyses on Digit Sequences II, as these participants indicated to have used unauthorized aids, such as a notepad. Fifty-nine scores (1.7% of all scores) were identified as outliers and excluded from analyses.

For the traditional sample, 19 scores (1.9% of all scores) were identified as outliers and excluded from analyses.

All removed scores were low performance scores that were unlikely to reflect true test performance in a sample of healthy individuals. These scores pointed to either incomprehension of test instructions, motor problems (6/19 observations in the traditional sample were from one participant who reported difficulties due to rheumatism), or possibly a moment of distraction. The vast majority of removed test scores were speed-based scores (19/19 from the traditional sample and 45/59 from the online sample), which were, especially in the unmonitored setting, vulnerable to small moments of distraction and motor problems.

Computer experience scores and demographics

Spearman's ρ correlations among the three computer skills tests showed that these were all significantly associated with one another in both the online and traditional sample (see Table S1). A composite score for the computer skills was constructed by calculating the mean of the reversed standardized completion times of the type, click and drag skills. From here on, all further analyses involve this performance-based composite measure of computer skills.

Spearman's ρ correlations showed that computer skills were strongly negatively associated with age – i.e., younger people showed better computer skills. Spearman's ρ correlations showed that both self-report measures of computer experience were associated with demographics, whereby men and higher-educated people generally reported more years of computer experience and hours of computer use per week, and younger people reported more hours of computer use and fewer years of computer experience (see Table S2).

Demographics and cognitive test results

Spearman's ρ correlations showed that age was negatively associated with performance on approximately all online tests as well as on several traditional tests, and that sex and education were associated with several online and traditional tests. See Table S3 and S4 for these results.

Selection of computer experience measures and cognitive tests for MRA

Spearman's ρ correlations showed that better computer skills were associated with better performance on most online and traditional tests – several correlations exceeded .30. Years of computer experience was not associated with any of the online and traditional tests (except for Corsiblock in the offline sample, r_s = -.29, p < .05). More hours of computer use per week were associated with better performance on several online tests, but not with any of the traditional tests (see Table S3 and S4). Based on these analyses, we selected performance-based computer skills and self-reported hours of computer use per week as predictors for the multiple regression analyses.

Influence of computer experience and computer skills on cognitive performance

Figure 1 and 2 show the results from the multiple regression analyses. No clear violations of the assumptions for regression analysis were identified. We found that after correcting for demographics, better performance-based computer skills were associated with better performance on five of the eleven online test outcomes, and also with better performance on four of the eleven traditional test outcomes. Specifically, participants' computer skills were associated with their performance on both the online and traditional versions of the TMT A, FePsy and Grooved Pegboard. For the TMT B and Corsi Block-tapping test, computer skills were only related to performance on the online versions of the tests; for the RAVLT Delayed Recall test, performance-based computer skills were only related to the traditional version. See Table S5 and S6 for all results.

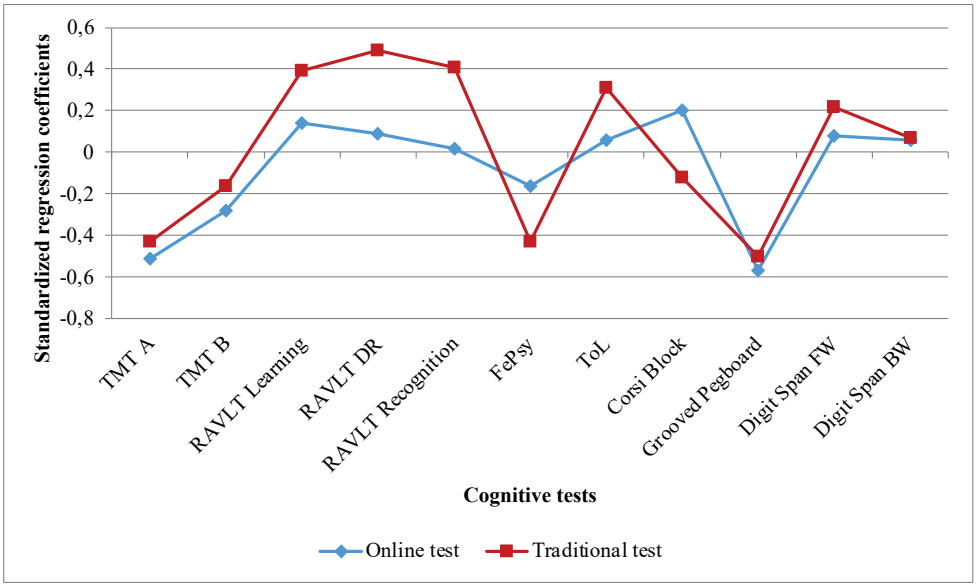


Figure 1. Regression coefficients MRA: performance-based computer skills and cognitive tests

This figure shows the results of the multiple regression analyses with performance-based computer skills as predictor corrected for demographic factors. On the x-axis, all cognitive tests are shown, from which raw outcome measures were used. On the y-axis, the standardized regression coefficients of performance-based computer skills are reported.
Positive regression coefficients show that better performance-based computer skills coincide with a higher raw test score, while negative coefficients show that better computer skills coincide with a lower raw test score.

As shown in Figure 2, more self-reported hours of computer use was associated with better performance on three of the eleven online test outcomes. Self-reported hours of computer use was associated with online versions of the TMT A, Corsi Block-tapping test, and Grooved Pegboard. No significant positive associations between this measure and performance on any of the traditional tests were observed; in fact, after correcting for demographics, self-reported hours of computer use correlated *negatively* to performance on the traditional Corsi Block-tapping test. See Table S7 and S8 for a detailed description of the results.

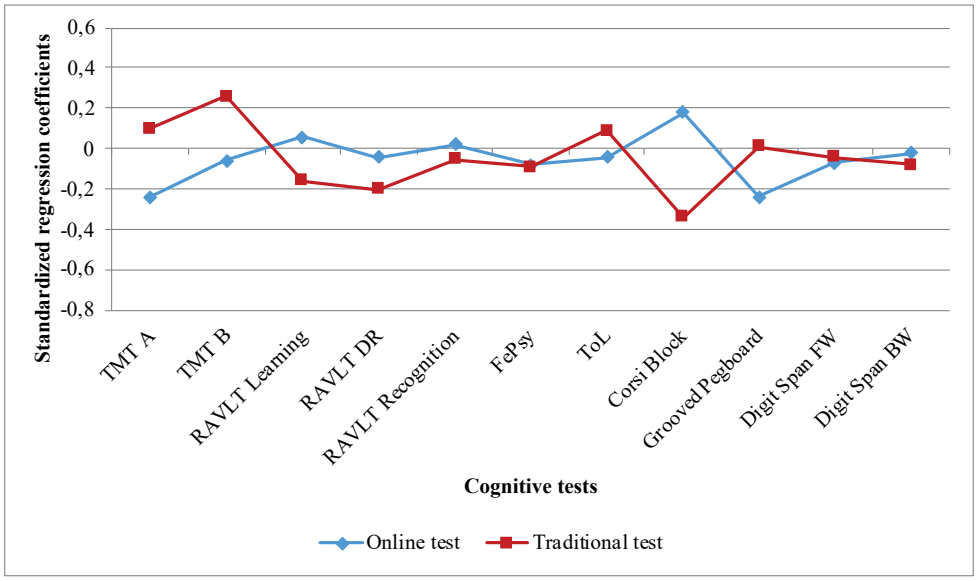


Figure 2. Regression coefficients MRA: self-reported hours of computer use and cognitive tests

This figure shows the results of the multiple regression analyses with self-reported hours of computer use as predictor corrected for demographic factors. On the x-axis, all cognitive tests are shown, from which raw outcome measures were used. On the y-axis, the standardized regression coefficients of hours of computer use are reported. Positive regression coefficients show that more self-reported hours of computer use coincide with a higher raw test score, while negative coefficients show that more hours of computer use coincide with a lower raw test score.

Conclusions

The aim of the current study was to investigate how to correct for computer experience in online cognitive testing by examining the influence of a performance-based and a self-report measure of computer experience on both online computerized test performance and traditional paper-and-pencil test performance.

After correction for sex, age, and education, better performance-based computer skills were associated with better performance on online

and traditional cognitive test outcomes, predominantly in the domains of motor coordination and information processing speed. Since our measure of computer skills is operationalized in terms of the speed with which one can perform computer actions, a relation between computer skills and our online tests is not a surprising finding. The fact that we also observe a relation between performance-based computer skills and performance on traditional neuropsychological tests that are theoretically independent of computer skills is less self-evident and suggests that our performance-based measure of computer skills also taps into certain domains of cognitive functioning.

We also found a higher number of self-reported hours of computer use per week to be associated with better performance on online cognitive test outcomes in the domains of motor coordination and processing speed, but not with the traditional cognitive test outcomes. This finding indicates that differences in performance on online cognitive tests can partly be explained by self-reported amount of hours of computer use per week, while differences in performance on traditional tests cannot. Additional analyses in the online sample stratified on age did not lead to other conclusions on the influence of performance-based computer skills and self-reported hours of computer use, which showed that our findings were not limited to older people (for results, see figures S1 and S2).

Based on these findings we suggest that correcting for computer experience in online cognitive testing is useful, predominantly for tests that measure motor coordination and processing speed. We suggest that in correcting for computer experience one should use a self-report measure. Performance-based computer skills do account for differential performance on certain computer actions, but are difficult to disentangle from motor coordination and processing speed. Self-reported hours of computer use is a more global measure but is not associated with tasks unrelated to computer experience. Additional support for favoring self-report measures of computer experience comes from the observation that, similar to cognitive

tests of motor coordination and processing speed, our performance-based measure of computer skills was strongly negatively associated with age (r = -.59-.62, p < .001 in both samples). In contrast, the correlation between age and self-reported hours of computer use was clearly weaker (r = -.28, p < .001 in the online sample, r = -.16, p = .20 in the traditional sample). Moreover, additional correlational analyses showed that there is merely a low to moderate positive correlation between the performance-based computer skills and self-reported hours of computer use (r = .38, p < .001 in the online sample, r = .13, p = .34 in the traditional sample), adding to the notion that both are two conceptually different measures. All in all, our results suggest that in order to correct computerized test results, using a self-report measure of computer experience (participants' self-reported computer use per week) would be the method of choice.

Clinical implications

The present results point to the need to correct for computer experience when using online computerized cognitive tests in clinical decision making. We suggest to do so using a self-report measure of computer experience: participants' self-reported computer use per week. It should be noted that because in particular patient groups current computer use can be affected by disorders, it may make sense to ask such patients for pre-morbid rather than current computer use.

Study limitations

A limitation of the current study was the age difference between the online and the traditional sample, leading to the possibility that the difference in the association between self-reported computer use and online and traditional tests could be due to age-related differences between the two samples. However, in the regression analyses we corrected for age. Also, we repeated the analyses in a sample where we matched the online sample on the traditional sample 1:1 with an age difference of one year (n=45, mean

age=42.9 years), which did not lead to major differences in results (see figures S9, S10 and S11 for detailed results). Another limitation was the difference in sample size between the online and the traditional sample. Basing conclusions on differences in statistical significance alone might be questionable, since these may be due to a difference in power. However, an examination of the magnitude of the regression coefficients led to the same conclusions. Also, the age-matching procedure equalized both sample sizes, but yielded no major differences in results. Another limitation was that analyses only have taken place in a sample of healthy controls. More research is needed to examine whether the results are also generalizable to clinical samples with impairment. It is also worth noting that the computerized tests produced more precise reaction time scores than the traditional tests, which possibly increased the chance of finding an association between these computerized scores and computer experience. Lastly, we made use of a binned self-reported measure of computer use per week. We do not know how well this measure correlated with actual hours of computer use. Therefore, in future research, it might be better to measure actual hours of computer use.

Conclusions

In sum, our study showed that correcting for (premorbid) amount of computer use per week could be useful in online cognitive testing.

References

1. Germine L, Reinecke K, Chaytor NS. Digital neuropsychology: challenges and opportunities at the intersection of science and software. *Clin Neuropsychol*. 2019;33(2):271-286.
2. Marcopulos B, Łojek E. Introduction to the special issue: are modern neuropsychological assessment methods really "modern"? Reflections on the current neuropsychological test armamentarium. *Clin Neuropsychol*. 2019;33(2):187-199.
3. Casaletto KB, Heaton RK. Neuropsychological assessment: Past and future. *J Int Neuropsychol Soc*. 2017;23(9-10):778-790.
4. Hansen TI, Haferstrom EC, Brunner JF, Lehn H, Haberg AK. Initial validation of a web-based self-administered neuropsychological test battery for older adults and seniors. *J Clin Exp Neuropsychol*. 2015;37(6):581-594.
5. Barak A, English N. Prospects and limitations of psychological testing on the Internet. *J Technol Hum Serv*. 2002;19(2-3):65-89.
6. Bilder RM. Neuropsychology 3.0: evidence-based science and practice. *J Int Neuropsychol Soc*. 2011;17(1):7-13.
7. Parsey CM, Schmitter-Edgecombe M. Applications of technology in neuropsychological assessment. *Clin Neuropsychol*. 2013;27(8):1328-1361.
8. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *Arch Clin Neuropsychol*. 2012;27(3):362-373.
9. Birnbaum MH. Human research and data collection via the Internet. *Annu Rev Psychol*. 2004;55:803-832.
10. Buchanan T, Smith JL. Using the Internet for psychological research: personality testing on the World Wide Web. *Br J Psychol*. 1999;90(1):125-144.
11. Caine C, Mehta MP, Laack NN, Gondi V. Cognitive function testing in adult brain tumor trials: lessons from a comprehensive review. *Expert Rev Anticancer Ther*. 2012;12(5):655-667.

12. Naglieri JA, Drasgow F, Schmit M, et al. Psychological testing on the Internet: new problems, old issues. *Am Psychol*. 2004;59(3):150-162.
13. Reips U-D. Internet-based psychological experimenting: five dos and five don'ts. *Soc Sci Comput Rev*. 2002;20(3):241-249.
14. Arrieux JP, Cole WR, Ahrens AP. A review of the validity of computerized neurocognitive assessment tools in mild traumatic brain injury assessment. *Concussion*. 2017;2(1):CNC31.
15. Gagnon MM, Laforce Jr R. Computerized vs. paper-pencil assessment of cognitive change following acute ischemic stroke. *J Neurol Disord*. 2016;4(8):317-322.
16. Chan JY, Kwong JS, Wong A, Kwok TC, Tsoi KK. Comparison of computerized and paper-and-pencil memory tests in detection of mild cognitive impairment and dementia: a systematic review and meta-analysis of diagnostic studies. *J Am Med Dir Assoc*. 2018;19(9):748-756.
17. Golan D, Wilken J, Doniger GM, et al. Validity of a multi-domain computerized cognitive assessment battery for patients with multiple sclerosis. *Mult Scler Relat Disord*. 2019;30:154-162.
18. Schagen S, Klein M, Reijneveld J, et al. Monitoring and optimising cognitive function in cancer patients: present knowledge and future directions. *Eur J Cancer Suppl*. 2014;12(1):29-40.
19. Bilder RM, Reise SP. Neuropsychological tests of the future: how do we get there from here? *Clin Neuropsychol*. 2019;33(2):220-245.
20. Anderson M. Technology Device Ownership: 2015. Pew Research Center; 2015. Available at: http://www.pewinternet.org/2015/10/29/technology-device-ownership-2015. [Accessed October 7th, 2019].
21. Strauss E, Sherman EM, Spreen, O. *A compendium of neuropsychological tests: administration, norms, and commentary.* 3rd ed. New York, NY: Oxford University Press; 2006.
22. Groth-Marnat G. *Handbook of psychological assessment*. 5th ed. Hoboken, NJ: John Wiley & Sons; 2009.
23. Fazeli PL, Ross LA, Vance DE, Ball K. The relationship between computer experience and computerized cognitive test performance among older adults. *J Gerontol B Psychol Sci Soc Sci*. 2012;68(3):337-346.

24. Feenstra HEM, Murre JMJ, Vermeulen IE, Kieffer JM, Schagen SB. Reliability and validity of a self-administered tool for online neuropsychological testing: the Amsterdam Cognition Scan. *J Clin Exp Neuropsychol*. 2018;40(3):253-273.
25. Feenstra HEM, Vermeulen IE, Murre JMJ, Schagen SB. Online self-administered cognitive testing using the Amsterdam Cognition Scan: establishing psychometric properties and normative data. *J Med Internet Res*. 2018;20(5):e192.
26. Iverson GL, Brooks BL, Ashton VL, Johnson LG, Gualtieri CT. Does familiarity with computers affect computerized neuropsychological test performance? *J Clin Exp Neuropsychol*. 2009;31(5):594-604.
27. Kuiper JS, Voshaar RCO, Verhoeven FE, Zuidema SU, Smidt N. Comparison of cognitive functioning as measured by the Ruff Figural Fluency Test and the CogState computerized battery within the LifeLines Cohort Study. *BMC Psychol*. 2017;5(15):1-12.
28. Lee JA. The effects of past computer experience on computerized aptitude test performance. *Educ Psychol Meas*. 1986;46(3):727-733.
29. Taylor C, Jamieson J, Eignor D, Kirsch I. *The relationship between computer familiarity and performance on computer-based TOEFL test tasks* (TOEFL Research Report 61). Princeton, NY: Educational Testing Service; 1998.
30. Smith B, Caputi P, Crittenden N, Jayasuriya R, Rawstorne P. A review of the construct of computer experience. *Comput Hum Behav*. 1999;15(2):227-242.
31. Verhage F. *Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar*. Assen, NL: van Gorcum; 1964.
32. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;67(6):361-370.
33. Smets E, Garssen B, Bonke B, De Haes, JCJM. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J Psychosom Res*. 1995;39(3):315-325.
34. Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. *Percept Mot Skills*. 1985;8(3):271-276.

35. van den Burg W, Saan R, Deelman B. *15-woordentest: Provisional Manual*. Groningen, NL: University Hospital, Department of Neuropsychology; 1985.
36. Alpherts W, Aldenkamp A. *FePSY:'The Iron Psyche'*. Heemstede, NL: Instituut voor epilepsiebestrijding; 1995.
37. Culbertson WC, Zillmer E. *The Tower of London DX (TOLDX) manual*. North Tonawanda, NY: Multi-Health Systems; 2001.
38. Kessels RP, van Zandvoort MJ, Postma A, Kappelle LJ, De Haan EH. The Corsi block-tapping task: standardization and normative data. *Appl Neuropsychol*. 2000;7(4):252-258.
39. Kløve H. *Grooved pegboard*. Lafayette, IN: Lafayette Instruments; 1963.
40. Wechsler D. *Wechsler Adult Intelligence Scale-III*. San Antonio, TX: The Psychological Corporation; 1997.
41. UNESCO Institute for Statistics. *International standard classification of education: ISCED 2011*. Montreal, Quebec: UNESCO Institute for Statistics; 2012.
42. Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J Exp Soc Psychol*. 2013;49(4):764-766.

Supplementary material

Spearman's ρ correlations between computer experience and demographic variables in the online and the traditional sample

	Online sample			Traditional sample		
	Sex	Age	Education	Sex	Age	Education
Performance-based computer skills	.05 (n=226)	-.59*** (n=226)	.04 (n=226)	.18 (n=60)	-.62*** (n=60)	.18 (n=60)
Self-reported years of computer experience	.17** (n=248)	.21** (n=248)	.16* (n=248)	.16 (n=70)	.67** (n=70)	.13 (n=69)
Self-reported hours of computer experience per week	.18** (n=235)	-.28** (n=235)	.24*** (n=235)	-.08 (n=68)	-.16 (n=68)	.54** (n=67)

NOTE. Performance-based computer skills is a reversed composite measure of the computer skills tests with higher scores indicating better computer skills. Sex (0=female, 1=male). Education (0=low, 1=high).
*$p < .05$. **$p < .01$. ***$p < .001$.

Table S3

Spearman's ρ correlation matrix: computer experience, demographic variables, and the online tests

Test	Demographics			Computer experience		
	Sex	Age	Education	Performance-based computer skills	Self-reported years of computer experience	Self-reported hours computer use per week
Connect the Dots I	-.15* (n=244)	.53** (n=244)	-.02 (n=244)	-.64** (n=225)	.03 (n=244)	-.38*** (n=231)
Connect the Dots II	-0.01 (n=243)	.60** (n=243)	-.11 (n=243)	-.54** (n=221)	.08 (n=243)	-.27*** (n=230)
Wordlist Learning	-.15* (n=247)	-.34** (n=247)	.10 (n=247)	.27*** (n=225)	-.04 (n=247)	.13 (n=234)
Wordlist Delayed Recall	-.11 (n=245)	-.22** (n=245)	.08 (n=245)	.21** (n=223)	-.07 (n=245)	.11 (n=232)
Wordlist Recognition	-0.05 (n=248)	-.24** (n=248)	-.08 (n=248)	.18** (n=226)	-.06 (n=248)	.03 (n=235)
Reaction Speed	-.07 (n=241)	.23** (n=241)	-.10 (n=241)	-.27** (n=221)	.01 (n=241)	-.17* (n=229)
Place the Beads	-.12 (n=248)	.19** (n=248)	-.09 (n=248)	-.13* (n=226)	-.07 (n=248)	-.13* (n=235)
Box Tapping	.19** (n=237)	-.30** (n=237)	.13* (n=237)	.36*** (n=216)	.07 (n=237)	.33*** (n=226)
Fill the Grid	-.13* (n=240)	.49** (n=240)	-.04 (n=240)	-.66** (n=221)	.02 (n=240)	-.37*** (n=227)
Digit Sequences I	.07 (n=248)	-.11 (n=248)	.15* (n=248)	.11 (n=226)	-.02 (n=248)	.03 (n=235)
Digit Sequences II	.09 (n=247)	-.20** (n=247)	.14* (n=247)	.18** (n=225)	.10 (n=247)	.08 (n=234)

NOTE. Test scores are raw cognitive test outcomes. For Connect the Dots I and II, Reaction Speed, Place the Beads, Fill the Grid, higher test scores indicate lower performance. For Wordlist Learning, Delayed Recall and Recognition, Box Tapping and Digit Sequences I and II, higher test scores indicate higher performance. Performance-based computer skills is a reversed composite measure of the computer skills tests with higher scores indicating better computer skills. Sex (0=female, 1=male). Education (0=low, 1=high). Hours computer use per week (1-4).

*$p < .05$. **$p < .01$. ***$p < .001$.w2

Table S4

Spearman's ρ correlation matrix: computer experience, demographic variables, and the traditional tests

Test	Demographics			Computer experience		
	Sex	Age	Education	Performance-based computer skills	Self-reported years of computer experience	Self-reported hours computer use per week
Trail Making Test A	.02 (n=67)	.33** (n=67)	-.07 (n=67)	-.47*** (n=59)	.11 (n=67)	-.01 (n=65)
Trail Making Test B	.16 (n=67)	.27* (n=67)	-.01 (n=67)	-.33* (n=59)	.12 (n=67)	-.13 (n=65)
RAVLT Learning	-.17 (n=38)	-.35* (n=38)	.21 (n=38)	.50** (n=31)	-.02 (n=38)	.01 (n=37)
RAVLT Delayed Recall	-.40* (n=38)	-.34* (n=38)	.05 (n=38)	.51** (n=31)	-.03 (n=38)	-.17 (n=37)
RAVLT Recognition	.02 (n=38)	-.17 (n=38)	-.21 (n=38)	.23 (n=31)	.11(n=38)	-.20 (n=37)
Visual Reaction Time	-.16 (n=69)	.22 (n=69)	-.09 (n=69)	-.44*** (n=60)	-.08 (n=69)	-.09 (n=67)
Tower of London	.03 (n=70)	-.00 (n=70)	-.06 (n=69)	.12 (n=60)	.15 (n=70)	-.02 (n=68)
Corsi Block-tapping Test	.08 (n=70)	-.44** (n=70)	.09 (n=69)	.32* (n=60)	-.29* (n=70)	-.09 (n=68)
Grooved Pegboard	.44*** (n=68)	.21 (n=68)	-.25* (n=68)	-.32* (n=60)	.06 (n=68)	-.18 (n=66)
WAIS III Digit Span forward	-.09 (n=70)	-.01 (n=70)	.35** (n=69)	.21 (n=60)	.04 (n=70)	.18 (n=68)
WAIS III Digit Span backward	.10 (n=70)	-.10 (n=70)	.08 (n=69)	.26* (n=60)	-.04 (n=70)	.01 (n=68)

Abbreviations: RAVLT, Rey Auditory Verbal Learning Test; WAIS III, Wechsler Adult Intelligence Scale, third edition.

NOTE. Test scores are raw cognitive test outcomes. For Trail Making Test A and B, Visual Reaction Time, Tower of London, Grooved Pegboard, higher test scores indicate lower performance. For RAVLT Learning, Delayed Recall and Recognition, Corsi Block-tapping Test and WAIS III Digit Span forward and backward, higher test scores indicate higher performance. Performance-based computer skills is a reversed composite measure of the computer skills tests with higher scores indicating better computer skills. Sex (0=female, 1=male). Education (0=low, 1=high). Hours computer use per week (1-4).

*$p < .05$. **$p < .01$. ***$p < .001$.

Table S5

Multiple regression analyses: performance-based computer skills and demographic variables, and the online cognitive tests

		Sex			Age			Education			Performance-based computer skills			Model
Model	Test	B	SE B	St. B	B	SE B	St. B	B	SE B	St. B	B	SE B	St. B	R^2
1	Connect the Dots I	-3.37**	1.08	-0.18	0.33***	0.04	.48	0.94	1.17	0.05				.25
2		-2.16*	0.97	-0.12	0.12**	0.05	0.18	1.47	1.04	0.07	-6.16***	0.79	-0.51	.41
1	Connect the Dots II	-1.88	1.89	-0.06	0.72***	0.07	0.57	-2.74	2.06	-0.07				.33
2		-0.72	1.85	-0.02	0.51***	0.09	0.40	-2.22	1.99	-0.06	-6.17***	1.53	-0.28	.38
1	Wordlist Learning	-2.44	1.30	-0.12	-0.27***	0.05	-0.34	1.22	1.41	0.05				.14
2		-2.81*	1.31	-0.14	-0.20**	0.06	-0.26	1.05	1.40	0.05	1.94	1.07	0.14	.15
1	Wordlist Delayed Recall	-0.68	0.35	-0.13	-0.04**	0.01	-0.22	0.46	0.37	0.08				.08
2		-0.74*	0.35	-0.14	-0.03*	0.02	-0.16	0.43	0.37	0.08	0.32	0.28	0.09	.08
1	Wordlist Recognition	-0.22	0.19	-0.08	-0.03**	0.01	-0.22	-0.38	0.21	-0.12				.07
2		-0.23	0.20	-0.08	-0.02*	0.01	-0.21	-0.38	0.21	-0.12	0.05	0.16	0.02	.07
1	Reaction Speed	-8.49	5.67	-0.10	0.77***	0.22	0.24	-5.62	6.16	-0.06				.07
2		-6.73	5.70	-0.08	0.45	0.27	0.14	-5.08	6.13	-0.05	-9.55*	4.75	-0.16	.08
1	Place the Beads	-3.01	1.91	-0.10	0.20**	0.07	0.18	-3.01	2.08	-0.10				.05
2		-3.24	1.94	-0.11	0.24*	0.09	0.21	-3.11	2.08	-0.10	1.18	1.58	0.06	.05
1	Box Tapping	0.93**	0.30	0.20	-0.05***	0.01	-0.29	0.43	0.32	0.09				.13
2		0.81**	0.30	0.17	-0.03*	0.01	-0.17	0.37	0.32	0.08	0.60*	0.24	0.20	.15
1	Fill the Grid	-3197.22*	1402.26	-0.14	433.63***	52.46	0.49	616.99	1532.58	0.02				.25
2		-1658.63	1209.79	-0.07	123.65*	56.46	0.14	1116.41	1310.20	0.04	-9088.37***	1007.06	-0.57	.45

Table continues

1	Digit Sequences I	0.33	0.30	0.07	-0.02*	0.01	-0.13	0.98**	0.33	0.20			.06
2		0.29	0.30	0.06	-0.02	0.01	-0.09	0.96**	0.33	0.20	0.23	0.25	0.08	.06
1	Digit Sequences II	0.50	0.36	0.09	-0.05***	0.01	-0.24	1.24**	0.40	0.20			.10
2		0.45	0.37	0.08	-0.04*	0.02	-0.20	1.22**	0.40	0.20	0.23	0.31	0.06	.10

Abbreviations: SE *B* = standard error of beta; St. *B* = standardized beta.

NOTE. Model 1 includes only demographic factors (block 1), i.e., sex, age and education. Model 2 includes demographics and performance-based computer skills (block 2). Performance-based computer skills is a reversed composite measure of the computer skills tests with higher scores indicating better computer skills. Sex (0=female, 1=male). Education (0=low, 1=high).

*$p < .05$. **$p < .01$. ***$p < .001$.

Table S6

Multiple regression analyses: performance-based computer skills and demographic variables, and the traditional cognitive tests

		Sex			Age			Education			Performance-based computer skills			Model
Model	Test	B	SE B	St. B	B	SE B	St. B	B	SE B	St. B	B	SE B	St. B	R^2
1	Trail Making Test A	0.94	1.76	0.07	0.15**	0.06	0.35	-0.96	1.99	-0.06				.14
2		1.49	1.69	0.11	0.03	0.07	0.07	-0.09	1.93	-0.01	-3.38*	1.34	-0.43	.23
1	Trail Making Test B	6.59	4.03	0.20	0.39**	0.12	0.40	-0.86	4.56	-0.02				.19
2		7.05	4.07	0.21	0.28	0.17	0.29	-0.13	4.64	-0.00	-2.82	3.17	-0.16	.20
1	RAVLT Learning	-2.59	2.98	-0.15	-0.25**	0.08	-0.50	3.74	2.95	0.21				.28
2		-2.69	2.86	-0.15	-0.12	0.11	-0.24	2.30	2.94	0.13	3.52	1.94	0.39	.36
1	RAVLT Delayed Recall	-1.90*	0.88	-0.35	-0.08**	0.02	-0.51	0.02	0.87	0.00				.33
2		-1.94*	0.81	-0.36	-0.03	0.03	-0.19	-0.53	0.83	-0.10	1.35*	0.55	0.49	.46
1	RAVLT Recognition	-0.44	0.45	-0.18	-0.02	0.01	-0.26	-0.20	0.45	-0.08				.10
2		-0.46	0.44	-0.19	0.00	0.02	0.01	-0.40	0.45	-0.17	0.50	0.30	0.41	.19
1	Visual Reaction Time	-10.38	8.04	-0.16	0.66*	0.25	0.34	3.95	8.93	0.06				.15
2		-7.91	7.76	-0.12	0.09	0.33	0.05	7.96	8.70	0.11	-15.23*	6.16	-0.43	.23
1	Tower of London	1.36	4.28	0.04	0.02	0.13	0.02	-1.65	4.76	-0.05				.00
2		0.48	4.26	0.02	0.22	0.18	0.23	-3.09	4.77	-0.09	5.45	3.38	0.31	.05
1	Corsi Block-tapping test	0.24	0.36	0.08	-0.04***	0.01	-0.46	0.14	0.40	0.04				.23
2		0.27	0.37	0.09	-0.05**	0.02	-0.54	0.19	0.41	0.06	-0.20	0.29	-0.12	.24

Table continues

1	Grooved Pegboard	7.72**	2.44	0.37	0.12	0.08	0.19	-6.05*	2.71	-0.26			.24
2		8.66***	2.28	0.42	-0.10	0.10	-0.16	-4.53	2.56	-0.20	-5.78**	1.81	-0.50	.36
1	WAIS III Digit Span forward	-0.50	0.49	-0.13	0.01	0.02	0.05	1.27*	0.54	0.30			.10
2		-0.58	0.49	-0.15	0.02	0.02	0.19	1.15*	0.55	0.27	0.46	0.39	0.22	.13
1	WAIS III Digit Span backward	0.26	0.53	0.07	-0.02	0.02	-0.14	-0.08	0.59	-0.02			.02
2		0.24	0.54	0.06	-0.01	0.02	-0.09	-0.12	0.60	-0.03	0.16	0.43	0.07	.03

Abbreviations: SE *B* = standard error of beta; St. *B* = standardized beta; RAVLT, Rey Auditory Verbal Learning Test; WAIS III, Wechsler Adult Intelligence Scale, third edition.

NOTE. Model 1 includes only demographic factors (block 1), i.e., sex, age and education. Model 2 includes demographics and performance-based computer skills (block 2). Performance-based computer skills is a reversed composite measure of the computer skills tests with higher scores indicating better computer skills. Sex (0=female, 1=male). Education (0=low, 1=high).

*$p < .05$. **$p < .01$. ***$p < .001$.

Table S7

Multiple regression analyses: self-reported hours computer use per week and demographic variables, and the online cognitive tests

		Sex			Age			Education			Self-reported hours of computer use per week			Model
Model	Test	B	SE B	St. B	B	SE B	St. B	B	SE B	St. B	B	SE B	St. B	R^2
1	Connect the Dots I	-3.26**	1.12	-0.16	0.39***	0.04	0.51	-0.17	1.18	-0.01				.28
2		-2.38*	1.11	-0.12	0.34***	0.04	0.44	1.01	1.18	0.05	-2.26***	0.57	-0.24	.33
1	Connect the Dots II	-3.87	1.99	-0.10	0.84***	0.08	0.59	-3.06	2.11	-0.08				.37
2		-3.50	2.03	-0.09	0.82***	0.08	0.58	-2.51	2.19	-0.06	-0.99	1.04	-0.06	.37
1	Wordlist Learning	-2.68*	1.29	-0.13	-0.30***	0.05	-0.37	1.11	1.36	0.05				.16
2		-2.93*	1.32	-0.14	-0.29***	0.05	-0.36	0.75	1.41	0.03	0.65	0.67	0.06	.17
1	Wordlist Delayed Recall	-0.65	0.35	-0.12	-0.05***	0.01	-0.26	0.36	0.37	0.06				.09
2		-0.69	0.36	-0.13	-0.05***	0.01	-0.25	0.30	0.38	0.05	0.12	0.18	-0.04	.09
1	Wordlist Recognition	-0.15	0.21	-0.04	-0.03***	0.01	-0.27	-0.32	0.22	-0.09				.09
2		-0.16	0.21	-0.05	-0.03***	0.01	-0.27	-0.34	0.23	-0.10	0.03	0.11	0.02	.09
1	Reaction speed	-11.05	5.71	-0.12	0.97***	0.22	0.29	-5.47	6.03	-0.06				.10
2		-9.65	5.83	-0.11	0.90***	0.23	0.26	-3.66	6.23	-0.04	-3.45	3.01	-0.08	.10
1	Place the beads	-3.84	1.99	-0.12	0.27***	0.08	0.23	-2.50	2.09	-0.08				.07
2		-3.59	2.03	-0.12	0.26**	0.08	0.22	-2.16	2.17	-0.07	-0.62	1.04	-0.04	.07
1	Box Tapping	0.89**	0.31	0.18	-0.05***	0.01	-0.28	0.48	0.32	0.09				.12
2		0.71*	0.31	0.15	-0.04***	0.01	-0.23	0.24	0.33	0.05	0.42**	0.16	0.18	.15

Table continues

1	Fill the grid	-2904.00*	1458.79	-0.11	512.05***	55.86	0.52	-1901.02	1555.93	-0.07			.28
2		-1722.77	1442.95	-0.07	444.85***	56.64	0.45	-297.21	1559.00	-0.01	-2983.07***	747.41	-0.24	.33
1	Digit Sequences I	0.55	0.30	0.12	-0.03*	0.01	-0.15	0.80*	0.32	0.16				.06
2		0.61*	0.31	0.13	-0.03*	0.01	-0.17	0.89**	0.33	0.18	-0.16	0.16	-0.07	.07
1	Digit Sequences II	0.69	0.37	0.12	-0.04**	0.01	-0.20	0.84*	-0.39	0.14				.07
2		0.71	0.37	0.12	-0.05**	0.01	-0.21	0.88*	0.40	0.14	-0.06	0.19	-0.02	.07

Abbreviations: SE B = standard error of beta; St. B = standardized beta.
NOTE. Model 1 includes only demographic factors (block 1), i.e., sex, age and education. Model 2 includes demographics and performance-based computer skills (block 2). Hours of computer use per week (1-4). Sex (0=female, 1=male). Education (0=low, 1=high).

*p < .05. **p < .01. ***p < .001.

Table S8

Multiple regression analyses: self-reported hours of computer use per week and demographic variables, and the traditional cognitive tests

		Sex			Age			Education			Self-reported hours of computer use per week			Model
Model	Test	B	SE B	St. B	B	SE B	St. B	B	SE B	St. B	B	SE B	St. B	R^2
1	Trail Making Test A	1.30	1.70	0.09	0.13*	0.05	0.31	-0.64	1.85	-0.04				.11
2		1.36	1.71	0.10	0.13*	0.05	0.32	-1.50	2.32	-0.10	0.64	1.04	0.10	.12
1	Trail Making Test B	5.72	3.97	0.17	0.30*	0.12	0.31	-2.00	4.29	-0.06				.14
2		5.91	3.90	0.18	0.32**	0.12	0.33	-6.77	4.94	-0.19	4.19	2.26	0.26	.18
1	RAVLT Learning	-1.95	2.49	-0.11	-0.27**	0.07	-0.53	4.38	2.56	0.24				.34
2		-1.81	2.50	-0.10	-0.28**	0.07	-0.55	6.01	3.08	0.33	-1.50	1.58	-0.16	.36
1	RAVLT Delayed Recall	-1.45	0.77	-0.27	-0.08**	0.02	-0.50	0.08	0.79	0.01				.34
2		-1.39	0.77	-0.26	-0.08**	0.02	-0.53	0.70	0.95	0.13	-0.57	0.48	-0.20	.37
1	RAVLT Recognition	-0.02	0.36	-0.01	-0.01	0.01	-0.22	-0.23	0.37	-0.10				.06
2		-0.02	0.37	-0.01	-0.01	0.01	-0.23	-0.17	0.45	-0.08	-0.05	0.23	-0.05	.07
1	Visual Reaction Time	-8.86	7.64	-0.14	0.53*	0.23	0.28	-5.71	8.13	-0.08				.10
2		-9.08	7.69	-0.14	0.52*	0.23	0.27	-2.28	9.78	-0.03	-2.90	4.56	-0.09	.11

Table continues

1	Tower of London	-0.36	4.02	-0.01	0.03	0.12	0.03	-2.35	4.28	-0.07			.01
2		-0.25	4.05	-0.01	0.03	0.12	0.04	-4.09	5.15	-0.12	1.48	2.40	0.09	.01
1	Corsi Block-tapping test	0.23	0.36	0.07	-0.04**	0.01	-0.41	0.35	0.39	0.11				.19
2		0.19	0.35	0.06	-0.04***	0.01	-0.44	0.97*	0.44	0.29	-0.52*	0.21	-0.34	.26
1	Grooved Pegboard	9.67***	2.31	0.43	0.17*	0.07	0.26	-5.36*	2.48	-0.23				.34
2		9.68***	2.34	0.43	0.17*	0.07	0.26	-5.45	3.05	-0.23	0.07	1.39	0.01	.34
1	WAIS III Digit Span forward	-0.62	0.43	-0.17	0.01	0.01	0.09	1.29**	0.46	0.33				.15
2		-0.63	0.44	-0.17	0.01	0.01	0.09	1.36*	0.56	0.35	-0.06	0.26	-0.04	.15
1	WAIS III Digit Span backward	0.46	0.51	0.11	-0.00	0.02	-0.04	0.35	0.54	0.08				.02
2		0.45	0.51	0.11	-0.01	0.02	-0.04	0.53	0.65	0.12	-0.15	0.30	-0.08	.02

Abbreviations: SE B = standard error of beta; St. B = standardized beta; RAVLT, Rey Auditory Verbal Learning Test; WAIS III, Wechsler Adult Intelligence Scale, third edition.
NOTE. Model 1 includes only demographic factors (block 1), i.e., sex, age and education. Model 2 includes demographics and performance-based computer skills (block 2). Hours of computer use per week (1-4). Sex (0=female, 1=male). Education (0=low, 1=high).

*p < .05. **p < .01. ***p < .001.

Table S9

Demographics of matched study population

Demographics	Matched online data (N=45)	Traditional data (N=70)
Women, n (%)	25 (55.6%)	46 (65.7%)
Age (years), M (SD)	42.9 (15.8)	37.3 (16.1)
Education level[a]		
Low, n (%)	12 (26.7%)	19 (27.1%)
High, n (%)	33 (73.3%)	51 (72.9%)

[a]Education is based on Verhage's classification 1 to 7 (Verhage, 1964). Low = Verhage 1 to 5; and High = Verhage 6 to 7.

Table S10

Multiple regression analyses: performance-based computer skills and demographic variables, and the online cognitive tests in matched sample

		Sex			Age			Education			Performance-based computer skills			Model
Model	Test	B	SE B	St. B	B	SE B	St. B	B	SE B	St. B	B	SE B	St. B	R^2
1	Connect the Dots I	-3.90	2.23	-0.22	0.32***	0.07	0.57	-0.52	2.54	-0.03				.37
2		-2.51	1.95	-0.14	0.11	0.08	0.20	0.93	2.22	0.05	-6.16***	1.60	-0.57	.55
1	Connect the Dots II	-3.35	3.41	-0.12	0.60***	0.11	0.66	-0.13	3.89	-0.00				.44
2		-2.75	3.48	-0.10	0.51**	0.15	0.56	0.51	3.95	0.02	-2.68	2.85	-0.15	.45

Table continues

1	Wordlist Learning	-0.30	2.94	-0.02	-0.17	0.10	-0.26	7.56*	3.35	0.33			.19
2		-1.30	2.90	-0.06	-0.02	0.12	-0.03	6.50	3.29	0.28	4.47	2.37	0.36	.26
1	Wordlist Delayed Recall	-0.83	0.80	-0.16	-0.02	0.03	-0.14	1.68	0.91	0.28			.13
2		-1.19	0.76	-0.22	0.03	0.03	0.18	1.31	0.86	0.22	1.59*	0.62	0.49	.26
1	Wordlist Recognition	-0.59	0.47	-0.20	-0.02	0.02	-0.17	-0.56	0.53	-0.16			.09
2		-0.64	0.48	-0.21	-0.01	0.02	-0.10	-0.60	0.55	-0.17	0.19	0.39	0.10	.10
1	Reaction speed	-12.25	10.02	-0.18	0.80*	0.34	0.35	-12.92	11.37	-0.17			.18
2		-10.77	10.36	-0.16	0.59	0.47	0.26	-11.84	11.58	-0.15	-5.49	8.57	-0.13	.19
1	Place the beads	-1.30	3.63	-0.05	0.17	0.12	0.22	-6.39	4.14	-0.23			.11
2		-1.47	3.74	-0.06	0.19	0.16	0.25	-6.57	4.25	-0.24	0.75	3.06	0.05	.11
1	Box Tapping	0.29	0.42	0.10	-0.03*	0.01	-0.36	0.18	0.48	0.06			.15
2		0.15	0.41	0.05	-0.01	0.02	-0.14	0.03	0.47	0.01	0.61	0.34	0.35	.21
1	Fill the grid	-5540.42	3150.46	-0.24	381.25**	100.34	0.52	-281.59	3662.48	-0.01			.32
2		-4071.30	2590.96	-0.18	59.78	109.39	0.08	-413.81	3011.68	0.05	-9719.93***	2194.57	-0.67	.56
1	Digit Sequences I	0.50	0.71	0.11	-0.02	0.02	-0.13	0.74	0.80	0.14			.05
2		0.61	0.72	0.14	-0.04	0.03	-0.25	0.86	0.82	0.17	-0.50	0.59	-0.18	.07
1	Digit Sequences II	0.59	0.74	0.10	-0.10***	0.02	-0.53	1.96*	0.86	0.29			.37
2		0.58	0.75	0.10	-0.10**	0.03	-0.53	1.96*	0.87	0.29	0.00	0.67	0.00	.37

Abbreviations: SE B = standard error of beta; St. B = standardized beta.
NOTE. Model 1 includes only demographic factors (block 1), i.e., sex, age and education. Model 2 includes demographics and performance-based computer skills (block 2).

*p < .05. **p < .01. ***p < .001.

Table S11

Multiple regression analyses: self-reported hours of computer use per week and demographic variables, and the online cognitive tests in matched sample

		Sex			Age			Education			Self-reported hours of computer use per week			Model
Model	Test	B	SE B	St. B	B	SE B	St. B	B	SE B	St. B	B	SE B	St. B	R^2
1	Connect the Dots I	-2.72	2.56	-0.14	0.38***	0.08	0.60	-2.34	2.83	-0.11				.39
2		-1.85	2.62	-0.09	0.36***	0.08	0.56	-1.04	2.97	-0.05	-1.77	1.33	-0.19	.42
1	Connect the Dots II	-4.32	3.32	-0.15	0.65***	0.11	0.70	0.01	3.67	-0.00				.50
2		-4.74	3.47	-0.16	0.66***	0.11	0.72	-0.62	3.93	-0.02	0.85	1.76	0.06	.51
1	Wordlist Learning	0.34	3.24	0.02	-0.13	0.10	-0.20	5.37	3.58	0.23				.10
2		-0.85	3.31	-0.04	-0.10	0.11	-0.15	3.59	3.75	0.16	2.42	1.68	0.24	.15
1	Wordlist Delayed Recall	-0.82	0.83	-0.15	-0.02	0.03	-0.09	1.18	0.92	0.20				.08
2		-1.16	0.84	-0.22	-0.01	0.03	-0.03	0.66	0.96	0.11	0.70	0.43	0.28	.14
1	Wordlist Recognition	-0.64	0.48	-0.21	-0.02	0.02	-0.16	-0.63	0.53	-0.19				.10
2		-0.81	0.49	-0.26	-0.01	0.02	-0.11	-0.88	0.55	-0.26	0.34	0.25	0.24	.14
1	Reaction speed	-12.01	10.58	-0.17	0.79*	0.35	0.34	-15.53	11.63	-0.20				.18
2		-13.77	11.12	-0.20	0.84*	0.37	0.36	-17.75	12.36	-0.23	3.21	5.62	0.10	.19

Table continues

1	Place the beads	-1.46	3.74	-0.06	0.17	0.12	0.22	-5.79	4.13	-0.22			.11
2		0.25	3.75	0.01	0.12	0.12	0.16	-3.23	4.24	-0.12	-3.48	1.91	-0.30	.18
1	Box Tapping	0.26	0.43	0.09	-0.03*	0.01	-0.35	0.09	0.47	0.03			.13
2		0.18	0.44	0.07	-0.03*	0.01	-0.32	-0.01	0.50	-0.00	0.15	0.23	0.11	.14
1	Fill the grid	-2006.59	3510.78	-0.08	452.91***	110.39	0.55	-5047.39	3933.58	-0.17				.35
2		-1292.99	3640.28	-0.05	431.81**	114.07	0.53	-4097.99	4128.88	-0.14	-1462.57	1833.56	-0.12	.37
1	Digit Sequences I	1.00	0.69	0.23	-0.02	0.02	-0.13	0.67	0.76	0.14				.08
2		1.14	0.71	0.26	-0.02	0.02	-0.15	0.89	0.81	0.18	-0.29	0.36	-0.14	.10
1	Digit Sequences II	0.83	0.76	0.15	-0.09**	0.03	-0.50	1.63	0.85	0.26				.33
2		0.75	0.79	0.13	-0.09**	0.03	-0.49	1.52	0.90	0.24	0.17	0.40	0.06	.33

Abbreviations: SE B = standard error of beta; St. B = standardized beta.
NOTE. Model 1 includes only demographic factors (block 1), i.e., sex, age and education. Model 2 includes demographics and performance-based computer skills (block 2).

*p < .05. **p < .01. ***p < .001.

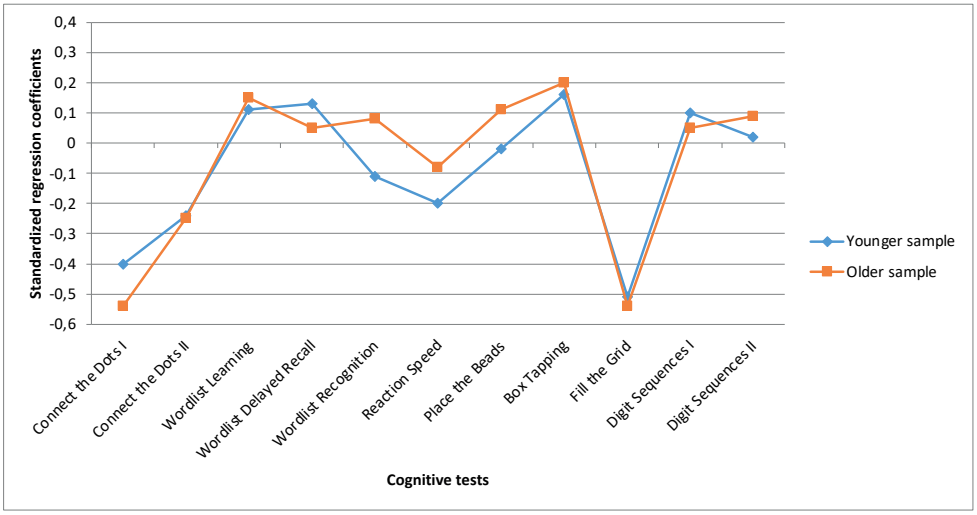


Figure S1. Regression coefficients performance-based computer skills in age-stratified online sample

This figure shows the results of the multiple regression analyses with performance-based computer skills as predictor corrected for demographic factors in the online sample stratified on age. The younger sample includes people with an age below 50 years, the older sample people 50 years and above. On the x-axis, all cognitive tests are shown, from which raw outcome measures were used. On the y-axis, the standardized regression coefficients of performance-based computer skills are reported.

Positive regression coefficients show that better performance-based computer skills coincide with a higher raw test score, while negative coefficients show that better computer skills coincide with a lower raw test score.

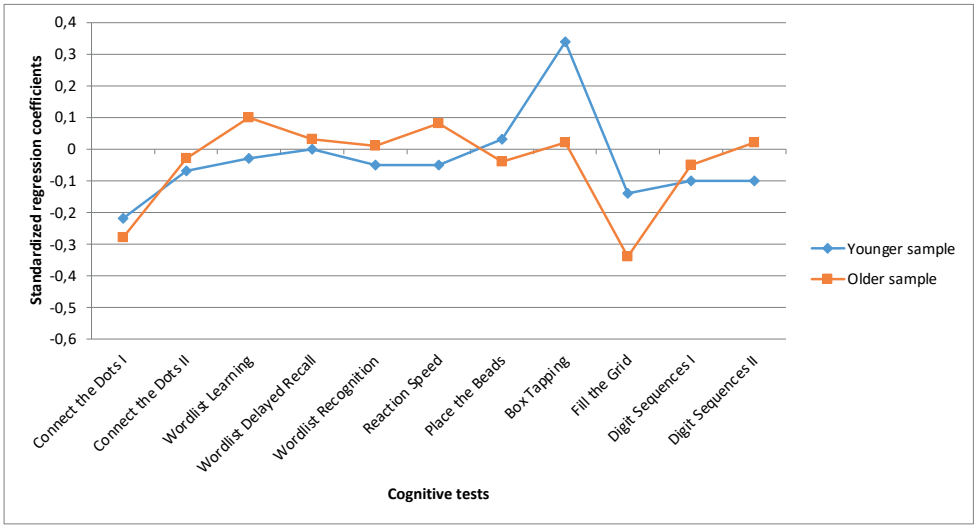


Figure S2. Regression coefficients self-reported hours of computer use in age-stratified online sample

This figure shows the results of the multiple regression analyses with self-reported hours of computer use as predictor corrected for demographic factors in the online sample stratified on age. The younger sample includes people with an age below 50 years, the older sample people 50 years and above. On the x-axis, all cognitive tests are shown, from which raw outcome measures were used. On the y-axis, the standardized regression coefficients of hours of computer use are reported.

Positive regression coefficients show that more self-reported hours of computer use coincide with a higher raw test score, while negative coefficients show that more hours of computer use coincide with a lower raw test score.

Chapter 3:
Standardized item selection for alternate computerized versions of Rey Auditory Verbal Learning Test(-based) word lists

Philippe R. Lee Meeuw Kjoe, Ivar E. Vermeulen, Joost A. Agelink van Rentergem, Elsken van der Wall, Sanne B. Schagen.

Journal of Clinical and Experimental Neuropsychology. 2022;44(9):681-701.

Abstract

Introduction: Despite an increasing need for new Rey Auditory Verbal Learning Test (RAVLT)-based word lists in computerized testing, no criteria or standardized procedures exist for its development. To lay a foundation for future development of new and alternate computerized RAVLT(-based) word lists, we present cross-lingual word criteria, developed new lists using the criteria and evaluated performance on the lists using online assessment.

Method: Based on psycholinguistic literature, we identified relevant word selection criteria. To validate the criteria, we developed two new American-English word lists and one new Dutch list, and administered the RAVLT using visual presentation of the new or original list in an online American (n=248) and Dutch sample (n=246) of healthy people. We compared performance of the new and original word lists on trial scores and serial position effects using Bayesian correlations and analyses of variance. Additionally, we compared proportions of correct responses per item, corrected for serial position.

Results: We identified 13 relevant word selection criteria. The criteria led to two new highly comparable American-English word lists with lower trial scores compared to the original American-English list, indicating that the criteria helped to develop parallel lists with fewer associations between items. The new Dutch word list showed similar trial scores, serial position effects, and proportions of correct responses per item corrected for serial position compared to the original Dutch version.

Conclusions: The systematic use of word selection criteria can facilitate development of new parallel word lists, including in new language areas. Future studies should evaluate the use of the word criteria for the other sections of the RAVLT (such as delayed recall and recognition), performance using original test modalities (auditory presentation and recall of words) as well as performance in clinical samples.

Introduction

Despite the increasing need for new and parallel versions of the Rey Auditory Verbal Learning Test (RAVLT)[1-2] in the context of computerized testing, no widely applied criteria or standardized procedures exist for development of new word lists. The RAVLT is a test of verbal learning and memory. Several procedural variations exist in the administration of the RAVLT, but in its simplest form, patients are presented with a list of 15 words and are asked to recall the words. This procedure is repeated five times. Words are primarily presented auditorily but visual presentation of words is also validated.[3] After a time interval of approximately 20 minutes, patients are asked to recall and to recognize the words. Over the years, many RAVLT word lists and alternate forms have been developed in numerous languages. Still, new and parallel RAVLT(-based) word lists are increasingly needed in the context of computerized assessments. However, the omission of widely applied criteria and standardized procedures complicates creation of new word lists and could potentially lead to suboptimal versions, as many word properties related to memorability of words may not be controlled for (Madan et al., 2020).[4] Word criteria for development of new word lists could help overcome this problem by guiding creation of new equivalent word lists using a uniform procedure. Criteria could also help creation of word lists in new language areas including developing countries and aggregation of data between international centers, which are future aims in neuropsychology (Bilder & Reise, 2019).[5] The current paper aims to develop a standardized procedure for developing new RAVLT(-based) word lists, thus laying the foundation for future development of equivalent word lists.

In this article, we propose a set of cross-lingual word criteria for RAVLT(-based) word lists that can be applied to improve uniformity of words and equivalence between new alternative word lists across languages. In a first step, relevant cross-lingual word criteria are identified. Secondly, based on these word criteria, new American-English and Dutch RAVLT-based

word lists are developed for use in the Amsterdam Cognition Scan (ACS), a self-administered online cognitive test battery (Feenstra, Murre, et al., 2018; Feenstra, Vermeulen, et al., 2018).[6-7] Lastly, online performance on the new word lists is compared with the online performance on the original American and Dutch RAVLT in a group of healthy people in terms of trial scores, serial position effects, and proportions of correct responses on the item-level corrected for serial position.

Cross-lingual word criteria for selection of RAVLT(-based) items

The general aim of developing word criteria for RAVLT(-based) items is that all items contribute to measurement of the same latent memory processes (encoding, consolidation and retrieval of verbal information) without influence of any other factor, such as education level, age or context effects. To reach this aim, items should be approximately equally difficult to learn and remember for all subjects. Differences in level of difficulty can obstruct application of alternate forms and modern use of RAVLT data such as item-level analyses that may shed light on subtle cognitive differences between groups,[8] computerized adaptive testing to shorten neuropsychological assessment, analysis of serial position effects, and data pooling of different word lists within and between countries.

To reach our aims, we applied three basic principles:

1) *Similar memorability of words*

This means that items should be similar in psycholinguistic word characteristics that affect memorability of words, such as valence and imageability.[e.g.,4,9]

2) *Unrelatedness of words*

This means that we aim to prevent inter-item associations between words that can lead to beneficial learning strategies, such as semantic word clustering[e.g.,10] and phonetic/syntactic clustering.

3) *Avoidance of differential item functioning (DIF)*

This means that we aim to minimize inter-individual differences of word memorability due to any factor such as age, educational level and/or cultural background.[5]

To construct a set of relevant word criteria, we examined word criteria formerly used in development of RAVLT(-based) word lists. At construction of the first word learning lists, the Test of Memory for Words by Édouard Clarapède, a small number of criteria were applied (for historical note, see Boake (2000)[11]). André Rey used these exact word lists for his RAVLT[2] and in some translations of this test, like the English version,[12] the original word list was directly translated. Word criteria for the original Dutch RAVLT translation were more elaborative.[3] Later, word criteria for alternate forms have been proposed[13] but were not widely adopted.

We studied 1. the original criteria of the very first French RAVLT word list,[11] 2. word criteria for alternate forms[13] and 3. criteria of the original Dutch word list.[3] The original French criteria were: concrete disyllabic nouns from different semantic categories.[11] Word criteria for alternate forms were: high word frequency, high imageability, mono- or disyllabic words and control for obvious semantic/phonetic associations and similarities between words.[13] Word criteria for the original Dutch RAVLT were: monosyllabic words, word frequency of 20 to 400 per million, age-of-acquisition below 6 years, words referring to concrete objects, control for auditory associations (not further specified), exclusion of potentially ambiguous words and words referring to

emotional factors, and control for semantic associations by even distribution of categories of words (such as animals).[3]

Next, we reviewed psycholinguistic and neuropsychological studies that examined certain word characteristics and their effects on memorability of words.[e.g.,4,10] This led to identification of 13 word criteria, grouped over four applied principles: similar memorability, mutual unrelatedness, avoidance of differential item functioning, and practical issues (see Table 1 for an overview of the criteria).

Table 1

Word criteria and examples

Criteria	Example
Nouns.	'Fridge', not 'tall'.
Monosyllabic words.	'Couch', not 'furniture'.
Words with high concreteness/imageability.*	'Hand', not 'luck'.
Words that only have one meaning.	'Knife', not 'pool'.
Neutral words.*	'Shelf', not 'blood'.
No semantic associations.	Not 'bird'-'owl'-'duck'.
No phonological associations.	Not 'hand'-'band'-'land'.
Words with a medium frequency of use.*	'Milk', not 'drone'.
Words that are easy to spell.	'Hand', not 'suite'.
Words with a low age of acquisition.*	'Shoe', not 'spine'.
Ethnic/religious/subcultural connotations and cognates/loanwords.	'Tree', not 'church'.
Exclude words from other verbal memory tests.	E.g., 'drum' (original RAVLT).
Words that contain at least 4 letters.	'Bird', not 'cat'.

*The word criterion is dependent upon available information in the specific language area.

Similar memorability of words

To ensure that items in the word list have similar memorability, we introduce five word selection criteria:

- Criterion 1: Use nouns.

Words with different functions (parts of speech) may differ in memorability. For example, nouns are easier to memorize than other parts of speech (such as adjectives and verbs),[14-15] possibly since nouns cross-linguistically are more imageable (see imageability criterion below). To avoid differences in memorability, only one word type should be used —because of their prevalence, nouns are the preferred candidate.

- Criterion 2: Use monosyllabic words.

According to the word length effect, shorter words are recalled better than longer words.[16] This is due to a slower encoding for longer words and therefore, higher proneness to decay for phonological memory traces. To avoid differences in memorability due to word length, and in accordance with previous word lists, it would make sense to include only monosyllabic words. If it is impossible to select 15 monosyllabic words that adhere to all criteria, only disyllabic words can be selected, as long as the number of syllables per item within word lists are consistent.

- Criterion 3: Use words with high concreteness and imageability.

Concrete nouns denote objects that can be perceived by the senses, while abstract nouns do not.[17] Imageable words (e.g., 'hat') generate a mental image while non-imageable words (e.g., 'luck') do not.[18] Concrete words are easier to remember than abstract words[19] and imageable words more than non-imageable words,[20] possibly since concrete and imageable words are encoded both in a verbal and a non-verbal system[21] or are more easily put in

a semantic context in a person's knowledge database.[22] To avoid differences in memorability and consistent with the original RAVLT word criteria, we chose to only include concrete and imageable words.

- Criterion 4: Use words that only have one meaning.

The memorability of a word with multiple meanings (such as homonyms, e.g., chair can mean furniture or chairperson) depends on which meaning one processes. Different meanings can differ for example in concreteness, resulting in differences in memorability.[23] Ambiguous words are also thought to induce more diffuse activation of the brain than concrete words.[24] To avoid these problems, words used should have one meaning only.

- Criterion 5: Use neutral words.

Valence refers to pleasantness of a stimulus. Valence is related to memorability of words.[9] Most studies show that negative and positive words are easier to memorize than neutral words.[e.g.,25] As in the category of nouns, there are much more neutral words than positive or negative words, it makes sense to choose neutral words.

Unrelatedness of words

Relatedness between words can influence their memorability,[10] and can make test performance more sensitive to executive functioning as application of encoding strategies can improve test performance.[26] The following two word selection criteria aim to reduce relatedness between selected words:

- Criterion 6: Avoid semantic associations.

Word lists consisting of words from the same semantic category (e.g., animals, body parts) are better recalled than lists of words from different categories: "the category effect".[10,27] A possible explanation is that when

words belong to the same semantic category, one word can work as a cue for other words narrowing down possibilities of other words (redintegration hypothesis). Compound cueing, i.e., combining items into a single item to aid memory (e.g., combining 'tea' and 'cup' to 'teacup'), is another way in which semantic associations may help retrieval of words.[28] To avoid such clustering strategies, semantic associations between items should be avoided.

- Criterion 7: Avoid phonological associations.

The use of phonologically similar words (e.g., words that rhyme or alliterate) can both positively affect word recall by cueing and negatively affect word recall by weakening of the memory trace through interference.[29-30] To avoid such complex effects, avoid phonological associations.

Avoidance of differential item functioning

To avoid differential item functioning[5] (e.g., people with a medical background might remember the word 'spleen' better), the following four word selection criteria are applied:

- Criterion 8: Use words with a medium frequency of use.

Word frequency is the rate that words occur per million words in natural language, often measured by means of corpora extracted from books, newspapers or movie subtitles. Higher frequency words (more than 100 per million words) are recalled better than low frequency words (less than 5 times per million), but this word frequency effect differs between people as it depends on one's vocabulary size.[31] The effect is highest in people with low vocabulary and diminishes when vocabulary grows. To avoid these inter-individual differences, we chose to include words with a word frequency between 5 and 100, as this range resembles medium word frequency and is more often applied in word list development.

- Criterion 9: Use words that are easy to spell.

Spelling transparency refers to the ease with which a word can be read/spelled, based on grapheme-to-sound correspondence. An example of a transparent/easy to spell word is 'hand' and an opaque/difficult word to spell is 'suite'. Spelling transparency is related to memorability but also depends on age and spelling ability. Transparent words are recalled significantly more easily than opaque words in younger readers and dyslexics, while opaque words are recalled better in older readers.[32] To avoid differences in memorability, we chose only words with high transparency as the outcome should not be related to spelling ability.

- Criterion 10: Use words with a low age of acquisition.

The age of acquisition (AoA) refers to the age at which a word is generally learned. It is unclear whether early AoA or late AoA words are better remembered but age seems to play a role in this effect.[33] To avoid age-related differences, we chose to include only early AoA words, since these words often are more concrete and imageable[34] and have higher word frequency[35] than late AoA words. We selected this specific age (4-6 years) range in correspondence with criteria from earlier RAVLT versions.[e.g.,3]

- Criterion 11: Avoid words with ethnic/religious/subcultural connotations and cognates/loanwords.

Words should be related as little as possible to any demographic/cultural background (e.g., gender-stereotyped words, archaic words, words related to religion), which can lead to differential item functioning (DIF) between different individuals. For example, women are shown to have better recall of feminine words ("dress") than masculine words ("beard"),[36] since these words lead to more elaborative semantic processing as words are seen as more relevant to their gender,[37] leading to DIF. Another example of DIF is in case of cognates, i.e., words that are phonologically similar and orthographically

identical in two languages such as the original French word 'croissant' in English. Studies have shown that bilinguals recognize cognates faster than non-cognates,[38] possibly influencing word recall too,[39] and thus leading to DIF between bilinguals and monolinguals.

Practical criteria

The last two criteria are practical rather than theory driven, aiming to reduce practice effects (12), and to facilitate automated test scoring in case of computerized testing (13).

- Criterion 12: Exclude words from other verbal memory tests.

To avoid overlap of items between new and existing word lists, words from other word lists are excluded as much as possible. Overlap in items would be a clear confound in assessing change over time, as 'parallel' word lists that include some of the same items can give rise to practice effects.

- Criterion 13: Words that contain at least 4 letters.

In computerized testing, participants are often asked to type in words they remember. Test scoring is typically automated, using a certain threshold of spelling mistakes that are allowed. In the ACS for example, we use a Levenshtein distance of 1, meaning that we treat incorrect responses as correct if the response only needs one character edit to become correct (e.g., 'halll' instead of 'hall'). To avoid treating incorrect responses as correct while a patient in fact meant another word, we included only words with at least 4 letters, since especially words that are shorter than 4 letters can be changed in many other words by one edit ('cat' can be changed to 'hat', 'mat', 'bat', 'cot', 'can' etc.).

Studies

To validate the criteria, we present two studies in which we develop two new American-English word lists (Study 1) and one new Dutch list (Study 2), and compare their performance with that of their original counterparts in terms of overall trial scores, serial position effects, and scores per individual item.

Study 1: materials and methods

Development of new American-English ACS versions of the RAVLT

This section describes the methods and results of the compilation of two new American-English word lists for application in the ACS.

We began our procedure by searching for American-English word corpora that categorized words on relevant linguistic characteristics. This led to three separate corpora: 1) Brysbaert, Warriner and colleagues (2014) categorized 37,058 words on concreteness (on a 5-point Likert scale) (criterion 3),[17] 2) Kuperman and colleagues (2012) categorized 30,121 words on age-of-acquisition ratings (criterion 10),[40] and 3) Warriner and colleagues (2013) categorized 13,915 words on valence (on a 9-point Likert scale) (criterion 5)[41] and included word frequency ratings from Brysbaert and New (2009) who categorized 74,286 words on word frequency (criterion 8).[42]

Based on all three databases, we created a pool of words from the word corpora that adhered to the word criteria, only including words that were included in all three databases. We firstly excluded words per database that did not adhere to criteria 1, 3, 5, 8 and 10, and combined words that remained. This led to a pool of 574 words. We removed words that did not adhere to criteria 2, 4, 11, 12 and 13. Regarding criterion 12, we only removed words that were included in existing RAVLT word lists in order to have enough items to choose from. This led to a pool of 117 words.

Based on this pool, we selected two sets of 15 words out of 117 eligible words, controlling for criteria 6, 7 and 9. We based word order on variety of vowels and avoidance of any association between items we could think of. Similar to the original RAVLT, we chose for a fixed word order between and within people. A fixed order helps to avoid between-subject differences in word clustering effects on memorability due to sequential presentation and facilitates analysis of serial position effects. A small group of research colleagues underwent the verbal memory test with the new word list; this pretest did not bring to light overseen word associations. See Figure 1 for the flowchart of inclusion of words.

Criterion 12 required us to exclude words that appear in other word lists, primarily with the traditional versions of the RAVLT in mind. As an anonymous reviewer has correctly pointed out, there are some words in the new lists that appear in versions of the California Verbal Learning Test and the Hopkins Verbal Learning Test-Revised. The new word list should ideally not be used as an alternate form of these tests for this reason. Another reason would be that these tests are fundamentally different, in that they deliberately employ semantic clustering between words, which we attempted to avoid for the new list with Criterion 6, as the RAVLT uses word lists consisting of unrelated words.[2]

Qualitative comparison of word characteristics of items from the new and the original American-English RAVLT(-based) word lists showed several differences. The two new American-English lists only contain monosyllabic words, whereas the original list contains both mono- and disyllabic words. Words from the new lists have on average a lower word frequency and also a narrower range of word frequency than the original list. The original list has a narrower range of valence but is on average more positively valenced. The original list has a narrower range of age-of-acquisition compared to the new lists and words on average had a younger age-of-acquisition in the original list than in the new lists. The original list contains words with semantic associations (words from the same category and words that can be

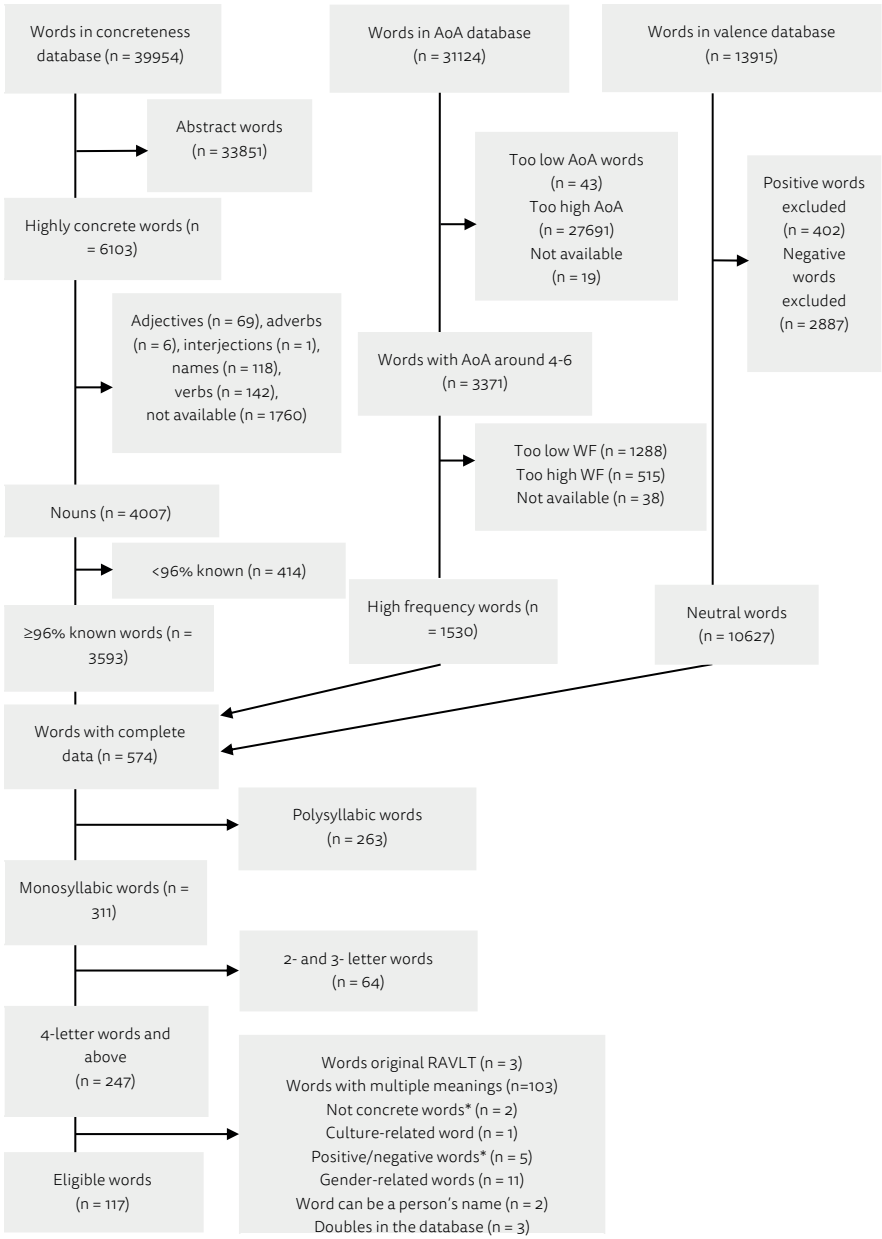


Figure 1. Flowchart of selection of American words
Abbreviation: AoA = age of acquisition, WF = word frequency, RAVLT = Rey Auditory Verbal Learning Test.
Concreteness rating 4.40 – 5.00.[17] Age of acquisition ratings in years, between 3.00 and 6.80 years of age.[40] Word frequency between 10 and 100.[42] Valence ratings between 4.00 and 7.30.[41]
*Some words adhered to the word criteria ratings, but were excluded based on the author's judgment (e.g., clown).

combined as in compound cueing) and phonetic associations (alliterations). Lastly, the original word list contained one word that can be interpreted as culture-related. See Table 2 for a comparison on applied word criteria per American-English word list. See Table S1 for psycholinguistic characteristics per item of the new and original American-English word lists.

Comparison of the new and original American-English RAVLT(-based) word lists

This section presents the methods and results of the comparison of the new and original American-English RAVLT(-based) word lists. We compared the word lists in terms of trial scores, reliability of trial scores, serial position effects, and proportions of correct responses per each item corrected for serial position.

Participants

Data were collected (between July and August 2021) from 248 American-English-speaking adults (see Table 3). Participants were recruited via MTurk; a crowdsourcing marketplace that is used as a data collection tool for social and behavioral sciences, which can provide high-quality data as long as several measures are taken.[43] All subjects were required to be 18 years or older and to have English as their first language. To maximize data quality, we used extra qualifications for participant selection: currently located in the United States, have a Human Intelligence Task (HIT; virtual task on MTurk) approval rate of 95%, and have more than 100 HITs approved on MTurk. These criteria helped to select people who previously completed online tasks on MTurk in a reliable manner. Exclusion criteria were a neurological/psychiatric disorder and a history of cancer as this can negatively influence performance on verbal memory tests.[44] We did not account for reading disabilities as reading tasks are frequent on MTurk, and an inherent part of being a functional member of the platform community.

Table 2

Comparison applied word criteria for the original and new American and Dutch word lists

Criteria	Original American	New American 1	New American 2	Original Dutch	New Dutch
Nouns	✔	✔	✔	✔	✔
Monosyllabic words	✘	✔	✔	✔	✔
Words with high concreteness/ imageability	✔	✔	✔	✔	✔
Words that only have one meaning	✘	✔	✘	✘	✘
Neutral words	✘	✔	✘	✘	✔
No semantic associations	✘	✔	✘	✘	✔
No phonological associations	✘	✔	✔	✘	✔
Words with a medium frequency of use	✘	✔	✔	✘	✔
Words that are easy to spell	✔	✔	✔	✔	✔
Words with a low age of acquisition	✔	✔	✔	✔	✔
Ethnic/religious/ subcultural connotations and cognates/loanwords	✘	✔	✘	✘	✔
Exclude words from other verbal memory tests	✘	✘	✔	✔	✘
Words that contain at least 4 letters	✘	✔	✔	✘	✔

Table 3

Demographics of American study population (study 1)

Demographics	Original American (N=80)	New American 1 (N=79)	New American 2 (N=89)
Women, n (%)	37 (46.3%)	30 (38.0%)	42 (47.2%)
Age (years), M (SD), range	37.4 (11.9), 18-71	35.4 (9.3), 21-65	38.1 (12.4), 20-71
Education level*			
Low, n (%)	43 (53.8%)	23 (29.1%)	35 (39.3%)
High, n (%)	37 (46.3%)	56 (70.9%)	54 (60.7%)
Device			
Desktop and laptop, n (%)	75 (93.8%)	74 (93.7%)	87 (97.8%)
Tablet and phone, n (%)	5 (6.3%)	5 (6.3%)	2 (2.2%)

*Education is based on classification 1 to 7: 1) 8th Grade or less, 2) 9 – 11th Grade, 3) GED,

4) High school graduate, 5) Associate degree / some college, 6) Bachelor's degree, 7) Advanced degree (Master's / Doctorate). Low education = 1 to 5; and high education = 6 to 7.

Ethical approval was given by the review board of the Netherlands Cancer Institute conform ethical guidelines for human experimentation stated in the Declaration of Helsinki (reference approval number: IRBd18-124). All participants provided online written informed consent prior to participation.

Materials

We administered a computerized version of either the original American-English version of the RAVLT or our newly developed version of the American-English RAVLT in a group of healthy adults. Allocation to the original or newly composed test was random with restriction of equal group sizes. In this study, we only administered the immediate recall phase

of the American-English RAVLT (trial 1 to 5), without the delayed recall and recognition phase.

Procedure

The test was administered online, using Qualtrics (Qualtrics, Provo, UT). Test instructions for the two versions were identical. Subjects were shown on their computer screen a list of 15 words, one by one, and were instructed to type in as many words as they could remember (in any order) one by one, as soon as the last word had been presented. This procedure was repeated four times with identical word order, leading to five trials in total. Words were presented for two seconds, with one second in between. All words were presented in black letters against a white background.

Assessment was done online without supervision. No downloads were needed and the test operated on any major Internet browser on a desktop, laptop, tablet or smartphone. An earlier study showed sufficient reliability and concurrent validity of the computerized, online version when compared to the original auditory version of the RAVTL for the Dutch version[6-7] and the American-English version (data not yet published). Participants were instructed to complete the test by themselves, alone, and in a quiet location with their tv, radio and phone notifications turned off. It was explicitly stated that use of any external aid, such as a notepad, was not allowed. The test took around 15 minutes to complete and subjects received a compensation of $2.30.

Data analysis

SPSS Statistics for Windows, version 22.0 (IBM Corp., Armonk, NY) and JASP, version 0.14 (JASP Team, 2020)[45] were used for statistical analyses. Probabilities of $p < .05$ were considered statistically significant. Only data from complete responders were analyzed. Between-group differences in demographic factors (sex, age and education) and device were analyzed using independent t-tests and χ^2-tests.

To ensure integrity of the online data, we undertook two additional steps. First, we removed error outliers, which we consider non-legitimate scores. These scores are possible because of the online administration of the test. Because participants are completing the test online at home, we have less control over test administration than in a face-to-face setting. Participants could have a computer malfunction that results in very low scores, or might click through in error to finish as quickly as possible. These low scores would be non-legitimate and would have an outsized effect on results. Therefore, we chose to remove them with a non-stringent minimal learning curve, which does not exclude any participants that have legitimate scores. We defined a minimal learning curve for the RAVLT using data from the Advanced Neuropsychological Diagnostics Infrastructure (ANDI) project, a Dutch database aggregating neuropsychological data of healthy controls from several research groups.[46] The minimal learning curve was set at the lowest trial scores normative participants could have. This resulted in a minimal learning curve of 2-3-3-4-5. Subjects were removed from our analyses when they scored below the minimum learning curve more than once, as we did not want to set the removal criteria too strict given the online and unsupervised nature of our study. We considered the minimal learning curve based on Dutch data to rule out computer malfunction to be appropriate for detecting computer malfunction in the American data as well. As a second step we took to ensure integrity of the data, we excluded subjects who entered responses with 8 words in order or reverse order as these scores might indicate use external aids to help learning performance.

To compare learning curves of the new and original word lists, we conducted a Bayesian mixed ANOVA, with version (original vs. new word lists) as between-subjects factor, trial (1-5) as within-subjects factor, and the total number of correct responses per trial as dependent variables (0-15), with demographic factors as covariates if preliminary analyses showed between-group differences on these variables. We used the default prior options for the effects (i.e., r = 0.5 for the fixed effects).

To determine reliability of trial scores of the new and original word lists, we conducted Bayesian correlation analyses, whereby we examined separate Pearson's correlations between the number of correct responses on each trial (0-15) and the total score over 5 trials (0-75), for both the new and original word lists. We used default prior options for the analyses (i.e., a stretched beta distribution with width = 1). Since we expected positive correlations, we specified one-sided alternative hypotheses and looked at 95% credible intervals for each correlation coefficient.

To compare serial position effects between the word lists, we conducted a Bayesian Repeated Measures ANOVA with version (original vs. new word lists) as independent variable, serial position as within-factor (primacy, middle, or recency), and percentage of correct responses at each serial position over all five trials as dependent variables, adjusted for demographic factors if significantly different between the groups. Based on the serial position classification,[47] primacy, middle, and recency represented percentages of correct responses of the first three words, the middle nine words, and the last three words in the list respectively over all five trials.

Given the strong impact of serial position on recall rates of words, we compared proportion of correct responses for each single word over all 5 trials between and within the original and new American-English RAVLT, corrected for serial position effects. To correct for serial position effects, we conducted a logistic regression model analysis, with the centered serial position (-7 to 7) and trial (1-5) as independent variables and correct (0=no, 1=yes) as dependent variable, and saved the expected values based on this serial position model. We then compared residuals from the expected values based on this serial position model. To obtain observed proportions of correct responses for the words over all five trials, we conducted a univariate ANOVA, with word (1-15) as independent variable and residuals of the logistic regression as dependent variables. To compare residuals between and within word lists and to identify items with uncharacteristically high or low scores, we visually inspected residuals per word.

Results study 1

Incomplete responders

Of the 556 participants who were randomized, 18.5% (n = 103) did not complete the online study; no reason for drop-out was provided. Of those 103 people who stopped earlier, 17.5% (n = 18) stopped after trial 1, 27.2% (n = 28) after trial 2, 15.6% (n = 16) after trial 3, 16.5% (n = 17) after trial 4, 22.3% (n = 23) after trial 5, and 1% (n = 1) at the demographic questions. This led to 453 complete responses. See Supplements Figure S1 for the flow-chart of participation and completion rates.

Data cleaning

Of the 453 complete responses, 45.3% (n=205) was excluded from analyses. Of those 205 excluded responses, 68.2% (n=140) was excluded as these responses with minimal 8 words in perfect sequence seemed to indicate use of a notepad and 26.8% (n=55) was excluded because these responses were below the minimal learning curve. Ten respondents were excluded as they had changed their answers to the screening questions post-hoc to continue the study. This led us to include 248 subjects in the final analysis.

Comparison on demographic characteristics

The three American samples did not differ in terms of sex, age and the device (desktop, laptop, tablet or smartphone) on which they completed the online memory test. However, on average, the sample who completed the original word list was lower educated (46.3% high education) compared to the samples who completed the first or second new word list (70.9% and 60.7% high education, respectively), χ^2 (2, n = 248) = 10.11, p < .01). There was no significant difference in education level between the samples who completed the two new word lists, χ^2 (1, n = 168) = 1.51, p = .22.

Comparison of learning curves between new and original American-English word lists

The Bayesian Repeated Measures ANOVA showed that, corrected for education level, the average learning curve differed between the American-English word lists. The learning curve of the original American-English word list was higher than the learning curves of the two new word lists while the two new word lists did not differ in scores. Data were best represented by a model that included the variables trial, version and education level. Compared to the same model without version, the BF_{10} was 63.53, indicating decisive evidence for differences between word lists. Post hoc comparisons (Bayesian t-tests controlled for multiplicity) showed that there is decisive evidence for differences between the original word list and the two new word lists (BF_{10} > 209.5). Post hoc comparisons showed that scores increased with each trial (BF_{10} > 226981.0), indicating absence of ceiling effects. See Figure 2 for average learning curves of the new and original American-English RAVLT.

Reliability of trial scores

The reliability of all trial scores of both new word lists were strongly similar to those of the original word list. Strong correlations were found between the scores of all trials and the total score summed over five trials. The 95% credible intervals of the correlations ranged from .70 to .95. See Table 4 for Bayesian Pearson's r correlations for all trials.

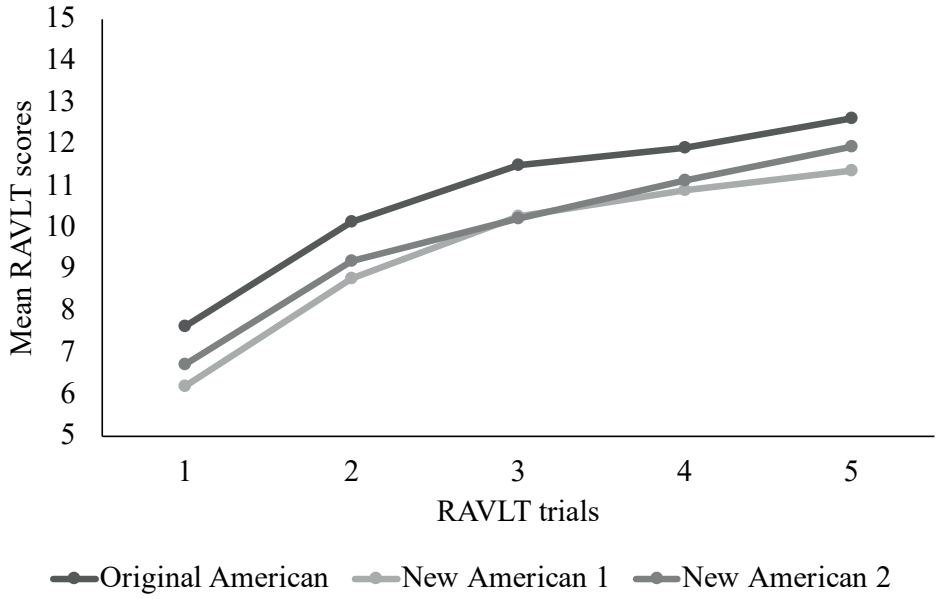


Figure 2. Learning curves for trials 1 to 5 on the new and original American RAVLT
Abbreviation: RAVLT = Rey Auditory Verbal Learning Test.

Comparison of serial position effects

Differences were found in serial position effects between the word lists. A Bayesian Repeated Measures ANOVA showed differences in primacy, middle and recency scores between the word lists. The model including the predictors serial position, version and an interaction between serial position and version fit the data better than this model without the interaction (BF=59.0). Both the new and the original American-English word lists showed primacy effects, but only the new American-English word lists showed a recency effect. The second new word list had a higher primacy score (95% Credible Interval = .81-.87) than the first new word list (.70-.80). The original word list had a higher middle score (95% Credible Interval = .64-.72) compared to the new word lists (.53-.62 and .54-.62). The word lists did not differ in recency scores. See Table 5 for percentages of

Table 4
Bayesian Pearson's r correlations between trial scores and total score of the American word lists

Version	Variable		Trial 1 score	Trial 2 score	Trial 3 score	Trial 4 score	Trial 5 score
Original American	Total score	Pearson's r	.82	.90	.90	.88	.87
		BF_{10}	$5.88*10^{17}$	$2.92*10^{26}$	$2.06*10^{26}$	$1.86*10^{24}$	$4.04*10^{22}$
		95% CI	.73-.88	.84-.93	.84-.93	.82-.92	.80-.91
New American 1	Total score	Pearson's r	.70	.87	.93	.90	.90
		BF_{10}	$9.25*10^{9}$	$1.55*10^{22}$	$5.18*10^{30}$	$6.65*10^{25}$	$1.14*10^{26}$
		95% CI	.55-.79	.80-.91	.88-.95	.84-.93	.84-.93
New American 2	Total score	Pearson's r	.80	.93	.89	.88	.86
		BF_{10}	$3.93*10^{17}$	$5.65*10^{34}$	$3.68*10^{27}$	$2.09*10^{26}$	$3.26*10^{24}$
		95% CI	.70-.86	.88-.95	.83-.92	.81-.92	.79-.91

NOTE. Total score is the summed score over all five trials. CI = credible interval.

correct response per serial position. See Figure 3 for serial position curves of the new and original word lists.

Table 5

Percentages correct responses per serial position of the American word lists

Version	Serial position effect	Mean % correct responses	95% CI
Original American	Primacy	.83	.79 - .87
	Middle	.68	.64 - .72
	Recency	.72	.68 - .77
New American 1	Primacy	.75	.70 - .80
	Middle	.57	.53 - .62
	Recency	.70	.66 - .74
New American 2	Primacy	.84	.81 - .87
	Middle	.58	.54 - .62
	Recency	.69	.65 - .74

NOTE. Total score is the summed score over all five trials. CI = credible interval.

Item-level comparison of proportion correct responses

Corrected for serial position, items from all three word lists were learned equally well. In general, items from both new word lists and the original word list showed similar residuals, with all data points between -.15 and .10, showing similarity in pattern of correct responses and a low margin of fluctuations of correct responses within word lists. Figure 4 depicts proportion correct responses over all trials per item per test, corrected for serial position.

In the second new American-English RAVLT version, performance on the 4th and the 10th word was lower than expected and lower than every other word, indicating that these words were more difficult to learn than expected based on their position and more difficult in comparison with the other words, possibly needing alteration.

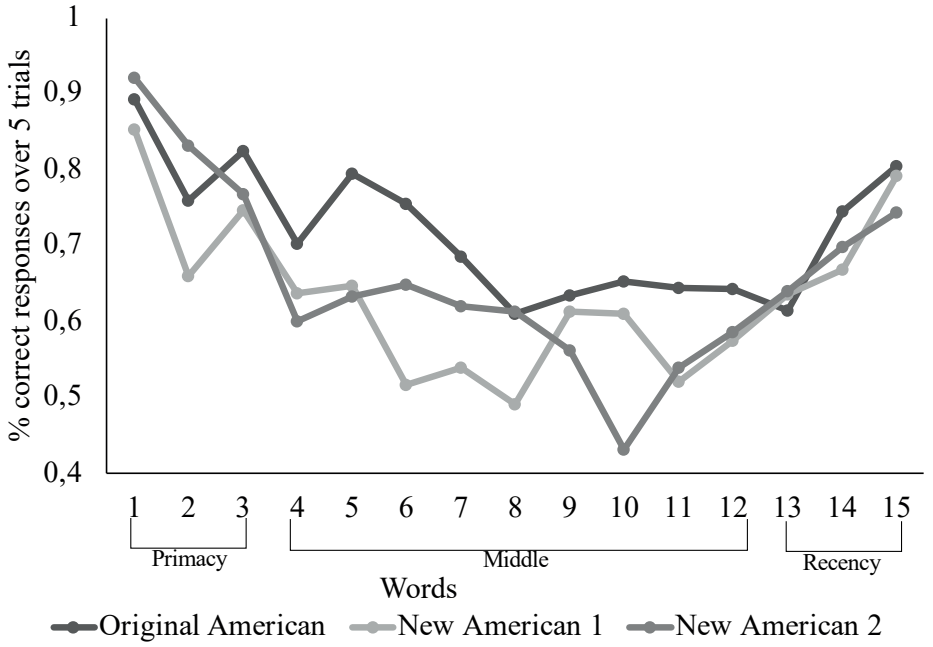


Figure 3. Serial position curves for the new and original American versions of the RAVLT
Abbreviation: RAVLT = Rey Auditory Verbal Learning Test.
NOTE. Word recall on the y-axis represents the percentage of correct responses per item over all five trials.

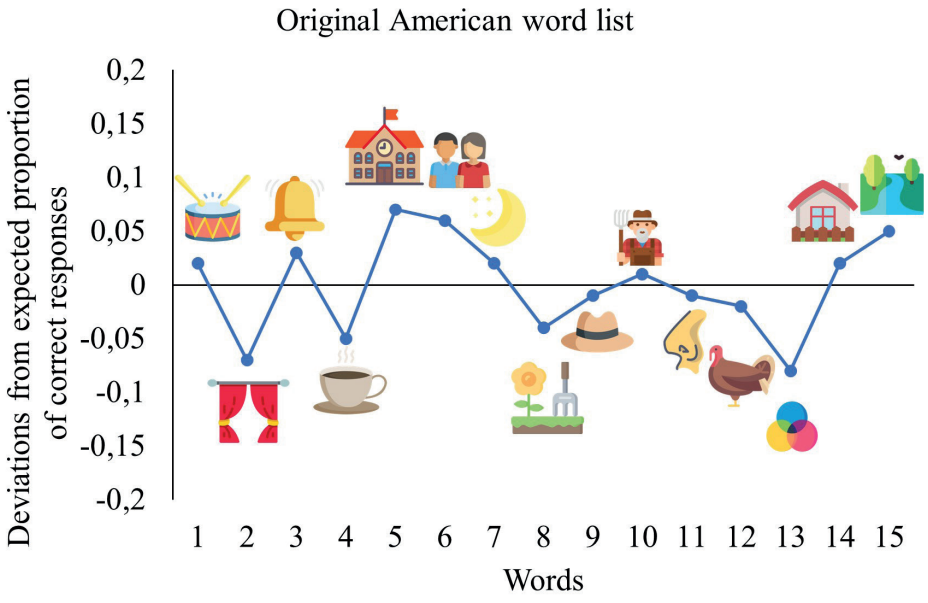
Study 2: materials and methods

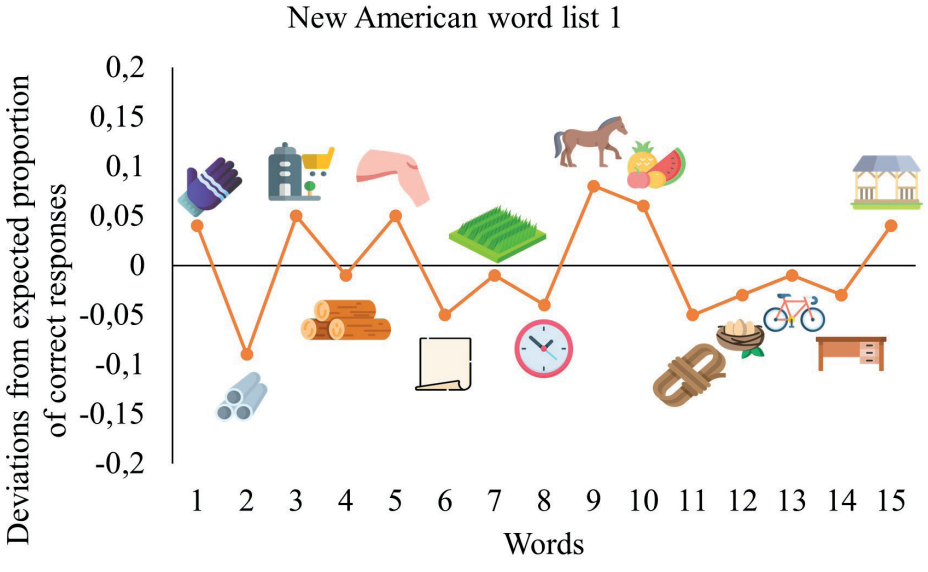
Development of a new Dutch ACS version of the RAVLT

This section describes the methods and results of the compilation of a new Dutch word list for application in the ACS.

We searched for Dutch word corpora that categorized words on relevant linguistic characteristics. We found three separate corpora: 1) Brysbaert, Stevens and colleagues (2014) categorized 30,071 words on concreteness (on a 5-point Likert scale) (criterion 3),[48] 2) Brysbaert, Stevens and colleagues (2014) aggregated age-of-acquisition ratings from 4 different studies for 31,178 words (criterion 10),[48] and 3) Moors and colleagues (2013)

categorized 4299 words on valence (on a 7-point Likert scale) (criterion 5)[49] and included word frequency ratings from Keuleers and colleagues (2010) who categorized 437,503 words on word frequency (criterion 8).[50]





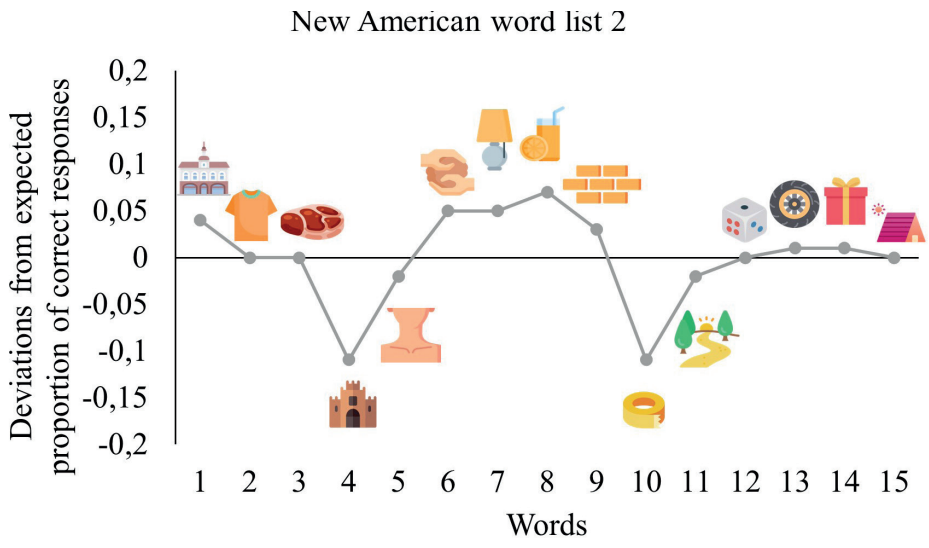


Figure 4. Deviations from expected proportion of correct responses based on serial position per word for the new and original American RAVLT
Abbreviation: RAVLT = Rey Auditory Verbal Learning Test.

NOTE. The word items are depicted by icons to avoid the word items to be publicly known by future subjects. Upon request, the new American word lists are freely available for researchers.

Based on the three word corpora, we built a pool of words that adhered to our word selection criteria. We started by excluding words that did not adhere to criteria 1, 2, 3, 5, 8 and 10, and combining words that remained, leading to a pool of 143 words. We removed words that did not adhere to criteria 4, 9, 11, 12 and 13, leading to a pool of 89 words. Based on this pool, we selected one list of 15 words, controlling for criteria 6 and 7. Similar to the original RAVLT, we chose for a fixed word between and within subjects. We decided word order controlling for variety of vowels and avoidance of any association between items. A pretest with a small group of research colleagues who underwent the verbal memory test with the new word list did not bring to light unwanted word associations. See Figure 5 for the flowchart of inclusion of words.

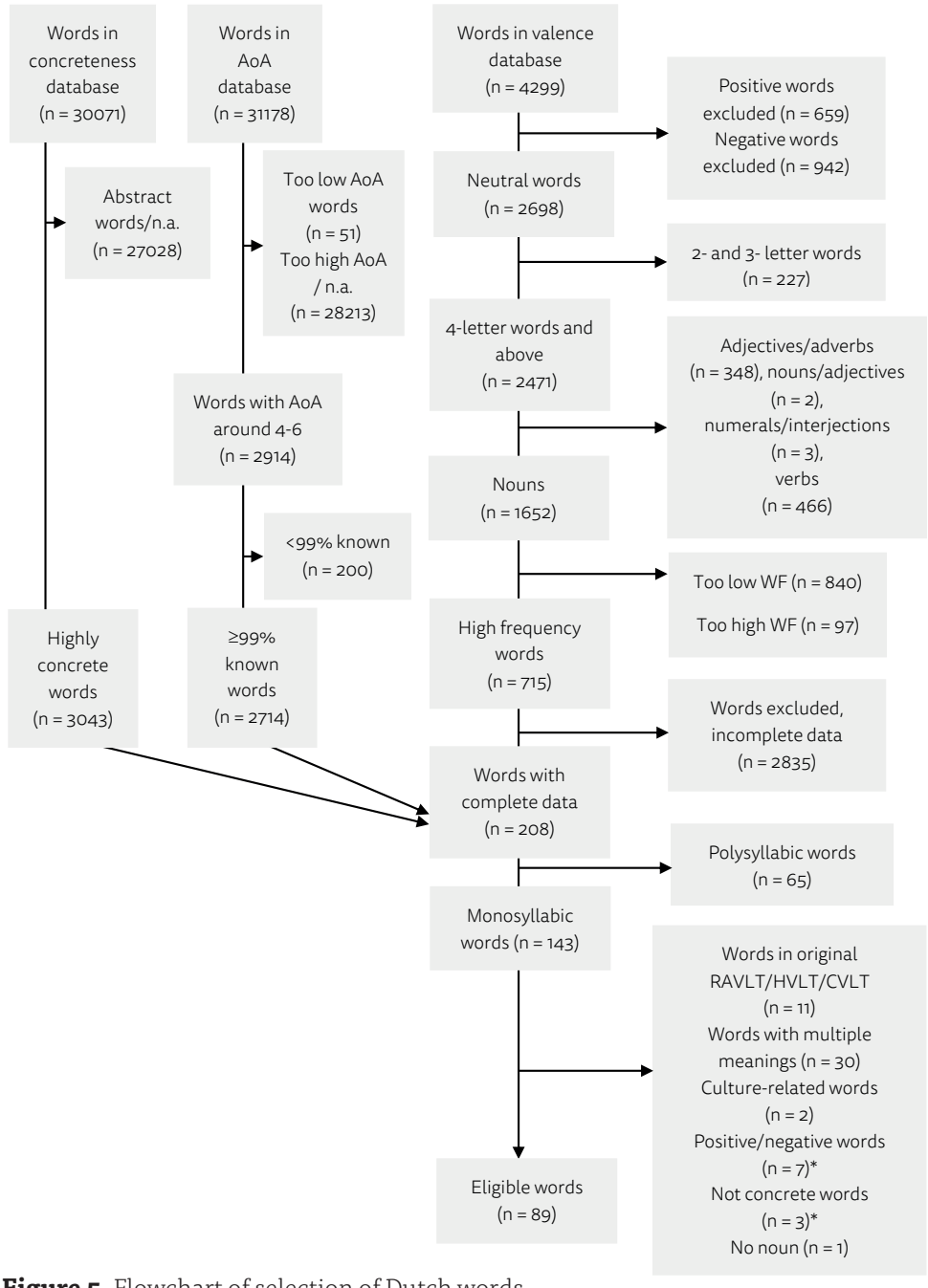


Figure 5. Flowchart of selection of Dutch words
Abbreviation: AoA = age of acquisition, WF = word frequency, RAVLT = Rey Auditory Verbal Learning Test, HVLT = Hopkins Verbal Learning Test, CVLT = California Verbal Learning Test.
Concreteness rating 4.5 – 5.[48] Age of acquisition ratings in years, between 3.90 and 6.99 years of age.[48] Valence ratings between 3.00 and 5.00.[49] Word frequency between 4 and 90.[50]
*Some words adhered to the word criteria ratings, but were excluded based on the author's judgment (e.g., clown).

A qualitative comparison of word characteristics of items in the new and in the original Dutch RAVLT(-based) word list[3] revealed several similarities. Both word lists only contained monosyllabic nouns that are easy to spell and had no clear mutual semantic associations. Also, both word lists contained concrete, neutral, medium-frequent, and words with a low age of acquisition.

The original word list did contain more words with multiple meanings, as well as some phonetic mutual associations between words (mainly alliterations). Lastly, the original word list contained one gender-related word (see criterion 11). See Table 2 for a comparison on applied word criteria for the new and original Dutch word lists. See Table S2 for psycholinguistic characteristics per item of the new and original Dutch word lists.

Comparison of the new and original Dutch RAVLT(-based) word lists

This section presents the methods and results of the comparison of online performance on the new and original Dutch RAVLT(-based) word lists. We compared the new and original word lists in terms of trial scores, reliability of trial scores, serial position effects, and proportions of correct responses per each item corrected for serial position.

Participants

Data were collected (between February and August 2020) from 246 Dutch-speaking adults (see Table 6). Participants were recruited via Academic Research Panel; an online platform of volunteers who previously participated in research (mostly involving self-report studies) and who signed up to be invited for new studies. All subjects were required to be 18 years or older and to have sufficient proficiency of the Dutch language. Exclusion criteria were a history of cancer and the presence of any disease. This last criterion was broad to avoid detailed follow-up questions of medical history in a taxing verbal learning study. Although we would have preferred to keep criteria the same between the American and Dutch sample, symptoms that could influence cognition were removed in both samples. We did not account for reading disabilities as these are unlikely in such a participant recruitment platform.

Table 6

Demographics of Dutch study population (study 2)

Demographics	Original Dutch (N=116)	New Dutch (N=130)
Women, n (%)	76 (65.5%)	79 (60.8%)
Age (years), M (SD), range	51.7 (14.8), 18-77	48.9 (17.0), 18-78
Education level*		
Low, n (%)	29 (25.0%)	26 (20.0%)
High, n (%)	87 (75.0%)	104 (80.0%)
Device		
Desktop and laptop, n (%)	79 (68.1%)	87 (66.9%)
Tablet and phone, n (%)	37 (31.9%)	43 (33.1%)

*Education is based on Verhage's classification 1 to 7.[51] Low = Verhage 1 to 5; and High = Verhage 6 to 7.

Ethical approval was given by the review board of the Netherlands Cancer Institute conform ethical guidelines for human experimentation stated in the Declaration of Helsinki (reference approval number: IRBd18-124). All participants provided online written informed consent prior to participation.

Materials, procedure and data analysis

The materials, procedure and analysis plan for the Dutch RAVLT were equivalent to those of Study 1. Instead of a small compensation for each participant, two randomly selected subjects received a gift card of €20,-.

Results study 2

Incomplete responders

See Supplements Figure S2 for the flow-chart of participation and completion rates. Fifty percent (n = 265) of people who started did not complete the verbal memory test. Most people who dropped out did so directly after trial 1 (87.2%, n = 231), whereas a few stopped after trial 2 (4.9%, n = 13), trial 3 (5.7%, n = 15), or trial 4 (2.3%, n = 6). Importantly, the new and well-established original word lists showed similar dropout rates (48.7% and 51.3% respectively). Reasons for dropping-out were unclear, but likely had to do with motivation or practical time constraints. Members of Academic Research Panel usually fill in questionnaires, which requires less and shorter sustained attention. Ultimately, this led to 265 complete responses.

Data cleaning

Of the 265 complete responses, 7.2% (n = 19) was excluded from analyses. Eleven subjects were excluded as they scored responses with minimal 8 words in perfect sequence and three subjects with a learning curve below the minimum (see Study 1 for exclusion criteria). Five subjects were excluded because of an age below 18 years. This led us to include 246 subjects in the final analysis.

Comparison on demographic characteristics

The two Dutch samples did not differ in terms of sex, age, education and device used to complete the test.

Comparison of learning curves between new and original Dutch word list

No differences in trial scores were found between the new and original Dutch word list. Data were best represented by a model that only included the main factor trial. Compared to the same model with version, the BF_{10} was 12.03, indicating strong evidence for similarity of scores between word lists. Post hoc comparisons (Bayesian *t*-tests controlled for multiplicity) showed that scores increased with each trial ($BF_{10} \geq 11898.29$). See Figure 6 for average learning curves of the new and original Dutch RAVLT.

Reliability of trial scores

The reliability of each trial score of the new word list was highly comparable to those of the original word list. We found strong correlations between all trial scores and the total score summed over five trials (95% credible intervals ranged from .68 to .94). See Table 7 for Bayesian Pearson's *r* correlations for all trials.

Comparison serial position effects

No differences were found in serial position effects between the new and original Dutch word lists. Both word lists showed primacy and recency effects. Data was best described by the model including serial position as predictor. This model fit the data better than the model including serial position, version and interaction between serial position and version (BF = 6.98), indicating substantial evidence for similarity of serial position effects. Both the original and new word lists showed a primacy effect (95% Credible Interval= .82-.87 vs. .60-.65 and .80-.85 vs .62-.67) and recency effects (.67-.73 vs. .60-.65 and .71-.77 vs. .62-.67). See Table 8 for percentages of correct response per serial position. See Figure 7 for the serial position curve of the new and original word lists.

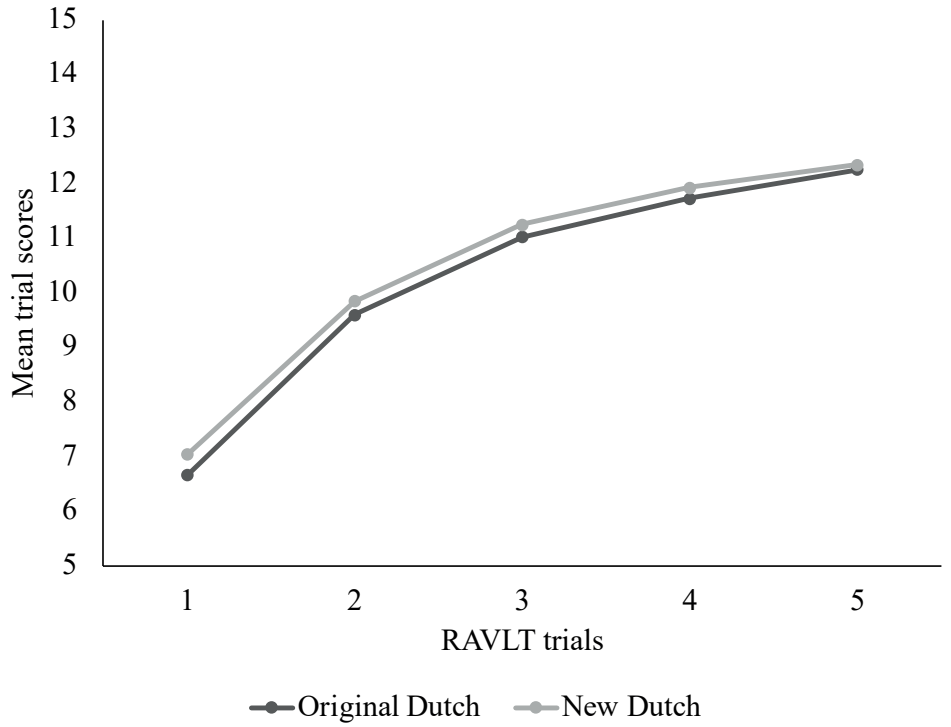


Figure 6. Learning curves for trials 1 to 5 on the new and original Dutch RAVLT

Abbreviation: RAVLT = Rey Auditory Verbal Learning Test.

Item-level comparison of proportion correct responses

Corrected for serial position, items from both word lists were learned comparably well. Overall, we found similar residuals for items from both word lists, with all data points between -.16 and .10. Figure 8 depicts proportion correct responses over all trials per item per test, corrected for serial position.

Table 7
Bayesian Pearson's r correlations between trial scores and total score of the Dutch word lists

Version	Variable		Trial 1 score	Trial 2 score	Trial 3 score	Trial 4 score	Trial 5 score
Original Dutch	Total score	Pearson's r	.77	.91	.88	.89	.81
		BF_{10}	$1.28*10^{21}$	$8.24*10^{41}$	$8.28*10^{35}$	$1.65*10^{36}$	$2.46*10^{25}$
		95% CI	.68-.83	.87-.94	.83-.92	.83-.92	.74-.86
New Dutch	Total score	Pearson's r	.78	.87	.87	.88	.84
		BF_{10}	$2.18*10^{24}$	$3.28*10^{37}$	$1.06*10^{37}$	$1.40*10^{39}$	$2.12*10^{32}$
		95% CI	.68-.83	.82-.90	.81-.90	.83-.91	.78-.88

NOTE. Total score is the summed score over all five trials. CI = credible interval.

Table 8
Percentages correct responses per serial position of the Dutch word lists

Version	Serial position effect	Mean % correct responses	95% CI
Original Dutch	Primacy	.84	.82 - .87
	Middle	.63	.60 - .65
	Recency	.70	.67 - .73
New Dutch	Primacy	.82	.80 - .85
	Middle	.64	.62 - .67
	Recency	.74	.71 - .77

NOTE. Total score is the summed score over all five trials. CI = credible interval.

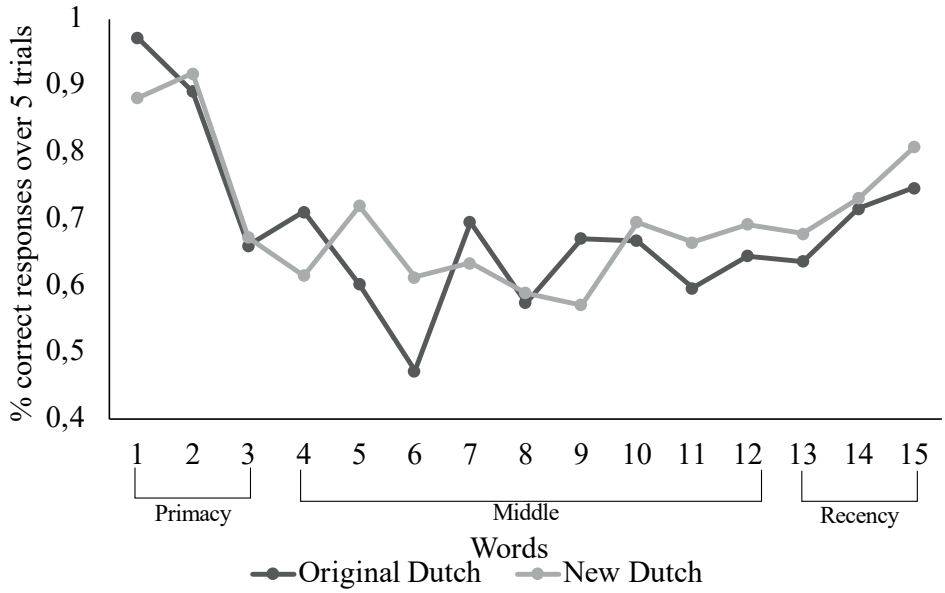


Figure 7. Serial position curves for the new and original Dutch version of the RAVLT NOTE. Word recall on the y-axis represents the percentage of correct responses per item over all five trials.

Figure 8. Deviations from expected proportion of correct responses based on serial position per word for the new and original Dutch RAVLT
Abbreviation: RAVLT = Rey Auditory Verbal Learning Test.

NOTE. The word items are depicted by icons to avoid the word items to be publicly known by future subjects. Upon request, the new Dutch word list is freely available for researchers.

However, in the new Dutch RAVLT version, the second word was better learned than the first word (and every other word), indicating an unexpected high recall score and need for alteration of word (order).

Discussion

The main goal of this study was to present cross-lingual word criteria for RAVLT(-based) word list development. We identified relevant word criteria, we developed two new American-English word lists and one new Dutch list based on these criteria, and we compared performance on the new word lists with the original lists on trial scores, serial position effects, and proportion

of correct responses per item corrected serial position. The findings showed, in general, that the two new American-English word lists were similar in most outcome measures, but had lower trial scores compared to the original American-English word list. The new and the original Dutch word list were similar in terms of all outcome measures. The results lead us to believe that the word criteria are useful for systematic development of parallel versions of RAVLT(-based) word lists.

Application of the newly developed word criteria in both American-English and Dutch were shown to be feasible and led to two new American-English word lists and one Dutch list, with —as expected from RAVLT(-based) word lists[52]— increasing learning curves and clear serial position effects with a U-shaped curve. Subjects tended to learn more words over time and words that were presented first (primacy) or last (recency) in the list were learned better than words that were presented in the middle of the list.

Compared to the original American-English word list, we found that the two new American-English word lists had lower scores on every trial. The difference in trial scores between the new and the original American-English word lists can possibly be explained by differences in applied word criteria. The original word list not only used different word criteria (i.e., mono and disyllabic words vs. only monosyllabic, and high word frequency vs. medium frequency), but also less word criteria (4 criteria vs. 13 criteria) and less stringent/specified word criteria (i.e., unspecified "high word frequency" vs. specified word frequency of 5-100 per million words). As a result, in the original list, several phonetic and semantic associations between words were not controlled for, possibly making the list easier to learn by grouping of items or compound cueing. Due to application of the word criteria in the new word lists, most word associations were controlled for, potentially resulting in a word list that is more difficult to learn. Consequently, the new word lists cannot be used as alternate versions of the original list, but this was not our goal. Our goal was to present word criteria that can help development of new lists for computerized assessment that are equivalent to one another.

Comparison between the two new American-English word lists showed that the word lists were similar in all outcome measures, except for the primacy effect, whereby the second new word list had a stronger primacy effect than the first new word list. This finding seems to be the result of the relatively high recall rate (92.1%) of the first item of the new American-English word list 2 and a relatively low recall rate (66.1%) of the second item of the new American-English word list 1. Although both word lists were made using the same word selection criteria, the second new American-English word list adhered to fewer criteria (see Table 2), possibly leading to unforeseen associations between words and/or executive function strategies explaining the difference in primacy effect.

Comparison between the new and original Dutch word lists showed that the new word list yielded highly similar learning curves and serial position effects to the Dutch original RAVLT word list, more so than the new American-English word lists did to their original. Higher similarity of performance on the new and original Dutch word lists can possibly be explained by higher similarity of word criteria used for development of the new and the original Dutch word list. In fact, the new word criteria were partly based on the criteria of the original Dutch list.[3] Moreover, qualitative analyses of characteristics of the original Dutch word list showed that in general, items of the original word list complied relatively well to the new word criteria.

Item-level analyses showed no large fluctuations in proportions of correct responses controlled for serial position in all word lists; however, the new American-English and Dutch word lists seemed to contain fewer words with deviating proportion of correct responses. This finding can possibly be explained as a consequence of stricter control of word characteristics of the new word lists. Due to extensive application of word criteria, it is likely that there was less variability in word memorability, word relatedness and differential item functioning. This made the serial position of words in the new word lists a better predictor of total word recall than in the original word

lists and thus less words with deviating proportions of correct responses were found.

Furthermore, item-level analyses showed that three words were recalled remarkably well. The first word of the new American-English word list 2 (92.1% over all 5 trials), the second word of the new Dutch list (92.0% over all 5 trials) and the first word of the original Dutch list (97.3% over all 5 trials). The high rate of recall for these three words could be explained as the combination of their serial position (these words were presented first and therefore learned better: primacy effect) and another factor. Their memorability was likely driven by a size effect (words referring to larger objects), an animacy effect (words referring to living entities) and a valence effect. All three characteristics have been shown to positively influence memorability.[4] By presenting a word that refers to larger objects, an animate word, or a positively valenced word first in the list, the rate of recall became even higher.

Two important implications of this study can be identified. First, application of our newly developed word criteria can lead to similar new RAVLT(-based) word lists across languages. Given that many word characteristics can influence memorability of words, we highly recommend strict control of word criteria per language over simple translations of word lists to other languages.[53] For the American-English and Dutch language, the criteria proved to be feasible although some loosening of the criteria was required in order for a sufficient number of words to remain. As the word criteria we defined are quite specific, it is probable that not all criteria can be followed perfectly in all language versions. To maximize equivalence between alternate forms and translations, however, approximation of the criteria should be pursued. Future studies will show the usability of the criteria in other languages. The word criteria are currently used for development of new RAVLT(-based) word lists for British-English and Danish versions of the ACS. Moreover, future studies should evaluate the usability of the word criteria for other sections of the RAVLT (such as the

distraction and recognition word list) and examine test performance on the new word lists on the delayed recall and recognition phase. Also, future studies could compare performance to other alternate versions and examine performance in clinical samples. Lastly, future studies should conduct cross-cultural comparisons of the new RAVLT-based word lists. We expect that word lists based on identical word criteria are equivalent across languages due to highly comparable word properties. This would need to be tested, for example in the context of an international data pooling initiative.

A second important implication is that different word lists can be equivalent in terms of trial scores/learning curves but simultaneously be inequivalent in terms of serial-position effects and on item-level. Given the contemporarily increasing use of serial-position effects and item-level analyses as a marker to find subtle group differences, this finding stresses the need for comparisons of new word lists on the item level prior to comparing or pooling results from different word lists.

A limitation of this study is that we did not administer the new and original word list in the same sample of participants. We did not do so to avoid practice effects or a complicated cross-over design to correct for test-retest effects. Thus, we cannot rule out that between-group differences influenced the results. However, analyses did not show any between-group differences in demographic variables in the American and Dutch samples, except for a difference in education level between the American samples. We, therefore, adjusted the repeated measures ANOVA's for education level. Another limitation is the absence of a test leader during the online test. Therefore, we cannot guarantee that all participants fully adhered to all test instructions and maximized their efforts. However, we applied several security checks and excluded participants who did not adhere to prespecified criteria. A last limitation is the application of the minimal learning curve based on Dutch data to the American sample. As we have not directly investigated the cross-cultural applicability of these cutoffs, it could be that the minimal learning curve was too liberal for the American data, resulting in the inclusion of

relatively low learning curves. It is recommended to use language-specific cutoffs in future studies.

In the evaluation of our word criteria, we made use of a computerized visual presentation of the words. Visual presentation has been used before in studies, but is not the typical way of presentation in neuropsychological practice, where the words are typically presented in an auditory mode. Previous research has shown that there are differences in performance between these presentation modes, with better performance on the first trial with auditory presentation, and better performance on trials 3 and 4 for visually presented words (van der Elst et al., 2005).[54] Therefore, we cannot assume that these presentation modes are equivalent. However, in formulating the word criteria, we had already considered that participants may use different strategies in visualizing auditory presentations or verbalizing visual presentations. Thus, we formulated criteria which should standardize words in a manner that is generic to both visual and verbal presentation (cf., criterion 2 monosyllabic, criterion 7 no phonological association between words).

In this study, we have chosen to limit our attention to the selection of words for the learning trials of the RAVLT, ignoring delayed recall and recognition trials. This was largely a practical consideration as we only administered a single neuropsychological test online via the computer, and thus did not have sufficient control or time to present a delayed recall and recognition trial after 15-30 minutes. However, we would argue that the same word criteria that would optimize the sensitivity of our measurement of learning would be appropriate for the measurement of delayed recall. We are more concerned with the measurement of recognition ability. The learning trials of the RAVLT have been designed to primarily measure encoding while delayed recall is associated with retrieval, but this distinction is not as clear-cut as the distinction between recall and recognition. Recognition trials show a ceiling effect, limiting their sensitivity to individual differences, which may be reduced by more careful selection of distractor words. Therefore,

we would suggest that in future research a similar effort is made for the formulation of criteria for the selection of words for recognition trials.

Also, in this study, we have compared performance on the new versions to performance on the traditional version of the RAVLT. There are several other word lists currently in use, especially for use as alternatives to the traditional version in repeated testing. Also, there are different verbal learning tests that are used with different words (notably, the Hopkins Verbal Learning Test and California Verbal Learning Test, although these tests have more differences with the RAVLT). Because the existing alternate versions were not necessarily constructed with the same criteria as the traditional versions (due to a lack of such criteria), it is difficult to say whether the results would be equivalent for these alternate versions. With our current set of criteria, we indeed hope to improve the a priori comparability of alternate versions by controlling the word characteristics more stringently across parallel versions.

In sum, the results of the current study provide cross-lingual word criteria for creation of new and alternate forms of RAVLT(-based) word lists. The word criteria will be used within our ACS, but can also be used for other RAVLT(-based) tests. Importantly, the word criteria could be adapted for improvement of other verbal learning and memory tests as well. Even more, the word criteria are just an example for modern improvements of stimulus selection in cognitive tests. As such, we can optimize cognitive tests to make them comparable internationally and over time.

References

1. Rey A. L'examen psychologique dans les cas d'encéphalopathie traumatique. *Arch Psychol*. 1941;28;215-285.
2. Rey A. *L'examen clinique en psychologie*. Paris, FR: Press Universitaire de France; 1958.
3. Brand N, Jolles J. Learning and retrieval rate of words presented auditorily and visually. *J Gen Psychol*. 1985;112(2):201-210.
4. Madan CR. Exploring word memorability: how well do different word properties explain item free-recall probability? *Psychon Bull Rev*. 2021;28(2):583-595.
5. Bilder RM, Reise SP. Neuropsychological tests of the future: how do we get there from here? *Clin Neuropsychol*. 2019;33(2):220-245.
6. Feenstra HEM, Murre JMJ, Vermeulen IE, Kieffer JM, Schagen SB. Reliability and validity of a self-administered tool for online neuropsychological testing: the Amsterdam Cognition Scan. *J Clin Exp Neuropsychol*. 2018;40(3):253-273.
7. Feenstra HEM, Vermeulen IE, Murre JMJ, Schagen SB. Online self-administered cognitive testing using the Amsterdam Cognition Scan: establishing psychometric properties and normative data. *J Med Internet Res*. 2018;20(5):e192.
8. Salthouse TA. Item analyses of memory differences. *J Clin Exp Neuropsychol*. 2017;39(4):326-335.
9. Hourihan KL, Fraundorf SH, Benjamin AS. The influences of valence and arousal on judgments of learning and on recall. *Mem Cognit*. 2017;45(1):121-136.
10. Brunet HE, Kramer JH, Lupas GJ, Foley JM. Strategy use and verbal memory in older adults: The role of intellectual functioning and the preferential impact of semantic clustering. *Clin Neuropsychol*. 2020;34(1):204-216.
11. Boake C. Edouard Claparede and the auditory verbal learning test. *J Clin Exp Neuropsychol*. 2000;22(2):286-292.
12. Taylor EM. *Psychological appraisal of children with cerebral defects*.

Cambridge, MA: Harvard University Press; 1959.
13. Shapiro DM, Harrison DW. Alternate forms of the AVLT: a procedure and test of form equivalency. *Arch Clin Neuropsychol*. 1990;5(4):405-410.
14. Ellis R. Factors in the incidental acquisition of second language vocabulary from oral input. In: Ellis R, ed. *Learning a second language through interaction*. Amsterdam, NL: John Benjamins Publishing Company; 1999:35-61.
15. Ellis N, Beaton A. Factors affecting the learning of foreign language vocabulary: imagery keyword mediators and phonological short-term memory. *Q J Exp Psychol*. 1993;46(3):533-558.
16. Bhatarah P, Ward G, Smith J, Hayes L. Examining the relationship between free recall and immediate serial recall: similar patterns of rehearsal and similar effects of word length, presentation rate, and articulatory suppression. *Mem Cognit*. 2009;37(5):689-713.
17. Brysbaert M, Warriner AB, Kuperman V. Concreteness ratings for 40 thousand generally known English word lemmas. *Behav Res methods*. 2014;46(3):904-911.
18. Paivio A, Yuille JC, Madigan SA. Concreteness, imagery, and meaningfulness values for 925 nouns. *J Exp Psychol*. 1968;76(1p2):1-25.
19. Walker I, Hulme C. Concrete words are easier to recall than abstract words: evidence for a semantic contribution to short-term serial recall. *J Exp Psychol Learn Mem Cogn*. 1999;25(5):1256-1271.
20. Klaver P, Fell J, Dietl T, et al. Word imageability affects the hippocampus in recognition memory. *Hippocampus*. 2005;15(6):704-712.
21. Paivio A. *Mental representation: a dual-coding approach*. New York, NY: Oxford Univ. Press; 1986.
22. Schwanenflugel PJ, Akin C, Luh WM. Context availability and the recall of abstract and concrete words. *Mem Cognit*. 1992;20(1):96-104.
23. McDougall S, Pfeifer G. Personality differences in mental imagery and the effects on verbal memory. *Br J Psychol*. 2012;103(4):556-573.
24. Pexman PM, Hargreaves IS, Edwards, JD, Henry LC, Goodyear BG. Neural

correlates of concreteness in semantic categorization. *J Cogn Neurosci*. 2007;19(8):1407-1419.
25. Adelman JS, Estes Z. Emotion and memory: a recognition advantage for positive and negative words independent of arousal. *Cognition*. 2013;129(3):530-535.
26. Unsworth N, Miller AL, Robison MK. Individual differences in encoding strategies and free recall dynamics. *Q J Exp Psychol*. 2019;72(10):2495-2508.
27. Poirier M, Saint-Aubin J. Memory for related and unrelated words: further evidence on the influence of semantic factors in immediate serial recall. *Q J Exp Psychol*. 1995;48(2):384-404.
28. Chance FS, Kahana MJ. Testing the role of associative interference and compound cues in sequence memory. In: Bower JM, ed. *Computational Neuroscience*. Boston, MA: Springer;1997:599-603.
29. Posner MI, Konick AF. On the role of interference in short-term retention. *J Exp Psychol*. 1966;72(2):221-231.
30. Li X, Schweickert R, Gandour J. The phonological similarity effect in immediate recall: Positions of shared phonemes. *Mem Cognit*. 2000;28(7):1116-1125.
31. Brysbaert M, Mandera P, Keuleers E. The word frequency effect in word processing: An updated review. *Curr Dir Psychol Sci*. 2018;27(1):45-50.
32. Baluch B, Danaye-Tousie M. Memory for words as a function of spelling transparency. *J Psychol*. 2006;140(2):95-104.
33. De Deyne S, Storms G. Age-of-acquisition differences in young and older adults affect latencies in lexical decision and semantic categorization. *Acta Psychol*. 2007;124(3):274-295.
34. Della Rosa PA, Catricalà E, Vigliocco G, Cappa SF. Beyond the abstract—concrete dichotomy: mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. *Behav Res Methods*. 2010;42(4):1042-1048.
35. Zevin JD, Seidenberg MS. Age of acquisition effects in word reading and other tasks. *J Mem Lang*. 2002;47(1):1-29.

36. Herrmann DJ, Crawford M, Holdsworth M. Gender-linked differences in everyday memory performance. *Br J Psychol*. 1992;83(2):221-231.
37. Baer A, Trumpeter NN, Weathington BL. Gender differences in memory recall. *Mod Psychol Stud*. 2006;12(1):11-16.
38. Rosselli M, Ardila A, Jurado MB, Salvatierra JL. Cognate facilitation effect in balanced and non-balanced Spanish–English bilinguals using the Boston Naming Test. *Int J Biling*. 2014;18(6):649-662.
39. Rodriguez TA. From the known to the unknown: using cognates to teach English to Spanish-speaking literates. *Read Teach*. 2001;54(8):744-746.
40. Kuperman V, Stadthagen-Gonzalez H, Brysbaert M. Age-of-acquisition ratings for 30,000 English words. *Behav Res Methods*. 2012;44(4):978-990.
41. Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav Res Methods*. 2013;45(4):1191-1207.
42. Brysbaert M, New B. Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav Res Methods*. 2009;41(4):977-990.
43. Hauser D, Paolacci G, Chandler J. Common concerns with MTurk as a participant pool: evidence and solutions. In Kardes FR, Herr PM, Schwarz N, eds. *Handbook of research methods in consumer psychology*. New York, NY: Routledge; 2019:319-337.
44. Wefel JS, Kesler SR, Noll KR, Schagen SB. Clinical characteristics, pathophysiology, and management of noncentral nervous system cancer-related cognitive impairment in adults. *CA Cancer J Clin*. 2015;65(2):123-138.
45. JASP Team (2020). JASP (Version 0.14.1) [Computer software].
46. de Vent NR, Agelink van Rentergem JA, Schmand BA, Murre JMJ, Huizenga HM, ANDI Consortium. Advanced Neuropsychological Diagnostics Infrastructure (ANDI): a normative database created from control datasets. *Front Psychol*. 2016;7(1601):1-10.

47. Glanzer M, Cunitz AR. Two storage mechanisms in free recall. *J Verb Learn Verb Behav*. 1966;5(4):351-360.
48. Brysbaert M, Stevens M, De Deyne S, Voorspoels W, Storms G. Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychol*. 2014;150:80-84.
49. Moors A, De Houwer J, Hermans D, et al. Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behav Res Methods*. 2013;45(1):169-177.
50. Keuleers E, Brysbaert M, New B. SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behav Res Methods*. 2010;42(3):643-650.
51. Verhage F. *Intelligentie en leeftijd: onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar*. Assen, NL: van Gorcum; 1964.
52. Lezak MD, Howieson DB, Loring DW, Fischer JS. *Neuropsychological assessment*. New York, NY: Oxford University Press; 2004.
53. Boenniger MM, Staerk C, Coors A, Huijbers W, Ettinger U, Breteler MM. Ten German versions of Rey's auditory verbal learning test: age and sex effects in 4,000 adults of the Rhineland Study. *J Clin Exp Neuropsychol*. 2021;43(6):637-653.
54. van der Elst WIM, van Boxtel MPJ, van Breukelen GJP, Jolles J. Rey's verbal learning test: normative data for 1855 healthy participants aged 24–81 years and the influence of age, sex, education, and mode of presentation. *J Int Neuropsychol Soc*. 2005;11(3):290-302.

Supplementary material

Table S1
Psycholinguistic characteristics of the original and new American word lists

Original American						
Word	Length	WF	Concreteness	AoA	% known	Valence
1	4	8.47	4.96	4.63	100	6.05
2	7	10.29	4.82	4.95	100	5.36
3	4	39.33	4.96	3.89	100	5.67
4	6	144.53	4.81	4.94	96	7
5	6	333.12	4.79	3.89	100	5.41
6	6	13.14	4.56	4.22	100	6.73
7	4	49.96	4.90	4.83	100	7
8	6	26.55	4.73	5.33	100	7.25
9	3	64.18	4.88	3.33	100	5.69
10	6	11.84	4.54	4.74	100	6.14
11	4	69.75	4.89	2.95	100	5.5
12	6	22.61	4.89	3.95	100	5.9
13	5	39.43	4.08	4.00	100	7.05
14	5	514.00	5.00	3.16	100	7.19
15	5	55.47	4.89	4.90	100	6.72
Mean	5.13	93.51	4.78	4.25	99.73	6.31

New American 1						
Word	Length	WF	Concreteness	AoA	% known	Valence
1	5	10.10	4.97	4.30	100	6.11
2	4	16.43	4.82	5.50	100	5.53
3	4	18.90	4.83	6.78	100	4.45
4	4	27.00	4.85	4.58	96	5.82
5	4	14.69	5.00	4.42	100	4.7
6	5	11.61	4.93	5.33	100	5.57
7	4	12.35	4.93	5.45	100	6.05
8	5	58.63	5.00	4.42	100	5.65
9	5	92.88	5.00	4.15	100	6.05
10	5	21.73	4.81	3.63	100	7
11	4	22.71	4.93	5.44	100	5.05
12	4	11.10	4.86	5.11	100	5.65
13	4	25.88	5.00	4.79	100	6.1
14	4	43.90	4.87	5.56	100	5.56
15	5	9.63	4.92	5.40	100	6.4
Mean	4.4	26.50	4.91	4.99	99.73	5.71

Table S1 continued

New American 2						
Word	Length	WF	Concreteness	AoA	% known	Valence
1	4	51.94	4.67	5.35	100	5.89
2	5	46.37	4.94	3.53	100	5.56
3	4	43.65	4.90	4.42	100	6.62
4	4	15.43	4.72	6.48	100	NA
5	4	59.51	5.00	3.00	100	5.44
6	4	12.00	4.93	5.32	100	5.45
7	4	12.88	4.97	4.00	100	5.74
8	5	26.88	4.89	4.40	100	6.9
9	5	10.18	4.83	6.43	100	4.65
10	4	68.84	4.90	4.42	100	5.5
11	4	24.55	4.41	6.11	100	5.71
12	4	10.45	4.86	6.37	100	5.8
13	5	27.06	4.86	4.40	100	5.9
14	4	64.51	4.56	5.05	100	7.27
15	4	35.65	4.79	5.00	100	4.48
Mean	4.27	33.99	4.82	4.95	100	5.78

Abbreviations: WF = word frequency, AoA = age of acquisition, NA = not available.

NOTE. For the word ratings, we made use of different word corpora: concreteness ratings,[17] AoA ratings,[40] valence ratings,[41] and word frequency ratings.[42]

Table S2
Psycholinguistic characteristics of the original and new Dutch word list

Original word list						
Word	Length	WF	Concreteness	AoA	% known	Valence
1	5	13.49	4.67	4.27	100	5.31
2	4	9.95	4.8	5.12	100	3.81
3	5	14.32	4.93	5.56	100	4.75
4	4	25.45	4.93	4.22	99.9	4.33
5	3	3.43	4.00	6.53	98.2	3.64
6	4	23.90	4.47	5.70	99.9	4.13
7	5	1.99	4.87	6.10	100	5.00
8	3	2.38	4.47	6.46	100	4.02
9	3	58.36	4.87	5.37	100	4.36
10	4	34.30	4.73	5.86	100	4.45
11	5	1.56	4.80	8.27	100	4.17
12	4	n.a.	4.91	9.11	87	NA
13	4	6.06	4.53	7.44	100	3.17
14	4	23.55	4.53	6.65	100	3.92
15	3	6.75	4.73	5.26	100	3.97
Mean	4.00	16.11	4.68	6.13	99	4.22

New word list						
Word	Length	WF	Concreteness	AoA	% known	Valence
1	5	58.20	4.93	5.89	100	4.53
2	4	8.99	4.87	4.81	100	4.41
3	4	13.88	4.67	4.89	100	4.36
4	4	14.22	4.87	3.90	100	4.55
5	4	42.10	4.93	5.34	100	4.81
6	4	70.04	4.93	3.95	100	4.06
7	5	11.21	4.73	6.08	100	4.03
8	4	45.71	4.93	4.91	100	4.30
9	4	8.83	4.60	5.95	100	4.16
10	4	18.75	4.93	4.12	100	4.84
11	4	14.48	4.80	4.92	100	3.83
12	5	73.15	5.00	4.89	100	4.25
13	4	5.19	4.87	4.54	100	4.05
14	4	44.07	4.73	5.47	100	4.39
15	4	4.16	4.73	6.87	100	4.11
Mean	4.20	42.67	4.83	5.10	100	4.31

Abbreviations: WF = word frequency, AoA = age of acquisition, NA = not available.

NOTE. For the word ratings, we made use of different word corpora: concreteness ratings and AoA ratings,[48] valence ratings,[49] WF ratings.[50]

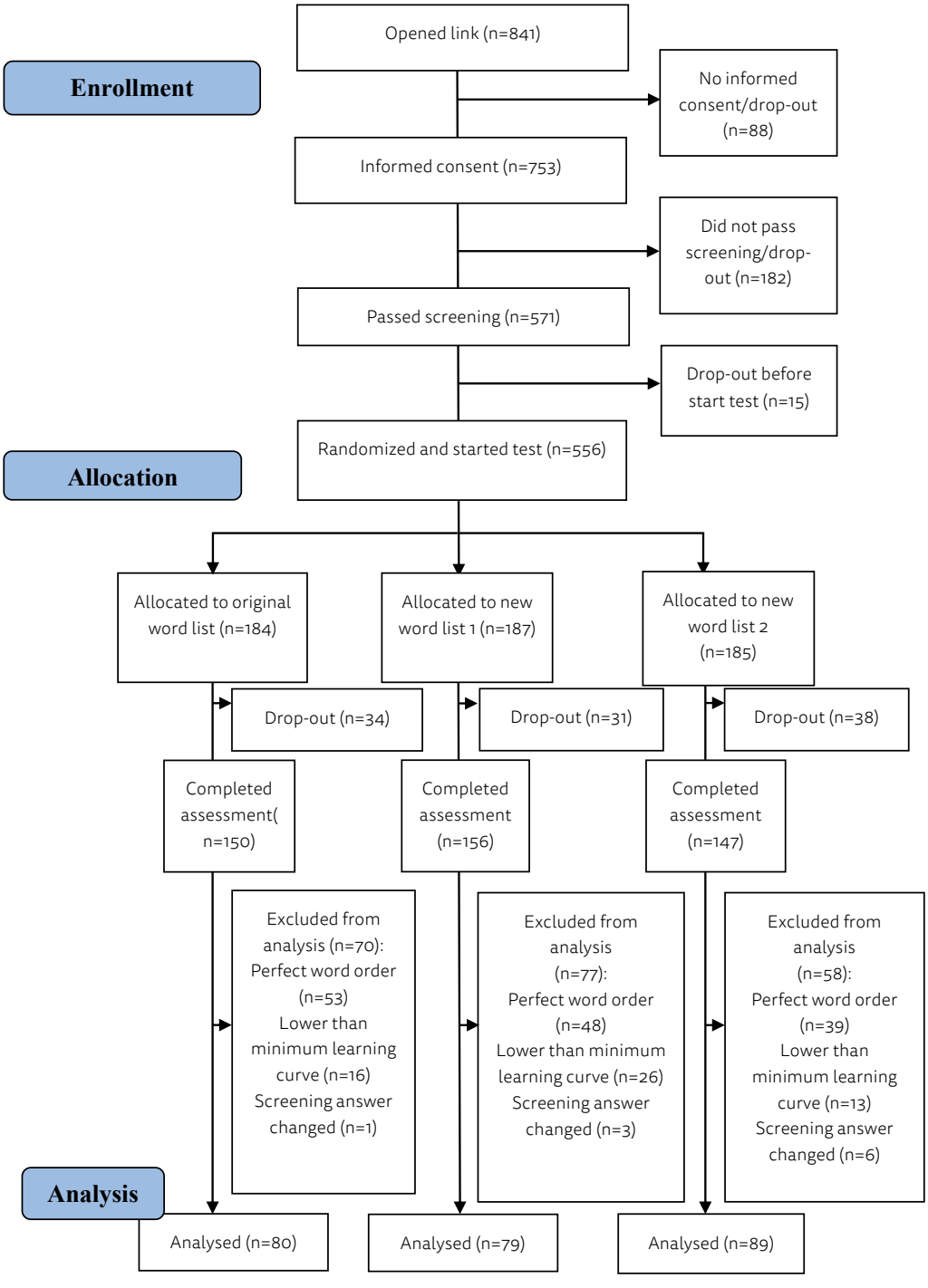


Figure S1. Flow-chart of participation and completion rates American sample

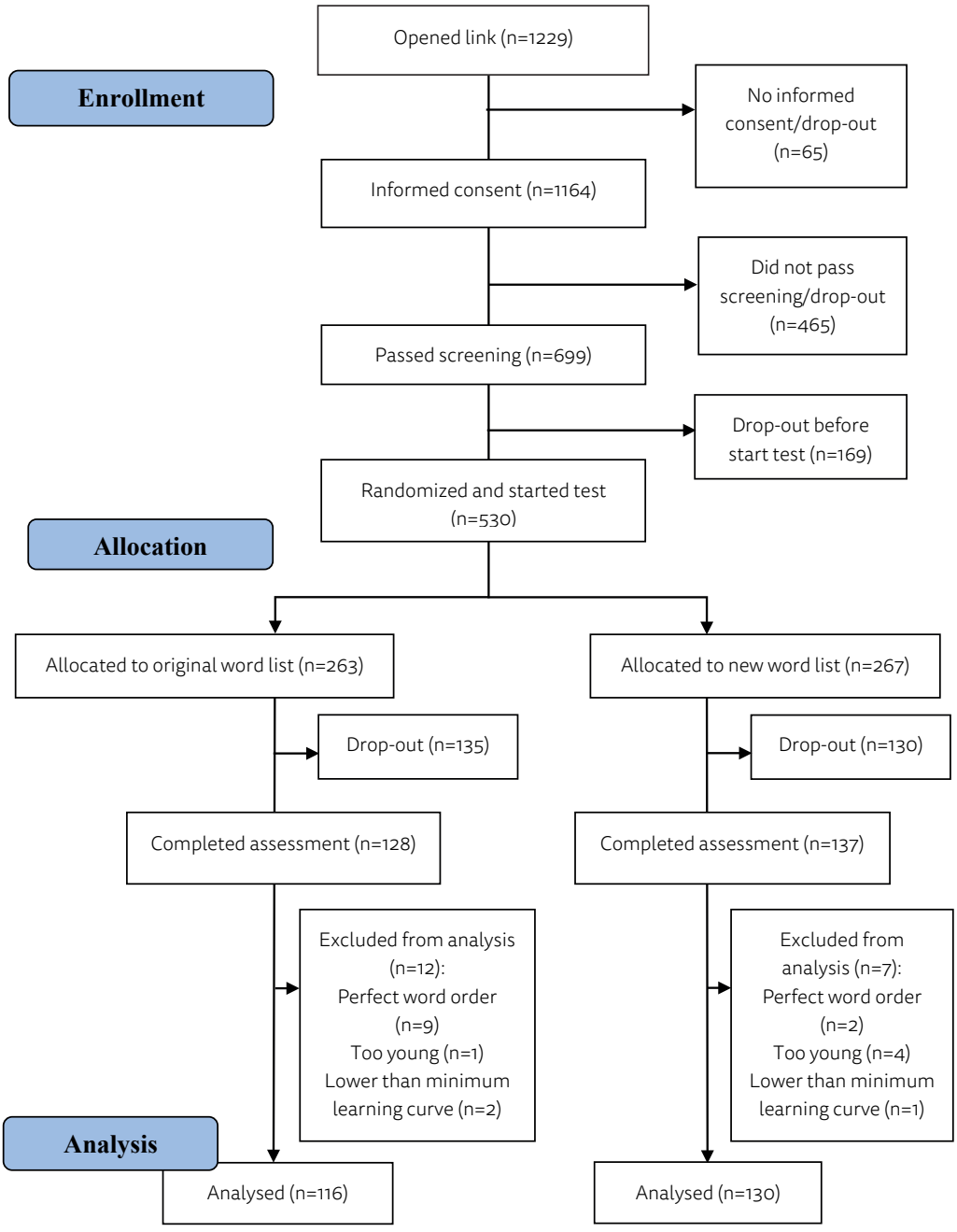


Figure S2. Flow-chart of participation and completion rates Dutch sample

Chapter 4:
Subgroups of cognitively affected and unaffected breast cancer survivors after chemotherapy: A data-driven approach

Joost A. Agelink van Rentergem*, Philippe R. Lee Meeuw Kjoe*,
Ivar E. Vermeulen, Sanne B. Schagen.

*Shared first authorship.
Journal of Cancer Survivorship. 2023 [published online ahead of print].

Abstract

Purpose

It is assumed that a segment of breast cancer survivors are cognitively affected after chemotherapy. Our aim is to discover whether there is a qualitatively different cognitively affected subgroup of breast cancer survivors, or whether there are only quantitative differences between survivors in cognitive functioning.

Methods

Latent profile analysis was applied to age-corrected neuropsychological data —measuring verbal memory, attention, speed, and executive functioning— from an existing sample of 62 breast cancer survivors treated with chemotherapy. Other clustering methods were applied as sensitivity analyses. Subgroup distinctness was established with posterior mean assignment probability and silhouette width. Simulations were used to calculate subgroup stability, posterior predictive checks to establish absolute fit of the subgrouping model. Subgrouping results were compared to traditional normative comparisons results.

Results

Two subgroups were discovered. One had cognitive normal scores, the other —45%— had lower scores. Subgrouping results were consistent across clustering methods. The subgroups showed some overlap; 6% of survivors could fall in either. Subgroups were stable, and described the data well. Results of the subgroup clustering model matched those of a traditional normative comparison method requiring small deviations on two cognitive domains.

Conclusions

We discovered that almost half of breast cancer survivors after chemotherapy form a cognitively affected subgroup, using a data-driven approach. This proportion is higher than previous studies using prespecified cutoffs observed.

Implications for Cancer Survivors

A larger group of cancer survivors may be cognitively affected than previously recognized, and a less strict threshold for cognitive problems may be needed in this population.

Introduction

Breast cancer and chemotherapy influence cognitive functioning.[1-2] Cognitive performance is lower in survivors compared to controls,[2] and decreases in survivors after treatment.[3] It is broadly assumed that only a subgroup of survivors is cognitively affected,[4-5] but it is unclear whether this subgroup concerns the lower end of a continuous scale of quantitative differences, or concerns a qualitatively different subgroup.[6] In case of the latter, we do not know how large this subgroup would be.

Researchers often evaluate cognitive effects of disease or treatments using mean difference comparisons, to compare groups of survivors and controls. Mean difference comparisons show that cancer survivors who received chemotherapy obtain lower scores than no-cancer controls, and lower scores than cancer survivors who did not receive chemotherapy, on average.[2] A second method to investigate cognitive effects is to use normative comparisons: Scores of individual survivors are compared to the distribution of scores in a reference group, to determine whether the survivors' scores are below normal. If so, survivors are defined as cognitively impaired. In general, breast cancer survivors who receive chemotherapy are more likely to be cognitively impaired than healthy controls.[3]

Both approaches have their flaws. Mean comparisons only reflect differences at the group level, which may not hold at the individual level. Normative comparisons are dependent on where the cutoff is placed between affected and unaffected cognitive functioning. Various cutoffs are in use, and these are not driven by the etiology of cancer-related cognitive impairment. Therefore, neither mean comparisons nor normative comparisons can substantiate whether there is cognitive impairment in only a subgroup of cancer survivors, nor can they reveal how large this subgroup would be.

Identification of subgroups of survivors is important for treatment of cognitive problems and determination of risk factors. If we can identify subgroups of survivors who are vulnerable to cognitive problems, we can develop more targeted interventions. When we know which characteristics are associated with subgroup membership, we can identify high-risk individuals, which can provide us with better informed decision-making regarding use of therapies.

To identify whether variation in scores suggests the existence of a subgroup, a different class of methods is required. Latent variable modeling and machine learning provide several approaches that use multivariate data to detect whether outcomes are best described by one, two, or even more subgroups. With these methods, the described number of subgroups and their scores are not predetermined, but estimated from the data at hand.

Methods that use multivariate data to form subgroups are common in other subfields of oncology, such as in subclassification of tumor types,[7-8] but they have not been used to form subgroups of tested cognitive functioning after cancer. There are some exceptions in pediatric oncology, where two to four subgroups were identified using both neuropsychological test scores and other variables.[9-11] Apart from examining a different population, these studies differed from ours in other ways; e.g., fewer cognitive tests were administered in these studies. Other studies that have examined subgroups differed from ours by examining the development of test scores over time, rather than the profile of test scores, focusing on a single outcome at a time.[12]

If subgroups exist, subgrouping methods aid in the practical formulation of cutoff points for normative comparisons, which is a point of concern.[13] This point is illustrated in Figure 1. In Panel a, a hypothetical ground truth is presented where scores come from a single distribution, which can be arbitrarily divided into two at any given point. In Panel b —a second hypothetical ground truth— scores come from one of two

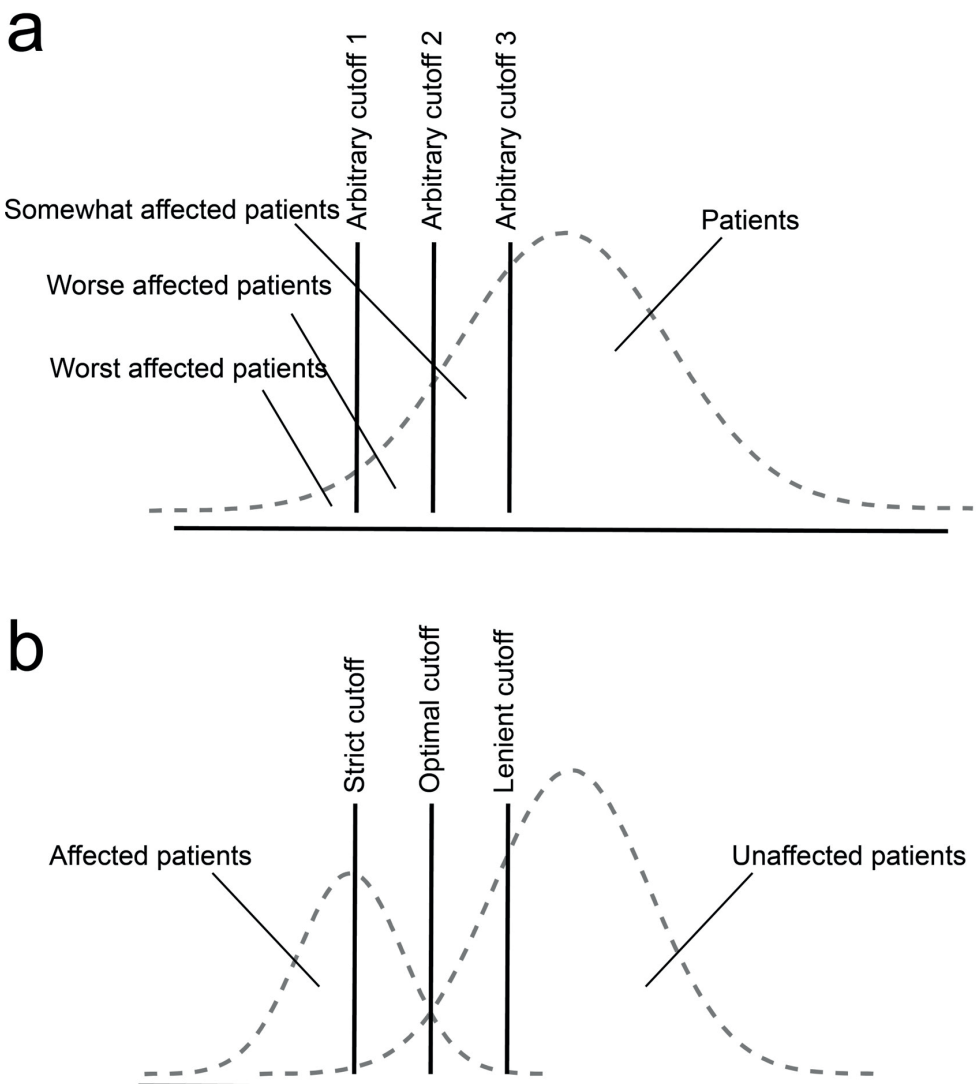


Figure 1. Illustration of hypothetical situations where patients' scores come from either one or two distributions

Fig.1a All scores come from a single distribution cut into two.
Fig.1b A minority of patients obtains scores from a second distribution.

distributions: an affected or an unaffected subgroup. In this latter scenario, separating survivors into two categories is reasonable, but where the cutoff point is placed may be optimal or suboptimal. In current applications of normative comparisons, we are unsure whether we are in scenario A and are dividing a continuous distribution, or whether we are in scenario B. If we are in scenario B, we may be employing suboptimal cutoff points to separate the two distributions.

The aim of this study is to use subgrouping methods to discover whether there truly are subgroups of breast cancer survivors that are cognitively affected or unaffected, or whether there is only continuous variation. Secondly, if there are affected subgroups, we aim to discover how large these subgroups are, and whether they correspond to the group of survivors identified with traditionally applied normative comparisons.

Methods

Participants

Data were extracted from a dataset that has previously been described, and was used to validate a computerized neuropsychological test battery.[14-15] For this dataset, a sample of 202 non-CNS cancer survivors participated after having been treated with intensive therapies, as well as a large sample of no-cancer controls. This study was approved by the review board of the Netherlands Cancer Institute. All participants provided written informed consent prior to assessment. For the current study, we selected women who were treated for breast cancer with chemotherapy (N=67), as the majority of studies on cognitive effects are conducted in this survivor population.[1] Additionally, we selected 248 participants without a history of cancer as a reference group for the purposes of age correction, standardization, and normative comparisons. To maximize the sample

size for the age correction, we kept both men and women in the reference group. Demographic information, after outlier removal (see Online Resource 1), is provided in Table 1.

Table 1

Demographic characteristics for the subsample under study and reference data

	Breast cancer patients treated with chemotherapy	Reference data
N	62	228
Age (mean, sd)	47.9, 9.3	48.3, 12.5
Percentage women	100	64
% education level[a] (low/ med/high)	0/35/65	0/30/70
% treated with chemotherapy	100	-
Mean years since end chemotherapy	3.3	-
% treated with endocrine therapy	74	-
% treated with immunotherapy	18	-

NOTE. [a]Low education = 8th Grade or less to GED; Medium = High school graduate to Associate degree / some college, and High = Bachelor's degree to Doctorate.

Clustering variables

Cognitive function was assessed with the Amsterdam Cognition Scan (ACS).[14] This test battery contains computer analogues of traditional paper-and-pencil tests that can be subdivided into four domains: verbal memory, attention, speed, and executive functioning (Online Resource 2). The reliability and the validity of the ACS have previously been shown to be sufficient to good.[15]

Analysis methods

Age corrections

For the present study, a state-of-the-art method for age corrections was used,[16] which combines the advantages of traditional tabular approaches — accommodating non-linearity and heteroscedasticity— with the advantages of regression-based approaches —continuity and better use of data. This method allows for age-dependent non-linear differences not only in the expected value, but also in the variance, skewness and kurtosis of scores. We used this method to transform raw scores into age-specific percentile scores, which we in turn converted into standard-normal scores.

Subgrouping methods

The main results were obtained using latent profile analysis. This analysis assumes that the data come from one or more multivariate normal distributions, and estimates the optimal number of these distributions, as well as summary statistics such as the mean and covariance matrices. To select the optimal number of subgroups, we used the Bayesian Information Criterion (BIC).[17]

To check whether results are dependent on the statistical method, we also performed k-means clustering and hierarchical clustering as machine learning methods. To select the optimal number of clusters for both k-means and hierarchical clustering, we used the Gap statistic as an index because it is able to discern whether a one-subgroup solution is optimal.[18]

Normative comparisons

Four criteria for traditional normative comparison were employed. The first —which we will refer to as "ICCTF"— is in line with the International Cognition and Cancer Task Force criteria,[13] where survivors are deemed

cognitively impaired when they have ≥1 test score of ≥2 standard deviation (SD) below the normative mean, or ≥2 test scores of ≥1.5 SD below the normative mean. The second and third criteria require either the 2 standard deviation criterion ("1x2SD"), or the two times 1.5 standard deviations criterion ("2x1.5SD"). The last criterion we used —"2x1SD, different domains"— requires ≥2 test scores of ≥1SD below the normative mean, with the additional requirement that the two test scores are from different cognitive domains. This last method has been used in defining affected cognition in some recent studies,[19-20] and was previously adapted to "4x1SD, different domains" for the case of 20 tests.[21]

Additional methods

As an index of convergence in classification, we used the Adjusted Rand Index (ARI),[22] with ARI≥0.65 indicating moderate recovery. Posterior probability of assignment and silhouette width were used as default measures of distinctness of subgroups.[23-24] To estimate the stability of the results given the sample size, data were simulated from the fitted model and the model refitted to these simulated data. Posterior predictive checks were performed to evaluate how well the model described the data in absolute terms; they involved comparing summary statistics of observed data to summary statistics of model-simulated data.[25]

For all analyses, the R statistical package was used, and specifically the mclust,[26] NbClust[27] and gamlss[16] packages.

Results

Outlier removal and age correction

Five survivors were removed based on outlying scores: Four had zero words correct on Word List Learning Delayed Recall or Box Tapping; one had zero correct on both. No survivors had missing values. Nineteen reference participants were removed based on outlying scores, one had missing values. The age corrections are depicted graphically in Online Resource 1 for all test variables.

Two subgroups: one cognitively affected, one cognitively unaffected

Latent profile analysis showed that the data is best described by two subgroups, one of 34 participants (55%), one of 28 survivors (45%). We refer to the first subgroup as "Cognitively unaffected". Mean z-scores in this subgroup ranged from 0.00 (on the Connect the Dots I) to 0.67 (on Word List Learning Delayed Recall). In this subgroup, there is a tendency towards better scores than expected on the two Word List Learning outcomes. On all other tests the scores in the "Cognitively unaffected" subgroup are as would be expected in a group of no-cancer controls.

We refer to the second subgroup as "Cognitively affected". Mean z-scores in this subgroup ranged from -0.33 (Box Tapping) and -0.48 (Fill the Grid), to -0.92 (Word List Learning Total Recall), -0.98 (Digit Sequences I), and -1.17 (Connect the Dots II). The distribution of raw scores and model-based distributions are provided in Figure 2.

A spherical model was selected, indicating that after separating the survivors into two subgroups, test scores were uncorrelated.[28] Demographic characteristics of the two subgroups are provided in Table 2.

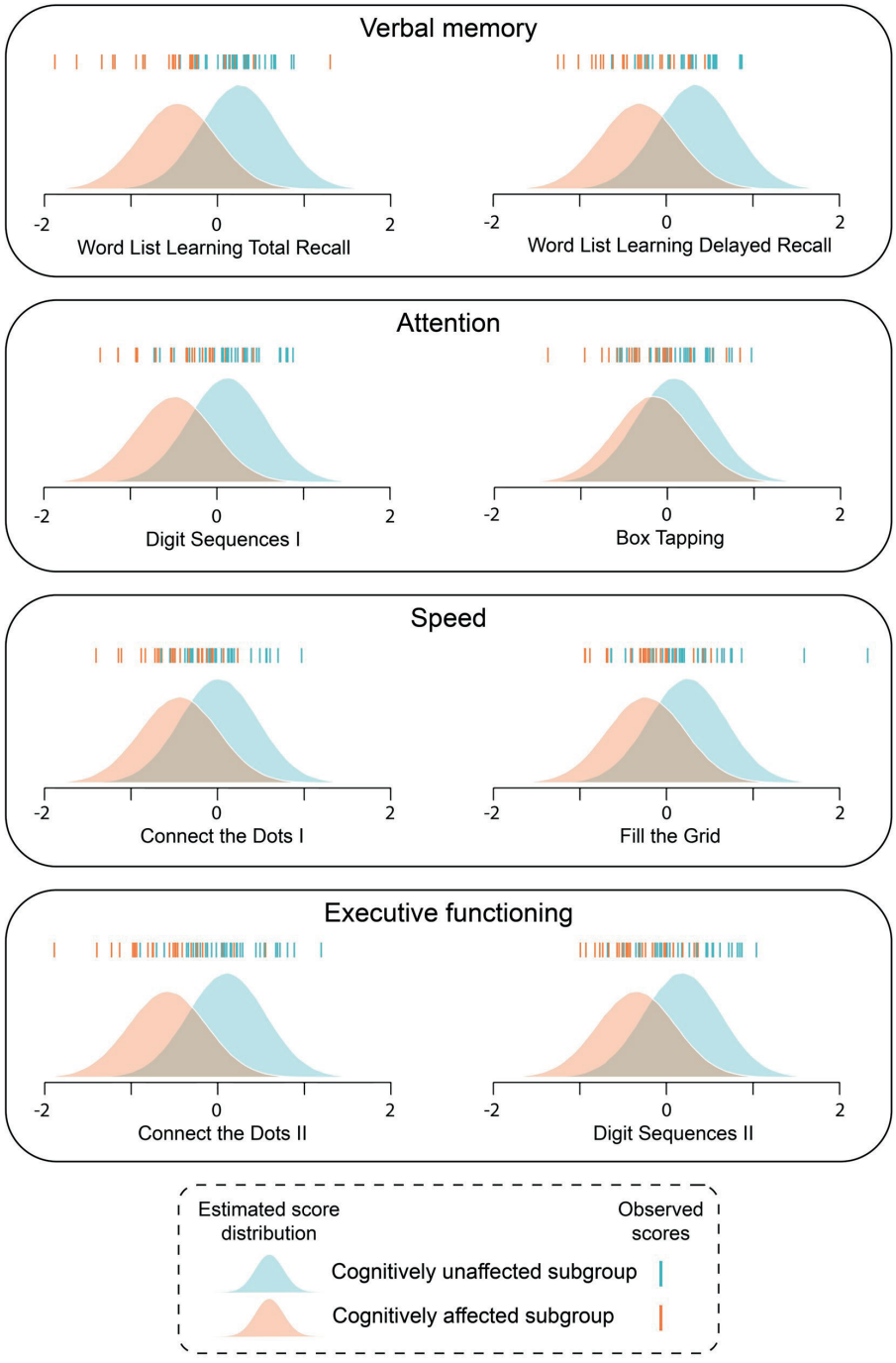


Figure 2. Modeled distribution of test scores in the two subgroups
The small bars at the top represent the observed unmodeled data.

Table 2

Demographic characteristics for the two identified subgroups

	Subgroup 1 Cognitively unaffected	Subgroup 2 Cognitively affected
N	34	28
Age (mean, sd)	49.3, 9.8	46.1, 8.5
Percentage women	100	100
% education level[a] (low/med/high)	0/29/71	0/43/57
% treated with chemotherapy	100	100
Mean years since end chemotherapy	3.1	3.6
% treated with endocrine therapy	78	64
% treated with immunotherapy	12	25

NOTE. [a]Low education = 8th Grade or less to GED; Medium = High school graduate to Associate degree / some college, and High = Bachelor's degree to Doctorate.

Subgroups replicated across subgrouping methods

k-means clustering also selected two subgroups, with group membership being equivalent to the latent profile analysis results for 61/62 participants (98%, ARI=0.94). Hierarchical clustering likewise selected two subgroups, with group membership being equivalent to the latent profile analysis results for 56/62 participants (90%, ARI=0.64). This convergence suggests that the subgrouping solution is relatively independent of the type of subgrouping method used.

Subgroups describe observed data pattern well

The graphical results of posterior predictive checks are provided in Online Resource 3. The observed maximum score was described well for all tests but two. For Fill the Grid, the observed maximum was higher than expected based on the model ($p<0.001$), indicating that a participant was faster than expected (a 67-year-old completing the test in just 49 seconds). For Word List Learning Delayed Recall, the observed maximum was lower than expected ($p=0.048$), indicating that the model expected the best-performer to do even better.

The minimum score was described well for all tests. Only for Word List Learning Total Recall, the observed minimum was marginally lower than expected ($p=0.050$), indicating that the model expected the worst performer to do better (a 45-year-old remembering 25 words out of 75). Means, medians and standard deviations were congruent with model-implied means, medians, and standard deviations for all tests. In general, the data was well approximated by the latent profile model.

Subgroups are recoverable, but show overlap

949 out of 1000 simulations recovered the same "two-subgroup spherical, equal volume" subgrouping result, suggesting that the sample size was

sufficient to arrive at a stable solution. The silhouette method and posterior uncertainty showed there was overlap between the subgroups. The average silhouette width was 0.24, just below the cutoff of 0.25 that suggests separable subgroups.[29] There were four survivors with uncertainty over 0.4, three assigned to the "cognitively affected" subgroup, and one to the "cognitively unaffected" subgroup. This indicates that for these participants, assignment to one of the two subgroups was at chance level. Without these four survivors, silhouette width was 0.27, and mean uncertainty of subgroup assignment was 3%.

Subgroups not different in age or education

The two subgroups did not differ in age, t(60)=1.355, p=0.181 or level of education t(60)=1.749, p=0.085, although this may also be due to a lack of statistical power; a Bayesian re-analysis[30] revealed there is only anecdotal evidence for the absence of an age difference (BF_{01}=1.783) and no evidence for an education difference or a lack thereof (BF_{01}=1.076).

Subgroups best approached by "2x1SD, different domains" criterion

When using normative comparisons to examine whether survivors are cognitively affected, the "ICCTF" criterion resulted in the same assignment as the subgrouping solution for 52/62 survivors (84%, ARI=0.45); the "1x2SD" criterion for 49/62 survivors (79%, ARI=0.33). The 13 survivors for whom assignment differed were not cognitively affected according to the "1x2SD" criterion, but were affected according to the latent profile analysis. For the "2x1.5SD" criterion, the result was exactly the same as for the "1x2SD" criterion, although the 13 mismatched survivors were different. The "2x1SD, different domains" criterion resulted in the same assignment as the subgrouping solution for 56/62 survivors (90%, ARI=0.64) and thus provided the best proxy for a subgrouping approach in the current dataset.

Discussion

Our objective was to discover whether there are subgroups of breast cancer survivors that are cognitively affected or unaffected, or in contrast, whether there is only continuous variation in cognitive functioning between survivors. We found that a subgroup —almost half— of breast cancer survivors is cognitively affected after chemotherapy; the other half is comparable to no-cancer controls.

The percentage of cognitively affected survivors in this sample —45%— is higher than the 24% that is typically found.[3] However, previous estimates were established using a top-down traditional normative comparisons approach, and were thus dependent on the definition of 'cognitively affected'. The primary benefit of our bottom-up data-driven approach is that we are not recovering what we put in.

Of the normative comparison approaches, the "2x1SD, different domains" criterion resulted in the optimal split between the two recovered subgroups. This supports the idea that the cognitive impact observed in breast cancer survivors after chemotherapy is best described as diffuse, and that deficits are relatively subtle.[1] This differentiates this particular type of cognitive effect from that observed in for example stroke, where large domain-specific effects can be observed.[31]

On two verbal memory outcomes, the "cognitively unaffected" subgroup performed better than would be expected based on the reference sample. This is most likely due to the inclusion of men in the reference sample, with no men in the survivor sample. Sex differences on neuropsychological tests are generally smaller than age differences,[32] but women outperform men on verbal learning tests.[33] As mentioned in the methods section, we kept both men and women in the reference group to maximize the sample size for the age correction, which may have introduced this effect.

The sample size of the present study was modest to establish a prevalence estimate; power calculations are not commonly performed for subgrouping methods.[34] Had the sample been larger, the delineation between affected and unaffected could have been more precise, and indeterminate subgroup assignment for four survivors might have been avoided. However, it could also be that there are simply survivors for whom cognitive status is not clear-cut. Results for these survivors are appropriately identified by our subgrouping method as being indeterminate, while a cutoff-based method would inappropriately force them into one of the two categories. Previous work from our group has also shown that forcing participants into either one category or the other with a binary classification strategy leads to unreliable classification.[35] This underscores that the use of a method that models uncertainty —such as latent profile analysis— is preferable over normative comparison methods no matter what cutoff is selected, although the analysis is more cumbersome.

Differences between the cognitively affected and unaffected subgroups varied across cognitive domains. On two tests, differences were small: Box Tapping, a spatial memory test, and Fill the Grid, a motor dexterity test. This is congruent with previous findings that suggest no difference in motor speed or visual memory between those treated with chemotherapy and those that are not.[36] Larger effects were found on Word List Learning —measuring verbal memory— and Connect the Dots —an executive functioning test that measures a variety of attentional processes;[37] variants of which are already included in the recommendations of the International Cognition and Cancer Task Force.[13] Digit Sequences I and II —which tap into executive functioning and attention— also showed larger differences between subgroups, and may be a good addition to standard test batteries for cancer-related cognitive impairment.

In conclusion, we provided support for the existence of two subgroups of breast cancer survivors after chemotherapy: one cognitively affected and one unaffected. We approached the problem of subgroups from

a new angle that does not rely on consensus-based cutoffs. In distinguishing the two subgroups, cutoff criteria that require small differences on multiple cognitive domains performed best. With this bottom-up method, the percentage of cognitively affected survivors was observed to be higher than previously thought.

References

1. Wefel JS, Kesler SR, Noll KR, Schagen SB. Clinical characteristics, pathophysiology, and management of noncentral nervous system cancer-related cognitive impairment in adults. *CA Cancer J Clin*. 2015;65(2):123-138.
2. Bernstein LJ, McCreath GA, Komeylian Z, Rich JB. Cognitive impairment in breast cancer survivors treated with chemotherapy depends on control group type and cognitive domains assessed: a multilevel meta-analysis. *Neurosci Biobehav Rev*. 2017;83:417-428.
3. Dijkshoorn ABC, van Stralen HE, Sloots M, Schagen SB, Visser-Meily JMA, Schepers VPM. Prevalence of cognitive impairment and change in patients with breast cancer: a systematic review of longitudinal studies. *Psycho-Oncol*. 2021;30(5):635-648.
4. Ahles TA, Root JC, Ryan EL. Cancer-and cancer treatment–associated cognitive change: an update on the state of the science. *J Clin Oncol*. 2012;30(30):3675.
5. Henneghan A. Modifiable factors and cognitive dysfunction in breast cancer survivors: a mixed-method systematic review. *Support Care Cancer*. 2016;24(1):481-497.
6. Borsboom D, Rhemtulla M, Cramer AOJ, van der Maas HLJ, Scheffer M, Dolan CV. Kinds versus continua: a review of psychometric approaches to uncover the structure of psychiatric constructs. *Psychol Med*. 2016;46(8):1567-1579.
7. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*. 2003;100(14):8418-8423.
8. Northcott PA, Korshunov A, Witt H, et al. Medulloblastoma comprises four distinct molecular variants. *J Clin Oncol*. 2011;29(11):1408.
9. Karlson CW, Sarver DE, Raiker JS, et al. The contribution of neurocognitive functions to academic and psychological outcomes in pediatric cancer: a latent profile analysis. *Child Neuropsychol*. 2020;26(7):881-899.

10. Partanen M, Phipps S, Russell K, et al. Longitudinal trajectories of neurocognitive functioning in childhood acute lymphoblastic leukemia. *J Pediatr Psychol*. 2021;46(2):168-178.
11. Sharkey CM, Mullins LL, Clawson AH, et al. Assessing neuropsychological phenotypes of pediatric brain tumor survivors. *Psycho-Oncol*. 2021;30(8):1366-1374.
12. Bender CM, Merriman JD, Sereika SM, et al. Trajectories of cognitive function and associated phenotypic and genotypic factors in breast cancer. *Oncol Nurs Forum*. 2018;45(3):308-326.
13. Wefel JS, Vardy J, Ahles T, Schagen SB. International cognition and cancer task force recommendations to harmonise studies of cognitive function in patients with cancer. *Lancet Oncol*. 2011;12(7):703-708.
14. Feenstra HEM, Vermeulen, IE, Murre, JMJ, Schagen SB. Online self-administered cognitive testing using the Amsterdam Cognition Scan: establishing psychometric properties and normative data. *J Med Internet Res*. 2018;20(5):e9298.
15. Feenstra HEM, Murre JMJ, Vermeulen IE, Kieffer JM, Schagen SB. Reliability and validity of a self-administered tool for online neuropsychological testing: the Amsterdam Cognition Scan. *J Clin Exp Neuropsychol*. 2018;40(3):253-273.
16. Stasinopoulos DM, Rigby RA. Generalized additive models for location scale and shape (GAMLSS) in R. *J Stat Softw*. 2008;23:1-46.
17. Raftery AE. Bayesian model selection in social research. *Sociol Methodol*. 1995;25:111-163.
18. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *J R Stat Soc Series B Stat Methodol*. 2001;63(2):411-423.
19. Witlox L, Schagen SB, de Ruiter MB, et al. Effect of physical exercise on cognitive function and brain measures after chemotherapy in patients with breast cancer (PAM study): protocol of a randomised controlled trial. *BMJ Open*. 2019;9(6):e028117.
20. Klaver KM, Duijts SF, Geusgens CA, et al. Internet-based cognitive

rehabilitation for WORking Cancer survivors (i-WORC): study protocol of a randomized controlled trial. *Trials*. 2020;21(1):1-12.
21. Gehring K, Sitskoorn MM, Gundy CM, et al. Cognitive rehabilitation in patients with gliomas: A randomized, controlled trial. *J Clin Oncol*. 2009;27(22):3712-3722.
22. Hubert L, Arabie P. Comparing partitions. *J Classif*. 1985;2(1):193-218.
23. Nagin DS. Analyzing developmental trajectories: a semiparametric, group-based approach. *Psychol Methods*. 1999;4(2):139.
24. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math*. 1987;20:53-65.
25. Rubin DB. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann Stat*. 1984;12(4):1151-1172.
26. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J*. 2016;8(1):289.
27. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: an R package for determining the relevant number of clusters in a data set. *J Stat Softw*. 2014;61:1-36.
28. Vermunt JK, Magidson J. Latent class cluster analysis. In: Hagenaars JA, McCutcheon AL, eds. *Applied Latent Class Analysis*. Cambridge, MA: Cambridge University Press; 2002:89-106.
29. Lletí R, Ortiz MC, Sarabia LA, Sánchez MS. Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Anal Chim Acta*. 2004;515(1):87-100.
30. Keysers C, Gazzola V, Wagenmakers EJ. Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nat Neurosci*. 2020;23(7):788-799.
31. Demeyere N, Riddoch MJ, Slavkova ED, et al. Domain-specific versus generalized cognitive screening in acute stroke. *J Neurol*. 2016;263(2):306-315.
32. Sherrill-Pattison S, Donders J, Thompson E. Influence of demographic

variables on neuropsychological test performance after traumatic brain injury. *Clin Neuropsychol*. 2000;14(4):496-503.

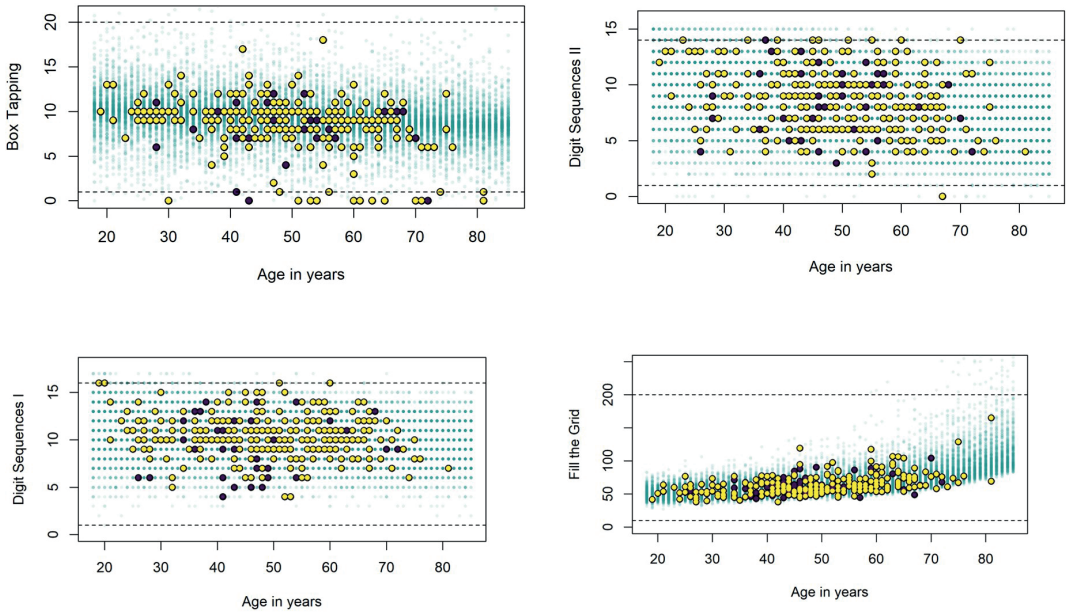
33. Gale SD, Baxter L, Connor DJ, Herring A, Comer J. Sex differences on the Rey auditory verbal learning test and the brief visuospatial memory test–revised in the elderly: normative data in 172 participants. *J Clin Exp Neuropsychol*. 2007;29(5):561-567.
34. Tein JY, Coxe S, Cham H. Statistical power to detect the correct number of classes in latent profile analysis. *Struct Equ Modeling*. 2013;20(4):640-657.
35. Luijendijk MJ, Feenstra HEM, Vermeulen, IE, Murre JMJ, Schagen SB. Binary classification threatens the validity of cognitive impairment detection [published online ahead of print]. *Neuropsychol*. 2022.
36. Jim HS, Phillips KM, Chait S, et al. Meta-analysis of cognitive functioning in breast cancer survivors previously treated with standard-dose chemotherapy. *J Clin Oncol*. 2012;30(29):3578.
37. Agelink van Rentergem JA, Vermeulen IE, Lee Meeuw Kjoe PR, Schagen SB. Computational modeling of neuropsychological test performance to disentangle impaired cognitive processes in cancer patients. *J Natl Cancer Inst*. 2021;113(1):99-102.

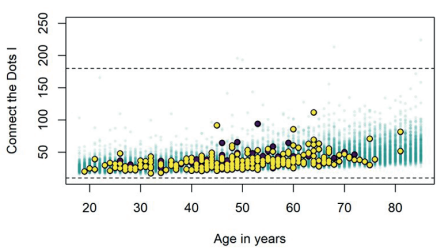
Supplementary material

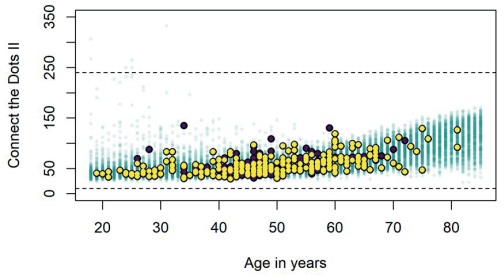
Online Resource 1

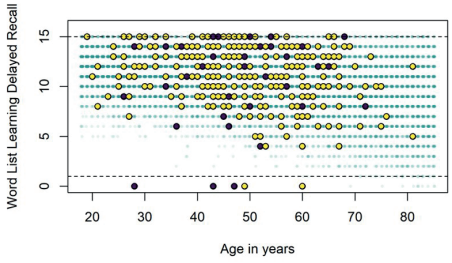
In blue, the estimated distribution of normative scores are displayed, based on GAMLSS estimation. The yellow dots represent the values of the reference participants of the ACS dataset (both men and women). The estimation of the normative distribution using GAMLSS was based on the yellow dots that do not fall outside of the range defined by the dashed lines. The purple dots represent the values of breast cancer patients.

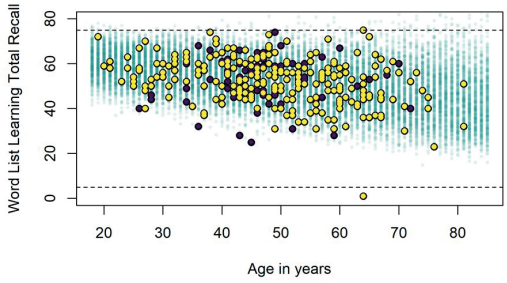
The dashed lines are the outlier cutoffs that were used to remove values. The outlier cutoff criteria were extreme, requiring the bare minimum of performance (or potentially removing impossibly good scores), and are given below the graphs. As mentioned in the main text, five participants were removed, based on a 0 score on Word List Learning Delayed Recall or Box Tapping (possibly due to computer error or misunderstanding).











Variable	Lower outlier removal criterion	Higher outlier removal criterion
Word List Learning – Total Recall	Fewer than 5 words remembered out of 75 (= fewer than 1 word remembered in every trial)	More than 75 words remembered out of 75 remembered
Word List Learning – Delayed Recall	Fewer than 1 word remembered out of 15	More than 15 words remembered out of 15
Digit Sequences I	Fewer than 1 point	More than 16 points
Box Tapping	Fewer than 1 point	More than 20 points
Connect the Dots I	Fewer than 10 seconds taken to complete the sequence of 25 clicks	More than 180 seconds taken to complete the sequence of 25 clicks
Fill the Grid	Fewer than 10 seconds taken to drag and drop 25 squares	More than 300 seconds taken to drag and drop 25 squares
Connect the Dots II	Fewer than 10 seconds taken to complete the sequence of 25 clicks	More than 240 seconds taken to complete the sequence of 25 clicks
Digit Sequences II	Fewer than 1 point	More than 14 points

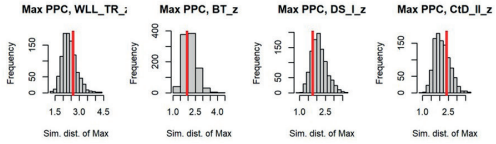
Online resource 2

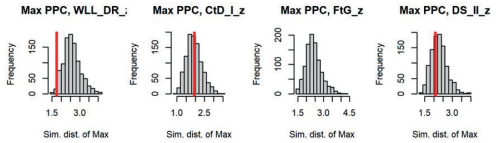
Cognitive tests and outcomes

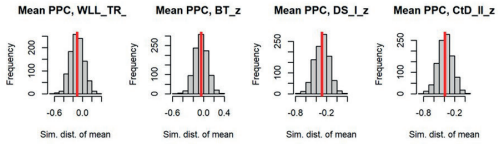
Cognitive Domains	Cognitive tests	Traditional equivalents	Main outcome measure
Verbal memory	Word List Learning – Total Recall	Verbal Learning Test – Total Recall	Total number of correct words (trial 1 to 5)
	Word List Learning – Delayed Recall	Verbal Learning Test – Delayed recall	Total number of correct words
Attention	Digit Sequences I	Digit Span Forwards	Total number of correctly repeated sequences
	Box Tapping	Corsi Block Tapping	Total number of correctly repeated sequences
Speed	Connect the Dots I	Trail Making Test A	Completion time in seconds
	Fill the Grid	Grooved Pegboard	Completion time in seconds
Executive functioning	Connect the Dots II	Trail Making Test B	Completion time in seconds
	Digit Sequences II	Digit Span Backwards	Total number of correctly repeated sequences

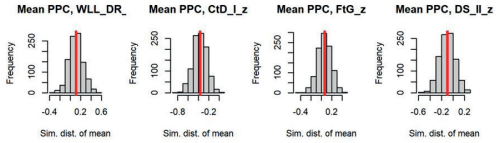
Online Resource 3

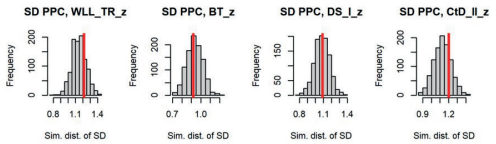
Posterior Predictive Checks for each of the eight test outcomes, for the following summary statistics: Maximum, Minimum, Mean, Median, and Standard Deviation. The histograms show the distribution of model-implied summary statistics, derived from 1000 simulations from the latent profile analysis. The red lines indicate the summary statistics that were observed in the real data. If the red line falls within the histogram, this indicates that this summary statistic is well described by the model. If the red line falls outside the histogram, this indicates some violation of the absolute fit of the model.

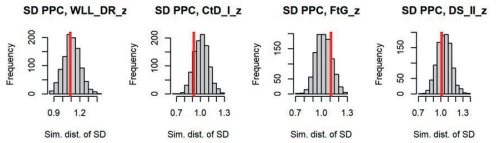

Max PPC, WLL_TR_z
Max PPC, BT_z
Max PPC, DS_I_z
Max PPC, CtD_II_z
Frequency
Sim. dist. of Max

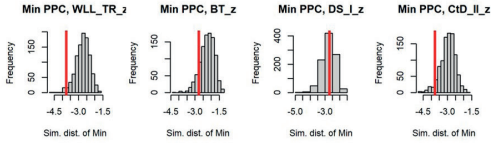

Max PPC, WLL_DR_z
Max PPC, CtD_I_z
Max PPC, FtG_z
Max PPC, DS_II_z
Frequency
Sim. dist. of Max

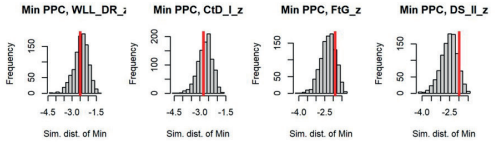

Mean PPC, WLL_TR_
Mean PPC, BT_z
Mean PPC, DS_I_z
Mean PPC, CtD_II_z
Frequency
Sim. dist. of mean

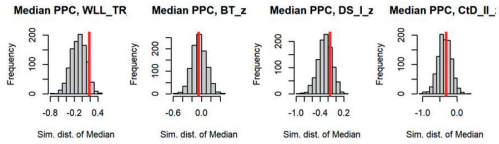

Mean PPC, WLL_DR_
Mean PPC, CtD_I_z
Mean PPC, FtG_z
Mean PPC, DS_II_z
Frequency
Sim. dist. of mean

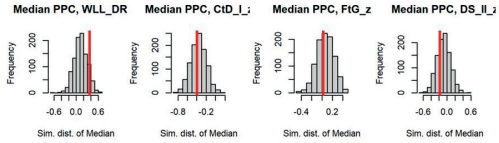

SD PPC, WLL_TR_z
SD PPC, BT_z
SD PPC, DS_I_z
SD PPC, CtD_II_z
Frequency
Sim. dist. of SD


SD PPC, WLL_DR_z
SD PPC, CtD_I_z
SD PPC, FtG_z
SD PPC, DS_II_z
Frequency
Sim. dist. of SD


Min PPC, WLL_TR_z
Min PPC, BT_z
Min PPC, DS_I_z
Min PPC, CtD_II_z
Frequency
Sim. dist. of Min


Min PPC, WLL_DR_z
Min PPC, CtD_I_z
Min PPC, FtG_z
Min PPC, DS_II_z
Frequency
Sim. dist. of Min


Median PPC, WLL_TR
Median PPC, BT_z
Median PPC, DS_I_z
Median PPC, CtD_II_z
Frequency
Sim. dist. of Median


Median PPC, WLL_DR
Median PPC, CtD_I_z
Median PPC, FtG_z
Median PPC, DS_II_z
Frequency
Sim. dist. of Median

Part II:
Implementation of cognitive tests in studies on cognitive effects of endocrine therapy

Chapter 5:
Endocrine therapy with or without CDK4/6 inhibitors in women with hormone-receptor positive breast cancer: What do we know about the effects on cognition?

Philippe R. Lee Meeuw Kjoe, Elsken van der Wall, Sanne B. Schagen.

Clinical Breast Cancer. 2022;22(3):191-199.

And adapted and translated to:
Lee Meeuw Kjoe PR, van der Wall E, Schagen SB. De mogelijke invloed van endocriene therapie en CDK4/6-remmers op cognitie bij het mammacarcinoom. *Nederlands Tijdschrift Voor Oncologie*. 2021;18(2):42-50.

Abstract

Adjuvant endocrine therapy (ET) is the cornerstone of treatment for hormone-receptor positive breast cancer. Recently, ET is increasingly combined with 'cyclin-dependent kinases 4 and 6' (CDK4/6) inhibitors. Given the importance of estrogens in neural processes and the role of cyclin D in hippocampal cell proliferation, it is plausible that these therapies affect cognition, but studies on these potential cognitive effects are sparse. In this review, we summarize existing knowledge on the cognitive effects of ET and CDK4/6 inhibitors in pre-, peri- and postmenopausal patients with breast cancer. We show that several clinical studies support adverse cognitive effects, especially on verbal memory, after ET-induced decrease of estrogen-levels or inactivation of estrogen-receptors. Clinical studies on the cognitive effects of CDK4/6 inhibitors are virtually non-existent and no conclusions can yet be drawn. Longitudinal studies on the cognitive effects of the combined ET-CDK4/6 inhibitors are highly needed to properly inform patients about potential short-term and long-term cognitive side effects. These studies should preferably include cognitive assessments (including a measurement prior to ET), and be designed in such a way that they can account for variables such as type and duration of ET, CDK4/6 inhibition, menopausal status, and other disease- and treatment-related symptoms that can impact cognition, such as fatigue and distress.

Introduction

Approximately 1 in every 8 U.S. women will be diagnosed with invasive breast cancer (BC) in their lifetime: In 2021, this equals to more than 281,000 U.S. women who are expected to develop BC.[1] Of these women, approximately 75% will be diagnosed with hormone-receptor positive (HR+) BC.[2] To reduce the risk of disease recurrence, women with primary HR+ BC are recommended to undergo (neo-)adjuvant systemic therapy such as (among others) endocrine therapy (ET). The preferred duration of ET is five years with, in case of a high-risk tumor, extension to seven to ten years. The duration of ET in women with metastatic BC might also be extended due to increasing promising combinations of new targeted therapies with ET. At present, these therapy strategies consist of the combination of common endocrine agents (e.g., selective estrogen receptor (SER) modulators such as tamoxifen, SER downregulators such as fulvestrant, steroidal and non-steroidal aromatase inhibitors, with the 'mammalian target of rapamycin (mTOR)' inhibitors, the 'cyclin-dependent kinases 4 and 6' (CDK4/6) inhibitors and the 'phosphoinositide 3-kinase (PI3K)' inhibitors. These targeted agents increase the number of ET options that hopefully results in longer duration of response. With this, knowledge of short-term and long-term side effects of these therapies become highly relevant. In this article, we focus on cognitive decline as a possible side effect.

Cognitive decline is frequently reported by BC patients during and after chemotherapy: More than half of BC patients report cognitive changes, primarily in memory and attention, and longitudinal studies found that up to 61% of BC patients performed worse on standardized neuropsychological tests from pre- to one year post-chemotherapy.[3] Risk factors for developing chemotherapy-related cognitive decline are diverse: Older age, genetic polymorphisms such as of the apolipoprotein E, and the type of chemotherapy (anthracycline-containing combinations) are all associated with an increased risk for cognitive decline.[4-5] In addition, the

severity of cognitive decline appears to be dose- and duration-dependent, also in other tumor types.[6-8] These risk factors, however, do not provide sufficient information for accurate risk prediction. Currently, the most important mechanisms of chemotherapy-related cognitive decline are thought to be dysregulation of neural circuits, downregulation of brain-derived neurotrophic factor (BDNF), increased neuroinflammation, and oxidative stress.[9] Neuroimaging studies have shown decreased gray matter density, changes in brain activation and cerebral circulation, decreased white matter integrity, and altered structural connectivity and morphological changes[10] in patients from pre to post chemotherapy, and even 20 years after treatment, differences are observed in brain structure between women who have undergone chemotherapy for BC and women without a cancer history.[11]

In contrast to studies on chemotherapy-related cognitive decline, studies on possible cognitive effects of ET in BC patients are sparse, both in the adjuvant and in the metastatic setting. Studies on the possible cognitive effects of the combined ET are virtually non-existent. Because of the important role of estrogens in maintaining functional integrity of the brain, an update of the literature regarding the cognitive effects of ET in both pre-, peri- and postmenopausal BC patients is opportune. The extended duration of ET (in the adjuvant setting already up to 10 years), and the rapidly increasing use of the combination of ET with CDK4/6 inhibitors make the study on the cognitive impact of these therapies even more relevant. Increasing awareness of harmful cognitive effects of these therapies is essential for optimal care for BC patients, and is also relevant for other (oncological) patients undergoing endocrine therapies. Such awareness contributes to better communications between health care providers and patients, adds to informed-decision making and spurs investigations of ways to intervene against cancer (treatment)-related cognitive impairment.

Methods

We collected literature through Pubmed and Google Scholar using the following search terms: (cancer OR chemotherapy OR estrogen OR tamoxifen OR aromatase inhibitor OR endocrine therapy OR CDK4/6 inhibitor OR palbociclib OR ribociclib OR abemaciclib) AND (mechanism OR cognition OR cognitive functioning OR cognitive impairment OR fatigue OR quality of life).

What is cognition?

Cognition is an umbrella term for processes related to thinking, such as concentration, memory, and executive functions (EF; planning, keeping an overview, and directing behavior).[12] These cognitive functions involve specific brain areas and neural circuits. Cognitive functions are further classified into several processes: For example, in memory, neuropsychologists make a distinction between short-term memory, working memory and long-term memory.

The clinical assessment of cognitive function includes the evaluation of both self-perceived cognitive complaints and cognitive function with formal neuropsychological tests. Cognitive complaints refer to subjective reports from a patient and/or a loved one of his/her cognitive ability in daily life, which are often measured using questionnaires, such as the Functional Assessment of Cancer Therapy – Cognitive Function (FACT-Cog).[e.g.,13] Research has shown that subjective cognitive measures not only reflect cognitive abilities, but also include for example, expectations, coping style, and the influence of fatigue and mood. Cognitive functioning assessed by standardized neuropsychological tests is less influenced by other factors. These tests traditionally consist of paper-and-pencil tests administered face-to-face by a neuropsychologist. An example of such a neuropsychological test is a verbal memory test, whereby a patient is asked to learn, remember

and recognize a list of words. Another example is an executive functioning test whereby mental flexibility is measured by asking patients to connect numbers and letters in numeric, alphabetic and alternating order.

Self-perceived cognitive complaints and tested cognitive functioning are often not strongly correlated, and in clinical practice, the assessment of both is critical to a differential diagnosis procedure and to the intervention plan that arises from this process. In research, the specific clinical research question will help guide the choice of objective performance-based neuropsychological tests of cognitive function and/or subjective inventories of cognitive complaints. Research aspiring to measure the effect of a treatment on cognitive function is best served by objective performance-based tests of cognitive function, a view espoused and shared by regulatory authorities.[14]

Hormone-receptor positive breast cancer

BC is regarded as HR+ when histopathological examination shows the presence of hormone receptors in ≥1% (USA) or >9% (the Netherlands) of tumor cells. In HR+ BC, estrogens play the most important role in the development and growth of the cancer cells by binding to the hormone receptors of the cancer cells (see Figure 1). Estrogens influence cell functioning by binding to the so-called estrogen receptors (ER) located in the nucleus or on the cell wall. As a result of binding, dimerization and phosphorylation of the ER occur allowing binding to deoxyribonucleic acid (DNA) of the target genes of ER. This binding to the so-called Estrogen Response Elements (EREs) leads to activation of these target genes which then induce growth, differentiation, apoptosis and angiogenesis. In addition, estrogens can also activate the target genes independently of binding to the EREs through protein-protein interactions, via for example the tyrosine kinase pathways (e.g., EGFR, HER2, IGFR).[15] Phosphorylation of ER also occurs by growth factors that increase the activation of protein kinases resulting in gene transcription.

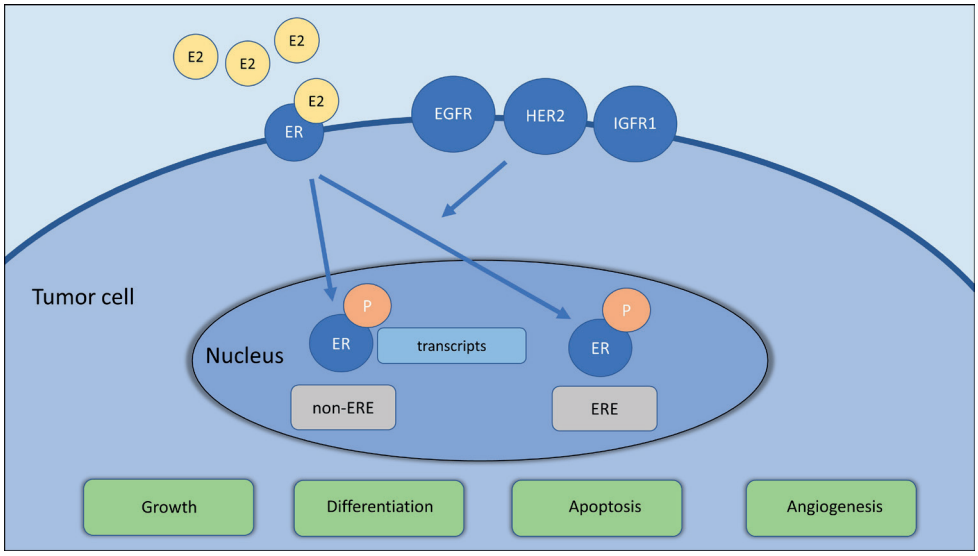


Figure 1. Interaction between estrogen and growth factor receptors in tumor cells

E2=estrogen, EGFR=epidermal growth factor receptor, ER=estrogen receptor, ERE=estrogen receptor element, HER2=growth factor receptor, HRG=heregulin, IGFR1=insulin-like growth factor receptor, P=phosporylation.

Adapted from reference 14.

Endocrine therapy

Multiple types of ET exist (see Figure 2). The choice of endocrine treatment is primarily determined by the patient's menopausal status, stage of disease, and any previous treatment with ET. In clinical practice, aromatase inhibitors (anastrozole, exemestane, and letrozole), tamoxifen, fulvestrant, and the luteinizing-hormone-releasing-hormone (LHRH) analogs are the most common types of ET.[16] With the exception of fulvestrant, all medications are prescribed in the (neo-)adjuvant and in metastatic setting.

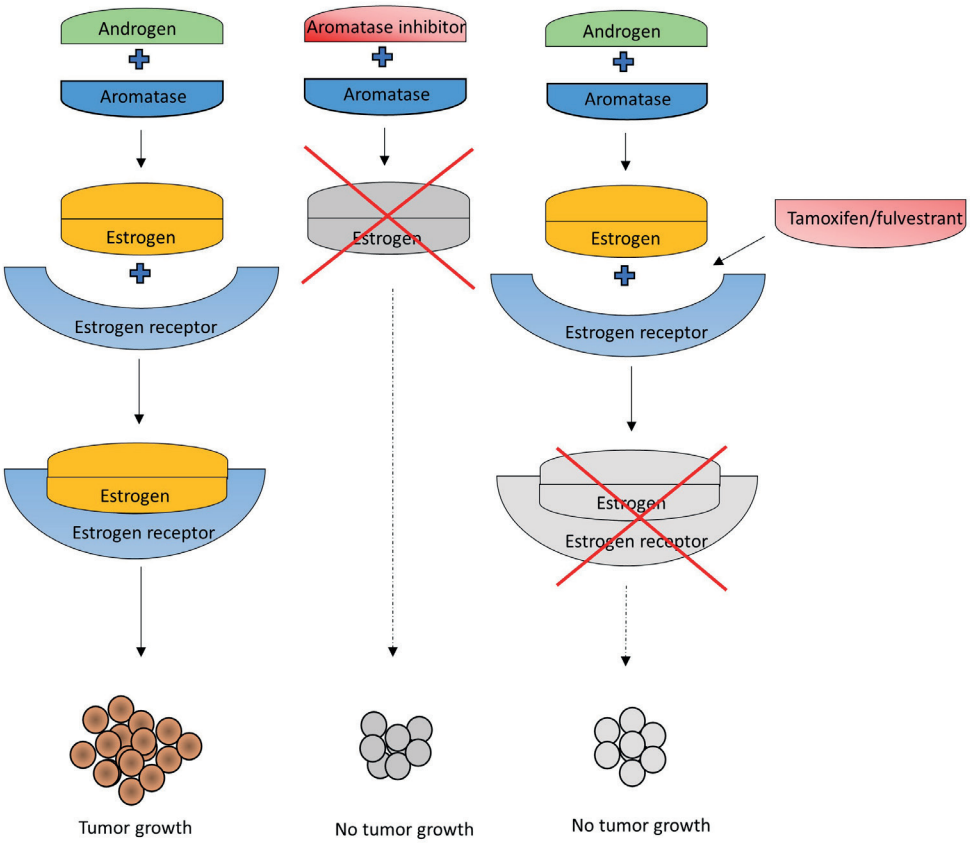


Figure 2. Mechanisms of aromatase inhibitors, tamoxifen, and fulvestrant in BC
Adapted from reference 17.

Tamoxifen is the most commonly used agent of all SERMs, drugs that have both estrogenic and anti-estrogenic effects, and is the longest known ET in the treatment of BC. Tamoxifen inhibits or stimulates the action of estrogens by binding to the ER and either blocking or stimulating ER, depending on the tissue. For example, tamoxifen has a stimulatory effect on endometrial and bone tissue, but an inhibitory effect on mammary tissue.[17]

Aromatase inhibitors inhibit the production of estrogens by inactivating the enzyme aromatase in peripheral adipose tissue. This

inactivation leads to blockage of the conversion of estrogen precursors, which are still produced in the adrenal glands even after menopause, to estrogens. This leads to a profound hypoestrogenism.[18] A distinction is made between steroidal and non-steroidal aromatase inhibitors. Steroidal inhibitors inactivate the enzyme aromatase irreversibly by forming a covalent bond. Non-steroidal aromatase inhibitors, on the other hand, bind with aromatase reversibly, reducing the activity of the enzyme. Both steroidal (such as exemestane) and non-steroidal aromatase inhibitors (such as anastrozole and letrozole) are prescribed primarily to postmenopausal women. The previously described profoundly hypoestrogenic status leads, with intact ovarian function, to a stimulation of ovarian estrogen production from the pituitary gland. Therefore, when treating premenopausal women with aromatase inhibitors, concurrent ovarian suppression is required by surgical removal of the ovaries or by LHRH analogs such as goserelin, buserelin, and leuproline.[19]

Fulvestrant belongs to the "selective estrogen receptor degrader" (SERDs). Fulvestrant inhibits the action of estrogens by degrading ER in tumor tissue. Fulvestrant binds with the ER causing inhibition of dimerization as well as translocation of the receptor to the nucleus. The receptors are then degraded preventing the intracellular effects of estrogens. The binding to the ER is much stronger than that of tamoxifen and activity has been described in patients resistant to tamoxifen. Fulvestrant is used as a single agent (in patients with phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha (PIK3CA) wild-type tumors who were previously treated with a CDK4/6 inhibitor) or as a combination therapy with a CDK4/6 inhibitor (if not used before) or alpelisib (if activating PIK3CA mutations have been identified in the tumor).[20-21]

In the adjuvant setting, multiple meta-analyses have shown that 5 years of treatment with tamoxifen reduces the risk of disease recurrence by half in the first 5 years, and by one-third in the years 5-10 when tamoxifen has already been discontinued. The risk of mortality is also reduced by one-

third. With extended adjuvant administration of tamoxifen to 10 years, indicated in high-risk women, there is an additional absolute reduction in mortality of almost 3%.[22] Compared with tamoxifen, (partial) replacement of tamoxifen with an aromatase inhibitor reduces the risk of relapsing disease by an additional 30%.[23]

In the first-line treatment of metastatic HR+ BC, monotherapy with aromatase inhibitors results in the prolongation of the progression-free interval by 13-16 months.[24]

How can ET affect cognition?

Based on increasing knowledge about the role of estrogens in cognitive functioning, it is expected that ET, which interacts with the function of estrogens or reduces estrogen levels, may be associated with cognitive decline.

Estrogens play a role in several biological systems that are known to be essential for normal cognitive function. Previous research has shown that estrogens affect cognition in healthy women during the menstrual cycle, menopausal transition, and menopause.[25] Estrogens are known to have a positive influence on neural plasticity and neuroprotection.[26-27] Estrogens stimulate neural plasticity both directly and indirectly. Animal models show that estrogens stimulate the production of dendritic spines (postsynaptic part of a synapse) in the hippocampus and prefrontal cortex.[28-29] In addition, estrogens stimulate neurotrophins such as BDNF which affect dendritic spines in the hippocampus.[30] Estrogens increase antioxidants, reduce free radicals and oxidative damage to mitochondrial DNA and have anti-inflammatory effects.[31-32] These latter processes are important for normal neuronal functioning and if altered, are associated with higher risk of neurodegenerative diseases such as Alzheimer's disease.[33]

Lastly, estrogens interact with neurotransmitters, including the brain's cholinergic and serotonergic systems, which play an important role in cognitive processes and mood.[34-35]

Indirect effects ET

ET could also indirectly lead to cognitive problems due to commonly reported side effects such as fatigue, insomnia, and mood disturbances.[36] These side effects can each separately have their impact on cognitive functioning but also reinforce each other in an adverse effect on cognition.

Differential effect types ET

Both aromatase inhibitors and tamoxifen can cross the blood-brain barrier.[37-39] It seems plausible for aromatase inhibitors and tamoxifen to have a differential effect on cognition given the previously described mechanisms of action of the agents whereby aromatase inhibitors cause a decrease in estrogen concentrations while tamoxifen blocks ER in breast tissue. In addition, aromatase inhibitors have an anti-estrogenic effect on all tissues while tamoxifen, depending on the tissue type, may also have a stimulatory effect on ER. Brain regions also differ in the balance between the two distinctive ER, ERα and ERβ. For example, there are more ERβ than ERα in the hippocampus and temporal cortex and, conversely, more ERα than ERβ in the amygdala and hypothalamus.[40]

Within the group of aromatase inhibitors, it is further useful to explore differences in cognitive effects, since steroidal inhibitors inactivate aromatase irreversibly, while non-steroidal aromatase inhibitors bind reversibly thereby reducing the activity of the enzyme.

What is known about the cognitive effects of endocrine therapy?

Several cross-sectional and prospective neuropsychological studies have examined the impact of ET on cognitive functioning. Although many studies have suboptimal study designs and varying study results, the data seems to suggest that the use of ET in BC patients adversely affects cognitive functioning, both self-reported cognitive complaints and tested impairment.[e.g.,41-42] For example, Boele and colleagues found lower scores on verbal memory and higher frequency of reported cognitive complaints in postmenopausal BC patients using tamoxifen compared to BC patients without systemic therapy and non-cancer controls.[43]

The cognitive domain that is most consistently found to be affected by ET is verbal memory (encoding, storage and retrieval of verbal information such as a shopping list or a newspaper article). A recent meta-analysis suggested, based on nine cross-sectional studies, that BC patients treated with ET (tamoxifen or aromatase inhibitors) scored lower on tests of verbal memory compared to both BC patients not treated with ET and healthy controls (small to medium effect, Hedges' g effect size: -.37).[44] Verbal memory is associated with brain regions with extensive expression of ER.[17]

Several randomized controlled trials (RCT) suggest that tamoxifen use is more often associated with cognitive decline than aromatase inhibitors. In the TEAM trial in which neuropsychological testing was used in postmenopausal BC patients, tamoxifen users scored lower on verbal memory and EF than non-cancer controls.[42] In contrast, exemestane users did not score lower than controls. The IBIS II study in which neuropsychological testing was used in postmenopausal women at high risk of developing BC showed that anastrozole users had similar test scores as control subjects.[46] Also, in the BIG 1-98 study, postmenopausal patients taking letrozole scored better on neuropsychological examination during the fifth year of treatment than those taking tamoxifen.[47]

Menopausal status might play an important role in the influence of ET on cognition, however, data of the influence of menopausal status are scarce.[48] In a side study of the TAILORx study, premenopausal women with BC treated with ET reported more often cognitive decline than postmenopausal women with BC.[13] It is possible that the effects are stronger in premenopausal patients because of the relatively large and sudden alterations in estrogen levels after ET. On the other hand, postmenopausal patients may have more often cognitive decline because higher age is associated with both a greater risk of BC and impaired cognitive functioning.

Methodological limitations of existing studies

Limitations in and differences between studies can lead to differences in cognitive test results independent of actual differences in cognitive functioning. The vast majority of studies on the impact of ET on cognition are cross-sectional in nature. Furthermore, even longitudinal studies often lack measurement before onset of ET, which makes it difficult to determine the origin of cognitive problems.[e.g.,49] In addition, most longitudinal studies have a short follow-up of up to 12 months. As a result, harmful long-term or late effects can be missed, as well as any improvements in cognition. The vast majority of studies has small sample sizes; Therefore, different ET classes are often taken together and/or combined with chemotherapy.[50] Also, the influence of menopausal status or chemotherapy-induced menopause is almost never examined.[51] Finally, studies differ in the assessed cognitive domains and in the methods by which cognitive impairment is defined.

What are CDK4/6 inhibitors?

CDK4/6 inhibitors have been added relatively recently to the arsenal of treatments for BC. They are currently approved as first- and second-line treatment for HR+, human epidermal growth factor receptor 2 (HER2)-negative metastatic BC, in combination with ET. At present, there are three different CDK4/6 inhibitors available for clinical application in

BC: palbociclib, ribociclib and abemaciclib.[52] Studies on use in the (neo-) adjuvant setting are ongoing.

The CDK4/6 inhibitors interfere with the cell cycle by inhibiting the CDK4/6-cyclin D1 signal transduction pathway. The cell cycle is divided into four phases: the G1 phase, the S phase, the G2 phase and the M phase. The CDK4-6/cyclin D1 signal transduction pathway regulates the transition from the G1 to the S phase. Cyclin D1 binds to CDK4/6, resulting in phosphorylation of the retinoblastoma (Rb) protein and the release of E2 transcription factors. These factors activate genes that regulate the transition from G1 to S phase. The CDK4/6 inhibitors exert their effect by inhibiting Cyclin D and CDK4/6 which prevent the release of transcription factors and achieve an arrest in cell division.[53]

Research also suggests that CDK4/6 inhibitors activate the immune system via a direct effect on tumor cells due to increased antigen presentation, and through an activating effect on immune cells in the tumor microenvironment.[54]

Recent RCTs have shown that the combination of ET with a CDK4/6 inhibitor in patients with metastatic HR+/HER2-negative BC prolongs the progression-free survival and thus appears to improve the patients' prognoses: An increase in the progression-free survival, is seen in first-line treatment of a CDK4/6 inhibitor in combination with an aromatase inhibitor or in second-line when added to fulvestrant.[55-60] In both treatment lines, the addition of the CDK 4/6 inhibitor to ET leads to a significant improvement in progression-free survival (PFS) compared to ET alone (average median PFS 13.6 months to 23.4 in the first-line and average median PFS 8.9 months to 15.5 months in the second-line) and of survival (37.3 months to 46.7 months).[61]

How might CDK4/6 inhibitors affect cognition?

CDK4/6 inhibitors appear to affect biological systems important for cognition. The arrest in cell division induced by the CDK4/6 inhibitors increases cytokine secretion, which already has been associated with cognitive problems and fatigue.[62-65] Fatigue is a frequently reported side effect of CDK4/6 inhibitors and may indirectly have a negative impact on cognitive functioning.[65-66]

The CDK4/6 inhibitors are not cell-specific. This means that the agents may interfere not only with the cell cycle of cancer cells, but also with the cell cycle of healthy cells, especially fast-growing, healthy cells such as progenitor cells in the brain.[67] Progenitor cells in the brain play an important role in neurogenesis, the production of new neurons. Cyclin D also appears to play a role in neurogenesis, especially in neural differentiation.[68] CDK4/6 inhibitors inhibit Cyclin D, which can lead to impaired neurogenesis and subsequently cognitive problems, such as learning and memory difficulties.

In other nonmalignant conditions, an adverse effect on cognition of pan-CDK inhibitors was as yet not observed. Limited preclinical studies in nonmalignant animal models of effects of first-generation pan-CDK inhibitors showed positive effects on cognition. In traumatic brain injury, it was found that CDK inhibitors (CR8, flavopiridol, roscovitine, and olomoucin) can reduce spatial memory impairment by decreasing neural loss, and astroglial and microglial activation.[69-70] In multiple sclerosis and schizophrenia, CDK inhibitors (flavopiridol) have been shown to reduce working memory problems by promoting remyelination by inhibiting microglia activation.[71]

What is known about the cognitive effects of CDK4/6 inhibitors?

Partly because of the only recent widespread application of CDK4/6 inhibitors, the possible effects of CDK4/6 inhibitors on cognition in BC

patients have hardly been investigated. To our knowledge, no study with formal cognitive measurements has been conducted to date.

Two recent RCTs with palbociclib (PALOMA-3) and abemaciclib (MONARCH 2) examined the effect on a subjective measure of cognition and suggested no additional effects of CDK4/6 inhibitors in both pre- and postmenopausal metastatic BC patients. In the PALOMA-3, both the combined ET group and the ET group reported significant deterioration in concentration and memory, without differences between the two groups in cognition.[72] In the MONARCH-2, the combined ET group reported complaints of cognition significantly later in their treatment than the ET group.[73] However, these findings should be interpreted with caution as cognition was only assessed by the subscale cognition (two questions) of the European Organization Research Treatment for Cancer-Quality of Life Questionnaire Core 30-questionnaire (EORTC-QLQ-C30).

Conclusion

Although no conclusions can yet be drawn, an increasing number of studies seem to indicate adverse effects of ET on cognitive functioning in pre-, peri- and postmenopausal BC patients, both self-reported and tested cognition. Reduced verbal memory is most consistently found. The type of ET seems to play a role: Tamoxifen may be the most harmful. The role of menopausal status on the impact of ET on cognition should be investigated since the effects might differ in severity between pre- and postmenopausal women.

As previously described, the potential effects of CDK4/6 inhibitors on cognition have hardly been investigated.

Research implications

Research on the cognitive effects of ET with or without CDK4/6 inhibitors is highly complex, not least because of the variety of treatment combinations. However, given the high incidence of HR+ BC and the increasing use of the aforementioned therapies, it is of great importance to gain more knowledge on the cognitive effects for clinicians to offer patients the best possible care now and in the future. In this paper, we focus on women with BC but in view of the increasing indications of ET ± CDK4/6 inhibitors, also in men and in other disease areas, such as prostate cancer, as well as ET combinations the relevance of evaluating this possible side effect becomes more adamant. Moreover, cognitive effects of ET combinations with targeted agents, such as PI3K and mTOR inhibitors, and next generation SERDs and Complete Estrogen Receptor Agonists should also be investigated, as cognitive effects of these agents are also biologically plausible.[74-75]

Prospective studies are needed to investigate the influence of ET with or without CDK4/6 inhibitors on cognition, preferably RCTs, with both objective and subjective cognition measurements pre- and post-treatment.

An example is the SONIA-EfFECT study (Evaluation of cognitive functioning in patients with metastatic BC treated with endocrine or combined therapy) that investigates cognitive effects of mono-ET and ET therapy plus CDK4/6 inhibitors in patients with metastatic HR+ BC (ClinicalTrials.gov Identifier: NCT03425838). Two hundred and sixty patients will undergo an online cognitive measurement at the start of treatment and nine months later (see Figure 3 for the study design). Cognition will be measured using the Amsterdam Cognition Scan (ACS); a validated online cognitive test battery that patients can independently complete from home (see Box 1 for more info).[76-77] Consistently combining subjective and objective measures of cognition provides a complete picture of potential cognitive effects.

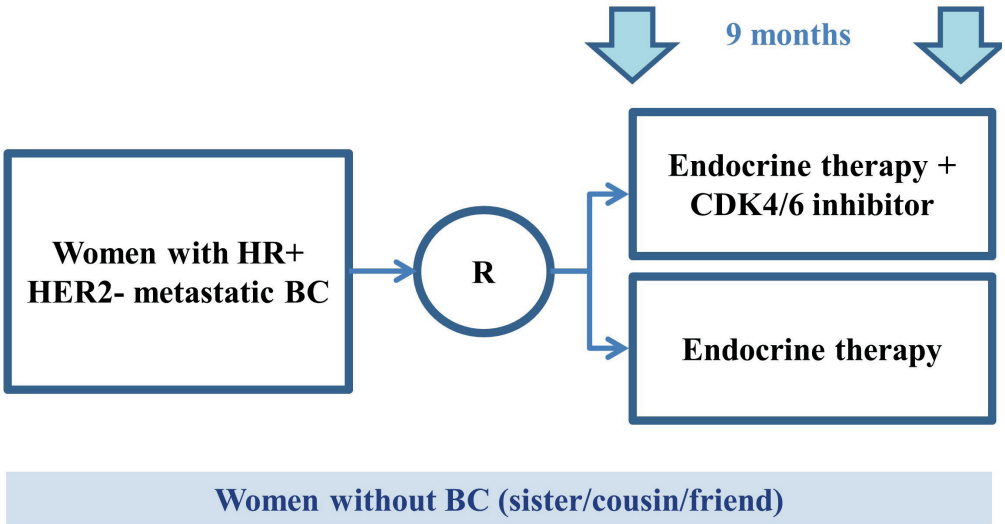


Figure 3. Study design SONIA-EfFECT study (ClinicalTrials.gov Identifier: NCT03425838)

Box 1. The Amsterdam Cognition Scan

The Amsterdam Cognition Scan (ACS) consists of a reliable and validated set of online cognitive tests that can be completed from home, unmonitored.[76-77] *The ACS assesses a broad spectrum of cognitive functions, including attention, verbal memory and executive functioning (see Figure 4). Online questionnaires on anxiety and depression, and fatigue are also incorporated. Normative data for the ACS are available. The ACS is proven a safe, patient-friendly, effective and practical tool to implement in clinical trials since hospital visits are not needed for cognitive assessments and completion time is relatively short. The ACS is translated to several languages, including American- and British-English and Swedish.*

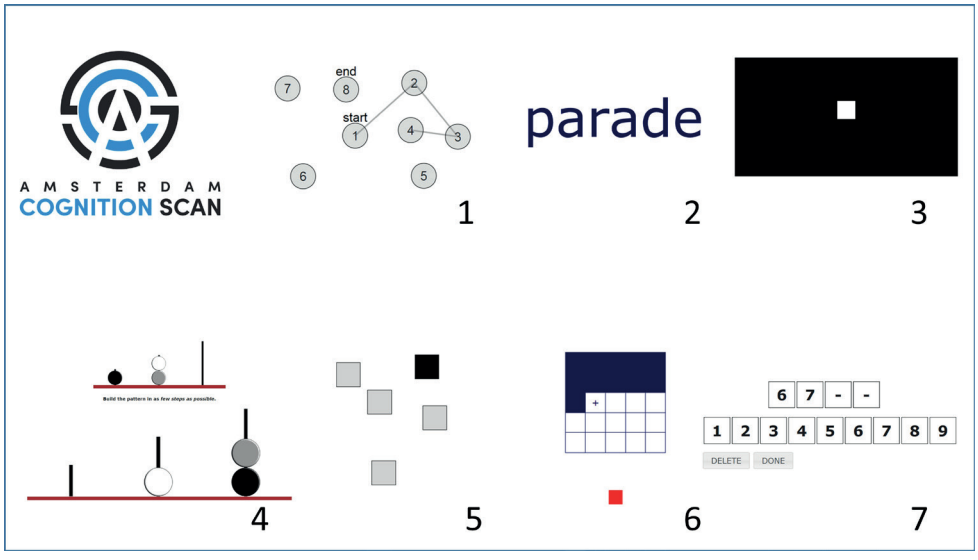


Figure 4. Screenshots of the Amsterdam Cognition Scan (ACS)

The ACS includes subtests assessing the cognitive domains: 1) processing speed, 2) learning and memory, 3) reaction speed, 4) executive function, 5) attention, 6) motor function and 7) working memory.

The severity of cognitive effects (and thereby the impact on daily life) should be investigated. To also observe late effects of ET, a follow-up duration up to several years after treatment is needed. Large samples are needed to ensure power and to account for the large heterogeneity in patient populations. Implementing online cognitive testing (such as the ACS) can help to achieve large research samples as cognitive assessments from home are less burdensome for patients, and time- and cost-effective compared to traditional paper-and-pencil tests.

Analyses should be stratified by factors that may also influence cognition. For example, the patient group (neoadjuvant, adjuvant, metastatic), menopausal status (pre-, peri, or postmenopausal), type and duration of ET (aromatase inhibitors, tamoxifen, fulvestrant, with or without LHRH agonist or ovarian suppression), whether or not combined with CDK4/6 inhibitor or other medication, pretreatment such as chemotherapy, and any presence of brain metastases should be taken into account. Patient-related symptoms such as fatigue, insomnia, and mood, and their relationship with cognition should be examined. Studies should use control groups including BC patients not treated with ET so that the effects on cognition of the BC itself can be taken into account.

The International Cancer and Cognition Taskforce (ICCTF) has developed guidelines on trial design and methods that can assist researchers who are interested in studying cancer (treatment)-related cognitive impairment.[78]

Consequences for clinical practice

Until more insight is gained about the cognitive effects of ET with or without CDK4/6 inhibitors, it is recommended in clinical practice to be aware of their possible detrimental influence on cognitive functioning, in patients with BC, but also with other disease areas.

When a patient has cognitive complaints in daily life, the patient can be helped by a referral to a clinical neuropsychologist. The clinical neuropsychologist will examine the medical background, administers neuropsychological tests and uses interview and various patient-reported outcomes to determine the presence of cognitive impairment, and the pattern and degree of cognitive strengths and weaknesses. This information can then be related to particular cognitive disorders or specific deficit clusters or, for example, to more general functioning problems with co-occurring cognitive complaints. Such a neuropsychological evaluation is critical to select the most optimal treatment, i.e., a treatment that targets the presumed biological and/or psychological basis of the problems and that will help the patient best. The cancer and cognition field is actively investigating ways to intervene against cognitive problems, showing promising results for cognitive rehabilitation using compensatory strategies and life style interventions such as exercise.[11]

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin*. 2021;71(1):7-33.
2. Osborne CK, Schiff R. Mechanisms of endocrine resistance in breast cancer. *Annu Rev Med*. 2011;62:233-247.
3. Wefel JS, Saleeba AK, Buzdar AU, Meyers CAJC. Acute and late onset cognitive dysfunction associated with chemotherapy in women with breast cancer. *Cancer*. 2010;116(14):3348-3356.
4. Sleurs C, Madoe A, Lagae L, et al. Genetic modulation of neurocognitive development in cancer patients throughout the lifespan: a systematic review. *Neuropsychol Rev*. 2019:29(2):190-219.
5. Kesler SR, Blayney DW. Neurotoxic effects of anthracycline-vs nonanthracycline-based chemotherapy on cognition in breast cancer survivors. *JAMA Oncol*. 2016;2(2):185-192.
6. Wefel JS, Vidrine DJ, Marani SK, et al. A prospective study of cognitive function in men with non-seminomatous germ cell tumors. *Psycho-Oncol*. 2014;23(6):626-633.
7. Collins B, MacKenzie J, Tasca GA, Scherling C, Smith A. Cognitive effects of chemotherapy in breast cancer patients: a dose–response study. *Psycho-Oncol*. 2013;22(7):1517-1527.
8. Schagen SB, Wefel JS. Chemotherapy-related changes in cognitive functioning. *EJC Suppl*. 2013;11(2):225.
9. Gibson EM, Monje M. Emerging mechanistic underpinnings and therapeutic targets for chemotherapy-related cognitive impairment. *Curr Opin Oncol*. 2019;31(6):531-539.
10. Li M, Caeyenberghs K. Longitudinal assessment of chemotherapy-induced changes in brain and cognitive functioning: a systematic review. *Neurosci Biobehav Rev*. 2018;92:304-317.
11. Wefel JS, Kesler SR, Noll KR, Schagen SB. Clinical characteristics, pathophysiology, and management of noncentral nervous system cancer-related cognitive impairment in adults. *CA Cancer J Clin*. 2015; 65(2):123-138.

12. Lezak MD, Howieson DB, Loring DW, Fischer JS. *Neuropsychological assessment*. New York, NY: Oxford University Press; 2004.
13. Wagner LI, Gray RJ, Sparano JA, et al. Patient-reported cognitive impairment among women with early breast cancer randomly assigned to endocrine therapy alone versus chemoendocrine therapy: results from TAILORx. *J Clin Oncol*. 2020;38(17):1875-1886.
14. Sul JK, Kluetz PG, Papadopoulos EJ, Keegan P. Clinical Outcome assessments in neuro-oncology: a regulatory perspective. *Neuro Oncol Pract*. 2016;3:4-9.
15. Pietras RJ, Márquez-Garbán DC. Membrane-associated estrogen receptor signaling pathways in human cancers. *Clin Cancer Res*. 2007;13(16):4672-4676.
16. Burstein HJ, Temin S, Anderson H, et al. Adjuvant endocrine therapy for women with hormone receptor–positive breast cancer: American Society of Clinical Oncology clinical practice guideline focused update. *J Clin Oncol*. 2014;32(21):2255.
17. Zwart W, de Leeuw R, Rondaij M, Neefjes J, Mancini MA, Michalides RJ. The hinge region of the human estrogen receptor determines functional synergy between AF-1 and AF-2 in the quantitative response to estradiol and tamoxifen. *J Cell Sc*. 2010;123(8):1253-1261.
18. Carpenter R, Miller W. Role of aromatase inhibitors in breast cancer. *Br J Cancer*. 2005;93(1):S1-S5.
19. Stocco C. Tissue physiology and pathology of aromatase. *Steroids*. 2012;77(1-2):27-35.
20. Gombos A. Selective oestrogen receptor degraders in breast cancer: a review and perspectives. *Curr Opin Oncol*. 2019;31(5):424-429.
21. Carlson RW. The history and mechanism of action of fulvestrant. *Clin Breast Cancer*. 2005;6:S5-S8.
22. Davies C, Pan H, Godwin J, et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet*. 2013;381(9869):805-816.

23. Early Breast Cancer Trialists' Collaborative Group. Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. *Lancet*. 2015;386(10001):1341-1352.
24. Reinert T, Debiasi M, Bines J, Barrios CH. Trends in progression-free survival (PFS) and time to progression (TTP) over time within first-line aromatase inhibitors trials in hormone receptor-positive advanced breast cancer. *Breast Cancer Res Treat*. 2018;168(2):457-465.
25. Luine VN. Estradiol and cognitive function: past, present and future. *Horm Behav*. 2014;66(4):602-618.
26. Arevalo M-A, Azcoitia I, Garcia-Segura LM. The neuroprotective actions of oestradiol and oestrogen receptors. *Nat Rev Neurosci*. 2015;16(1):17.
27. Fester L, Prange-Kiel J, Jarry H, Rune GM. Estrogen synthesis in the hippocampus. *Cell Tissue Res*. 2011;345(3):285.
28. Luine V, Frankfurt M. Interactions between estradiol, BDNF and dendritic spines in promoting memory. *Neurosci J*. 2013;239:34-45.
29. Sheppard PA, Choleris E, Galea LA. Structural plasticity of the hippocampus in response to estrogens in female rodents. *Mol Brain*. 2019;12(1):1-17.
30. Srivastava DP. Two-step wiring plasticity–a mechanism for estrogen-induced rewiring of cortical circuits. *J Steroid Biochem Mol Biol*. 2012;131(1-2):17-23.
31. Irwin RW, Yao J, Hamilton RT, Cadenas E, Brinton RD, Nilsen J. Progesterone and estrogen regulate oxidative metabolism in brain mitochondria. *Endocrinology*. 2008;149(6):3167-3175.
32. Sunday L, Osuna C, Krause DN, Duckles SP. Age alters cerebrovascular inflammation and effects of estrogen. *Am J Physiol Heart Circ*. 2007;292(5):H2333-H2340.
33. Brinton RD. The healthy cell bias of estrogen action: mitochondrial bioenergetics and neurological implications. *Trends Neurosci*. 2008;31(10):529-537.
34. Dumas J, Hancur-Bucci C, Naylor M, Sites C, Newhouse P. Estrogen treatment effects on anticholinergic-induced cognitive dysfunction in normal postmenopausal women. *Neuropsychopharmacology*. 2006;31(9):2065-2078.

35. Henderson JA, Bethea CL. Differential effects of ovarian steroids and raloxifene on serotonin 1A and 2C receptor protein expression in macaques. *Endocr J*. 2008;33(3):285-293.
36. Cella D, Fallowfield LJ. Recognition and management of treatment-related side effects for breast cancer patients receiving adjuvant endocrine therapy. *Breast Cancer Res Treat*. 2008;107(2):167-180.
37. Ito KI, Ito T, Okada T, et al. A case of brain metastases from breast cancer that responded to anastrozole monotherapy. *Breast*. 2009;15(4):435-437.
38. Kil K-E, Biegon A, Ding Y-S, et al. Synthesis and PET studies of [11C-cyano] letrozole (Femara), an aromatase inhibitor drug. *Nucl Med Biol*. 2009;36(2):215-223.
39. Lien EA, Solheim E, Ueland PM. Distribution of tamoxifen and its metabolites in rat and human tissues during steady-state treatment. *Cancer Res*. 1991;51(18):4837-4844.
40. Azcoitia I, Yague J, Garcia-Segura LM. Estradiol synthesis within the human brain. *J Neurosci*. 2011;191:139-147.
41. Ganz PA, Petersen L, Castellon SA, et al. Cognitive function after the initiation of adjuvant endocrine therapy in early-stage breast cancer: an observational cohort study. *J Clin Oncol*. 2014;32(31):3559.
42. Schilder CM, Seynaeve C, Beex LV, et al. Effects of tamoxifen and exemestane on cognitive functioning of postmenopausal patients with breast cancer: results from the neuropsychological side study of the tamoxifen and exemestane adjuvant multinational trial. *J Clin Oncol*. 2010;28(8):1294-1300.
43. Boele FW, Schilder CM, de Roode M-L, Deijen JB, Schagen SB. Cognitive functioning during long-term tamoxifen treatment in postmenopausal women with breast cancer. *Menopause*. 2015;22(1):17-25.
44. Underwood E, Rochon P, Moineddin R, et al. Cognitive sequelae of endocrine therapy in women treated for breast cancer: a meta-analysis. *Breast Cancer Res Treat*. 2018;168(2):299-310.
45. Gonzalez M, Cabrera-Socorro A, Pérez-García CG, et al. Distribution patterns of estrogen receptor α and β in the human cortex and hippocampus during development and adulthood. *J Comp Neurol*. 2007;503(6):790-802.

46. Jenkins VA, Ambroisine LM, Atkins L, Cuzick J, Howell A, Fallowfield LJ. Effects of anastrozole on cognitive performance in postmenopausal women: a randomised, double-blind chemoprevention trial (IBIS II). *Lancet Oncol*. 2008;9(10):953-961.
47. Phillips K-A, Ribi K, Sun Z, et al. Cognitive function in postmenopausal women receiving adjuvant letrozole or tamoxifen for breast cancer in the BIG 1-98 randomized trial. *Breast J*. 2010;19(5):388-395.
48. Ganz PA, Van Dyk K. Cognitive Impairment in Patients With Breast Cancer: Understanding the Impact of Chemotherapy and Endocrine Therapy. *J Clin Oncol*. 2020;38(17):1871-1874.
49. Van Dyk K, Crespi CM, Bower JE, Castellon SA, Petersen L, Ganz PA. The cognitive effects of endocrine therapy in survivors of breast cancer: A prospective longitudinal study up to 6 years after treatment. *Cancer*. 2019;125(5):681-689.
50. Blaustein JD. Treatments for breast cancer that affect cognitive function in postmenopausal women. *Policy Insights Behav Brain Sci*. 2017;4(2):170-177.
51. Bakoyiannis I, Tsigka E-A, Perrea D, Pergialiotis V. The impact of endocrine therapy on cognitive functions of breast cancer patients: a systematic review. *Clin Drug Investig*. 2016;36(2):109-118.
52. De Groot A, Kuijpers C, Kroep J. CDK4/6 inhibition in early and metastatic breast cancer: a review. *Cancer Treat Rev*. 2017;60:130-138.
53. Hamilton E, Infante JR. Targeting CDK4/6 in patients with cancer. *Cancer Treat Rev*. 2016;45:129-138.
54. Chaikovsky AC, Sage J. Beyond the cell cycle: enhancing the immune surveillance of tumors via CDK4/6 inhibition. *Mol Cancer Res*. 2018;16(10):1454-1457.
55. Finn RS, Crown JP, Lang I, et al. The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. *Lancet Oncol*. 2015;16(1):25-35.
56. Finn RS, Martin M, Rugo HS, et al. Palbociclib and letrozole in advanced breast cancer. *N Engl J Med*. 2016;375(20):1925-1936.

57. Hortobagyi GN, Stemmer SM, Burris HA, et al. Updated results from MONALEESA-2, a phase III trial of first-line ribociclib plus letrozole versus placebo plus letrozole in hormone receptor-positive, HER2-negative advanced breast cancer. *Ann Oncol*. 2018, 29(7), 1541-1547.
58. Cristofanilli M, Turner NC, Bondarenko I, et al. Fulvestrant plus palbociclib versus fulvestrant plus placebo for treatment of hormone-receptor-positive, HER2-negative metastatic breast cancer that progressed on previous endocrine therapy (PALOMA-3): final analysis of the multicentre, double-blind, phase 3 randomised controlled trial. *Lancet Oncol*. 2016;17(4):425-439.
59. Sledge Jr GW, Toi M, Neven P, et al. MONARCH 2: abemaciclib in combination with fulvestrant in women with HR+/HER2– advanced breast cancer who had progressed while receiving endocrine therapy. *J Clin Oncol*. 2017;35(25):2875-2884.
60. Slamon DJ, Neven P, Chia S, et al. Phase III randomized study of ribociclib and fulvestrant in hormone receptor-positive, human epidermal growth factor receptor 2-negative advanced breast cancer: MONALEESA-3. *J Clin Oncol*. 2018;36(24):2465-2472.
61. Sledge G, Toi M, Neven P, et al. MONARCH 2: overall survival of abemaciclib plus fulvestrant in patients with HR+, HER2-advanced breast cancer. *Ann Oncol*. 2019;30:v856.
62. Klein ME, Kovatcheva M, Davis LE, Tap WD, Koff A. CDK4/6 inhibitors: the mechanism of action may not be as simple as once thought. *Cancer Cell*. 2018;34(1):9-20.
63. Ohtani N, Takahashi A, Mann DJ, Hara E. Cellular senescence: a double-edged sword in the fight against cancer. *Exp Dermatol*. 2012;21:1-4.
64. Wardill HR, Mander KA, Van Sebille YZ, et al. Cytokine-mediated blood brain barrier disruption as a conduit for cancer/chemotherapy-associated neurotoxicity and cognitive dysfunction. *Int J Cancer*. 2016;139(12):2635-2645.
65. Joly F, Lange M, Dos Santos M, Vaz-Luis I, Di Meglio A. Long-Term Fatigue and Cognitive Disorders in Breast Cancer Survivors. *Cancers*. 2019;11(12):1896.

66. Lasheen S, Shohdy KS, Kassem L, Abdel-Rahman O. Fatigue, alopecia and stomatitis among patients with breast cancer receiving cyclin-dependent kinase 4 and 6 inhibitors: a systematic review and meta-analysis. *Expert Rev Anticancer Ther*. 2017;17(9):851-856.
67. Liu D-Z, Ander BP. Cell cycle inhibition without disruption of neurogenesis is a strategy for treatment of aberrant cell cycle diseases: an update. *Sci World J*. 2012;2012:491737.
68. Urbach A, Witte OW. Divide or commit–Revisiting the role of cell cycle regulators in adult hippocampal neurogenesis. *Front Cell Dev Biol*. 2019;7:55.
69. Skovira JW, Wu J, Matyas JJ, et al. Cell cycle inhibition reduces inflammatory responses, neuronal loss, and cognitive deficits induced by hypobaria exposure following traumatic brain injury. *J Neuroinflammation*. 2016;13(1):299.
70. Di Giovanni S, Movsesyan V, Ahmed F, et al. Cell cycle inhibition provides neuroprotection and reduces glial proliferation and scar formation after traumatic brain injury. *Proc NatI Acad Sci U S A*. 2005;102(23):8333-8338.
71. Mi G, Gao Y, Liu S, et al. Cyclin-dependent kinase inhibitor flavopiridol promotes remyelination in a cuprizone induced demyelination model. *Cell Cycle*. 2016;15(20):2780-2791.
72. Harbeck N, Iyer S, Turner N, et al. Quality of life with palbociclib plus fulvestrant in previously treated hormone receptor-positive, HER2-negative metastatic breast cancer: patient-reported outcomes from the PALOMA-3 trial. *Ann Oncol*. 2016;27(6):1047-1054.
73. Kaufman PA, Toi M, Neven P, et al. Health-Related Quality of Life in MONARCH 2: Abemaciclib plus Fulvestrant in Hormone Receptor-Positive, HER2-Negative Advanced Breast Cancer After Endocrine Therapy. *Oncologist*. 2020;25(2):e243.
74. Bandaru SS, Lin K, Roming SL, Vellipuram R, Harney JP. Effects of PI3K inhibition and low docosahexaenoic acid on cognition and behavior. *Physiol Behav*. 2010;100(3):239-244.

75. Halloran J, Hussong SA, Burbank R, et al. Chronic inhibition of mammalian target of rapamycin by rapamycin modulates cognitive and non-cognitive components of behavior throughout lifespan in mice. *Neuroscience*. 2012;223:102-113.
76. Feenstra HEM, Murre JMJ, Vermeulen IE, Kieffer JM, Schagen SB. Reliability and validity of a self-administered tool for online neuropsychological testing: the Amsterdam Cognition Scan. *J Clin Exp*. 2018;*40*(3):253-273.
77. Feenstra HEM, Vermeulen IE, Murre JMJ, Schagen SB. Online self-administered cognitive testing using the Amsterdam Cognition Scan: establishing psychometric properties and normative data. *J Med Internet Res*. 2018:20(5), e192.
78. Wefel JS, Vardy J, Ahles T, Schagen SB. International Cognition and Cancer Task Force recommendations to harmonise studies of cognitive function in patients with cancer. *Lancet Oncol*. 2011;12(7):703-708.

Chapter 6:
Effects of tamoxifen and exemestane on cognitive function in postmenopausal patients with breast cancer

Philippe R. Lee Meeuw Kjoe*, Jacobien M. Kieffer*, Brent J. Small, Willem Boogerd, Carolien M. Schilder, Elsken van der Wall, Elma Meershoek-Klein Kranenbarg, Cornelis J.H. van de Velde, Sanne B. Schagen.

*Shared first authorship.
JNCI Cancer Spectrum. In press.

Abstract

Background

Cognitive effects of tamoxifen have been described. We augment data from a previous short-term (ST) follow-up study with long-term (LT) data to evaluate ST and LT cognitive effects of tamoxifen followed by exemestane and exemestane in breast cancer patients.

Methods

Patients from the Tamoxifen and Exemestane Adjuvant Multinational trial received five years exemestane (Exemestane group, n=114), or 2.5 years tamoxifen followed by 2.5 years exemestane (Sequential group, n=92). Neuropsychological performance was assessed pre-endocrine therapy, after one year (ST follow-up) and at five years (LT follow-up). Controls (n=120) were assessed with parallel intervals. With random effects modeling we evaluated cognitive changes from baseline to ST and LT follow-up. Statistical tests were two-sided.

Results

After controlling for age, IQ, attrition, menopausal symptoms, anxiety/depression, and/or fatigue, the Sequential group showed ST and LT decline compared to controls on verbal memory (effect size (ES) .26; p = .01, and ES .34; p = .003) and executive function (ES .27; p = .007 and ES .38; p = .002). Compared with the Exemestane group, the Sequential group demonstrated ST decline on information processing speed (ES .33; p = .01) and executive function (ES .32; p = .01) and LT decline on verbal memory (ES: .33; p = .02). The Exemestane group showed no cognitive decline compared to controls.

Conclusion(s)

Cognitive adverse effects of tamoxifen alone and after switching to exemestane were observed, suggestive of a carry-over effect of tamoxifen. Our results underline the need for well-controlled, prospective trials studying cognitive effects of endocrine therapy.

Introduction

Adjuvant endocrine therapy (ET) is standard of care in the treatment of hormone receptor-positive (HR+) breast cancer (BC). Commonly two types of ET are given, depending, largely, on the patient's menopausal status. Selective Estrogen Receptor Modulators SERMs (e.g., tamoxifen) block estrogen receptors on BC cells, and Aromatase Inhibitors (AI, e.g., anastrozole, exemestane, and letrozole) inhibit production of estrogens by inactivating the enzyme aromatase in peripheral adipose tissue. Premenopausal women often receive tamoxifen plus ovarian function suppression (OFS) or an AI plus OFS for five to ten years post-surgery. In postmenopausal women, common treatments include five years of AI, five years of tamoxifen followed by an AI for two to three years, or tamoxifen for two to three years followed by an AI up to five years.[1]

ET may come with side-effects. Cognitive problems are a frequently reported symptom in BC patients using ET. The brain is widely responsive to estrogens. Important areas for cognition such as the hippocampus and frontal lobes are sensitive to estrogens. Therefore, downregulation of estrogen production or blocking its activity through ET could impact cognition. [2-4]

Several observational studies and randomized controlled trials (RCT's) using cognitive tests indicate that cognitive adverse effects of ET may exist and may differ between ET agents.[5-9] A recent comprehensive review reported frequency rates of cognitive dysfunction in 32% to 64% of patients receiving ET, with conflicting results on the differential impact of ET types.[10-11] Unfortunately, many studies had limitations including small sample sizes, short observation period, heterogeneity of ET and duration of use, and interference of other potentially neurotoxic therapies such as chemotherapy. Also, few studies have directly examined differences in cognitive effects between ET agents using RCT's and no study investigated cognitive effects following a switch.

The current study is a neuropsychological side study of the Tamoxifen and Exemestane Adjuvant Multinational (TEAM) trial in which the impact of five years of adjuvant exemestane (monotherapy) was compared with two and a half to three years of tamoxifen followed by two to two and a half years of exemestane (sequential treatment strategy) in postmenopausal HR+ early BC patients.[12-13] The neuropsychological study included only women who did not receive chemotherapy.

In an earlier publication, short-term (ST) follow-up results of this side study were published.[14] We included data of all patients who completed an assessment pre-ET and after one year of ET use. We found that at this one year follow-up, thus prior to switching, tamoxifen users (n = 80) performed worse on several cognitive domains, and reported more attention problems than cancer-free controls (n = 120) and exemestane users (n = 99).[14] Exemestane users did not differ in tested and self-reported cognition from controls.[15]

In the current study, we augment data from the ST follow-up[14] with data from the long-term (LT) follow-up (i.e., five years of exemestane monotherapy or sequential treatment of tamoxifen followed by exemestane). To make use of the longitudinal character of this study, we will not only use data from the LT follow-up, but also use and report on data from the baseline and ST follow-up. In addition, we did not only use data from complete cases, as we did in the prior publication,[14] but from all cases.

The current study aims therefore to describe the cognitive performance of BC patients using all three time points, e.g., from pre-ET to ST follow-up (one year after treatment) and LT follow-up (five years after treatment). Self-reported outcomes (anxiety/depression, menopausal symptoms, fatigue and cognitive function) are also evaluated. This is the first study that investigates cognitive effects of tamoxifen followed by an AI in BC patients.

Methods

Participants

Participants were Dutch postmenopausal HR+ BC patients who participated in the TEAM trial. They were randomly allocated to either five years of adjuvant exemestane (Exemestane group) (25 mg/day), or to two and a half to three years of tamoxifen (20 mg/d) followed by two to two and a half years of exemestane (25 mg/day; Sequential group). Eligibility criteria have been described in detail elsewhere.[13] Briefly, patients were included if they had histologically confirmed adenocarcinoma of the breast, positive estrogen and/or progesterone receptor status, and had undergone curative surgery. Additional exclusion criteria for this side study were: adjuvant chemotherapy, insufficient command of the Dutch language, Central Nervous System (CNS) disease or signs of dementia according to a dementia screening tool.[16] The control group consisted of female friends or relatives without a cancer history of about the same age as the patients. Controls were included if they had a postmenopausal status, no history of CNS disease, sufficient command of the Dutch language and no signs of dementia according to the dementia screening tool.[16] This neuropsychological study was approved by the central review board (Erasmus MC, Rotterdam) and the local medical ethics committees of all participating hospitals. All participants provided written informed consent.

Neuropsychological assessment

We used a battery of 18 cognitive tests that represent eight cognitive domains (Table 1).[17-25] All scores were coded such that higher scores indicate better performance.

Table 1

Summary of cognitive outcome measures

Cognitive domain	Cognitive tests	Outcome variable	Score range
Verbal memory	Rey auditory verbal learning test[15] Immediate recall	1. Total of 3 trials	0-45
	Rey auditory verbal learning test[15] Delayed recall	2. Total for long delay trial	0-15
	Visual Association Test[16]	3. Total of 2 trials	0-24
Visual memory	Wechsler Memory Scale visual memory subtest[17] Immediate recall	4. Points awarded according to scoring criteria	0-41
	Wechsler Memory Scale visual memory subtest[17] Delayed recall	5. Points awarded according to scoring criteria	0-41
Information processing speed	Stroop Card 1[18]	6. Seconds to complete	0+
	Stroop Card 2[18]	7. Seconds to complete	0+
	Trail making test part A[19]	8. Seconds to complete	0+
Executive functioning	Stroop Card 3[18]	9. Seconds to complete	0+
	Trail making test part B[19]	10. Seconds to complete	0+
Manual motor speed	Fepsy finger tapping[20] Dominant hand	11. Mean score of 5 trials of 10 sec	0+
	Fepsy finger tapping[20] Non-Dominant hand	12. Mean score of 5 trials of 10 sec	0+
Verbal fluency	Letter fluency (D,A,T)[21]	13. Total score of 3 letters/1 minute each	0+
	Category Fluency (Animals)[22]	14. Total score animals/1 minute	0+
	Category Fluency (professions)[22]	15. Total score professions/1 minute	0+
Reaction speed	Fepsy reaction times[20] Dominant hand	16. Mean m/score/30 trials	0+
	Fepsy reaction times[20] Non-Dominant hand	17. Mean m/score/30 trials	0+
Working memory	WAIS III Letter-number sequencing[23]	18. Total correct trials	0-21

Neuropsychological assessments were performed prior to start of ET (baseline), after one year (ST follow-up), and at five years (LT follow-up).

Patient reported outcomes

The 25-item Hopkins Symptom Checklist (HSCL_25) was used to assess anxiety and depression.[26] The scale has a one week time frame and the items were rated from 'not at all' (= 1) to 'extremely' (= 4). The outcome variable is the mean of all items (range 0-4). A mean score was calculated if participants answered 20 or more items.[27-28]

The 18-item endocrine subscale of the Functional Assessment of Cancer Therapy – Breast questionnaire (FACT B-ES) was used to assess menopausal symptoms.[29] This subscale has a four week time frame and consists of 18 items scored on a 5-point scale ranging from 'not at all' (= 0) to 'very much' (= 4). Outcome variable is the sum of reversed scores (0-72) so that higher scores indicate fewer endocrine symptoms. A mean score was calculated if at least half of the items were answered.[30]

We used the Fatigue symptom scale (three items) and the Cognitive Function scale (two items) of the 30-item EORTC QLQ-C30 version 3.0.[31-32] The scales have a one week time frame, and the items were rated from 'not at all' (= 1) to 'very much' (= 4). These scale scores were calculated according to standard EORTC scoring procedures and linearly transformed to a 0-100 scale. Missing values were replaced by the average score of the completed items in the same scale for each individual, provided that at least 50% of the items in that scale had been completed.[32] A higher score indicates a higher level of fatigue and a higher level of cognitive functioning.

Statistical analyses

Descriptive statistics were used to characterize the study sample. All raw cognitive test scores were converted into standardized z-scores based on the

baseline mean and standard deviation of the control group. Eight cognitive domain scores were calculated by the mean of the z-scores of the tests that belonged to the particular cognitive domain. The data of all patients and controls participating in the study were used; Attrition patterns across the three assessments were compared between groups. We evaluated between-group differences in change over time on anxiety/depression, menopausal symptoms, fatigue and cognitive function. All statistical tests were two-sided and significance was set at .05.

To analyse between-group differences in change over time on cognitive test performance, we conducted baseline to follow-up analyses (ST effect: T0 to T1, and LT effect: T0 to T2) using a mixed-effects modelling approach with a random intercept, maximum likelihood solution and an autoregressive covariance structure.[33] We chose this modelling approach as it can handle missing data, contrary to the earlier publication, in which cases with incomplete observations were discarded because the modelling procedure could not handle. For the primary analyses the control group was the reference category. If significant, we evaluated differences in mean change from baseline to short-term and baseline to long-term follow-up between the two patient groups and controls, and the two patient groups.

We investigated the impact of the following possible confounders: age, IQ, study attrition, and the following time dependent variables: fatigue (EORTC QLQ-C30), menopausal symptoms (FACT B-ES), and anxiety/depression (HSCL). We included confounders one by one in the model for every outcome to see if including a confounder would yield a better fit. These models were compared with Bayesian Information Criterion (BIC) and Akaike's Information Criterion (AIC).[34-35] Models with lower BIC or AIC values are considered better fitting models.[36]

Differences in mean change scores over time between the treatment groups and the control group were accompanied by standardized effect

sizes (ES) calculated based on the t-test statistic: (2*t)/(√degrees of freedom). ES of 0.2 was considered small, 0.5 moderate, and 0.8 large.[37]

Analyses were conducted on an intention-to-treat (ITT) basis. Additionally, we performed a per-protocol (PP) analysis on data from patients who met the criteria for minimal adherence with the intervention(s): Excluded from the PP analysis were Sequential group patients who continued with tamoxifen (instead of switching to exemestane) or switched to exemestane or another AI prematurely, Exemestane group patients who switched to tamoxifen, and patients, either from the Sequential or Exemestane group, who quit prematurely or went without ET more than a month at the time of cognitive assessment.

For all analyses, SPSS for Windows version 27 (IBM Corp., Armonk, NY) was used.

Results

In total, 206 patients (92 patients from the Sequential group and 114 patients from the Exemestane group) and 124 control women underwent cognitive assessment at baseline. Three (1 Sequential group, 1 Exemestane group, and 1 Control group) of the in total 330 participants were undergoing Methotrexate (e.g., for rheumatism or psoriasis) at baseline and were excluded from analyses. See Figure 1 for the inclusion flow-chart.

Compared to controls, the Sequential group and the Exemestane group were older (p = .01 and p = .02, respectively) and the Exemestane group had a lower estimated premorbid IQ (p = .02). See Table 2 for sociodemographic and clinical characteristics of the study population.

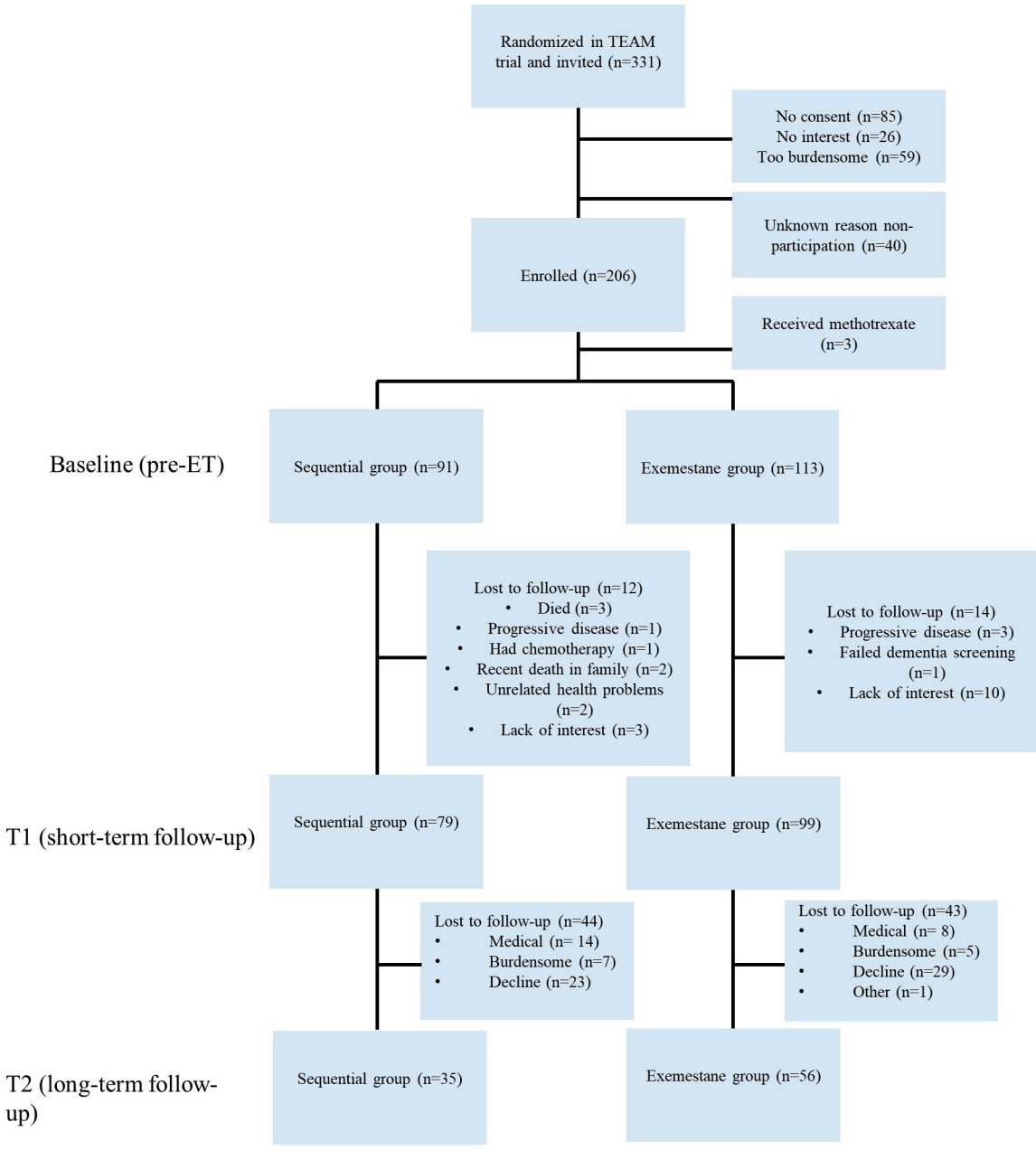


Figure 1. Flow-chart inclusion neuropsychological side study TEAM trial

Abbreviations: ET, endocrine therapy; T1, 1 year follow-up assessment; T2, 5 year follow-up assessment.

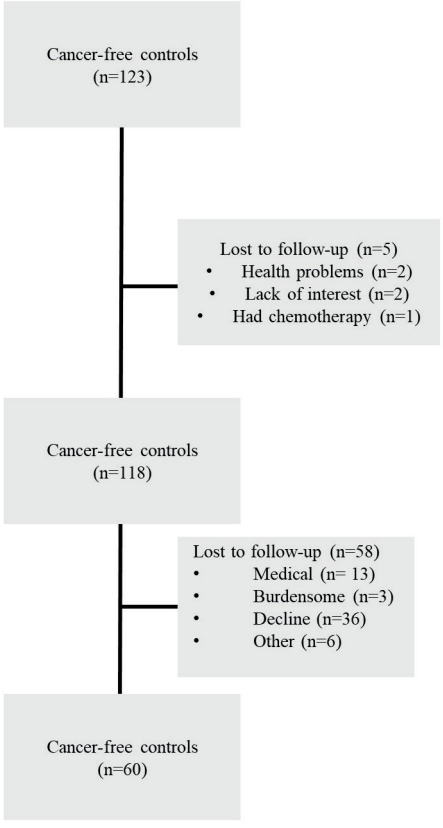


Figure continues

Table 2

Baseline sociodemographic and clinical characteristics of patients receiving tamoxifen followed by exemestane (Sequential group), patients receiving only exemestane (Exemestane group), and the control group

Characteristic	Sequential group (n=91)	Exemestane group (n=113)	Control group (n=123)	p
Age at randomisation (year), M (SD)	69.6 (7.9)	69.2 (7.0)	66.9 (8.0)	.02[a]
IQ, M (SD)	100.1 (19.9)	99.4 (18.8)	105.3 (19.0)	.04[b]
Time since surgery (month), M (SD)	1.3 (0.7)	1.5 (0.7)		.13
Age at menopause (year), M (SD)	49.3 (5.3)	49.7 (4.5)	48.2 (6.1)	.07
Radiotherapy at baseline or later, n (%)	49 (57)	75 (68)		.17
Ever use of HRT, n (%)	14 (15)	20 (18)	23 (19)	.10

Abbreviations: M, mean; SD, standard deviation; HRT, hormone replacement therapy.

[a]Post-hoc test: TAM/EXE vs. control $p = .01$, EXE vs. control $p = .02$, TAM/EXE vs. EXE $p = .68$.

[b]Post-hoc test: TAM/EXE vs. control $p = .05$, EXE vs. control $p = .02$, TAM/EXE vs. EXE $p = .79$.

Patient-reported outcomes

At baseline, the Sequential and the Exemestane group reported more fatigue compared to controls ($p < .001$, for both). This difference diminished over time as fatigue scores decreased ($p = .02$, and $p = .01$, respectively). At baseline, the Sequential group reported more anxiety/depression ($p < .001$) and endocrine symptoms ($p = .03$), compared to controls, and more endocrine symptoms compared to the Exemestane group ($p = .03$). During the trial the Exemestane group showed an increase in endocrine symptoms compared to controls ($p = .006$). At baseline, the Sequential group reported lower cognitive function compared to controls ($p < .001$) and the Exemestane group ($p = .002$). Changes over time in self-reported cognitive function did not differ between the Sequential group and the Exemestane group compared to controls ($p = .79$ and $p = .45$, respectively). The patient-reported outcome scores are depicted in Figure 2.

Compliance to neuropsychological assessment

More women in the Sequential and the Exemestane group completed only baseline compared to women in the control group ($p = .01$ and .008, respectively). Three drop-out patterns were distinguished (see Supplementary Table 1): 1. Completed only baseline, 2. Completed baseline and first follow-up (T0 and T1), 3. Completed all three assessments (T0, T1, and T2). Women who completed only baseline and who completed baseline and first follow-up were older than those who completed all assessments ($p = .007$ and $p < .001$), and women who completed only baseline had a lower intelligence quotient (IQ) than those who completed all assessments ($p = .002$), meaning that relatively more younger patients with a higher IQ completed all cognitive assessments. No differences were found between patients from the three drop-out patterns in anxiety/depression, menopausal symptoms, fatigue and self-reported cognition.

a

b

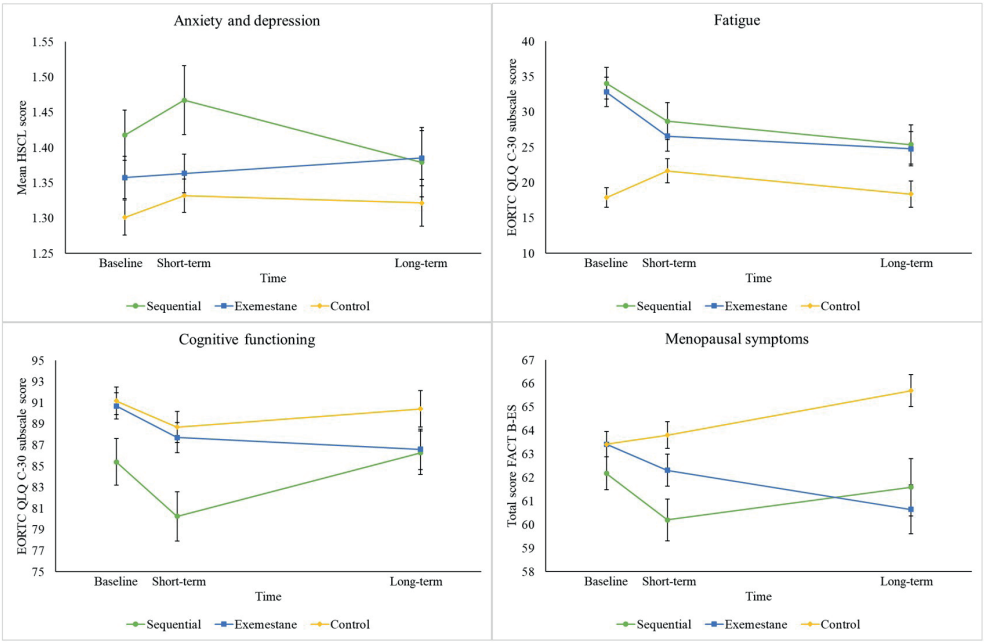


Figure 2. Change over time in patient-reported outcomes (anxiety/depression, fatigue, cognitive functioning and menopausal symptoms) in the Sequential, the Exemestane and the control group

Abbreviations: FU, follow-up.

Anxiety and depression were measured by the HSCL, Fatigue and Cognitive functioning by the EORTC QLQ C-30 subscales Fatigue and Cognitive functioning), and Endocrine symptoms by the FACT B-ES.

NOTE. For the upper two figures (Anxiety and depression, and Fatigue), higher scores represent more complaints. For the lower two figures (Self-reported cognitive functioning and Menopausal symptoms), higher scores represent less complaints.

Adherence to trial protocol

See Supplementary Figure 1 for an overview of participants and dropouts in the per-protocol analyses.

Model selection

For all analyses, adjustment was required based on AIC and BIC values. Most models were at least adjusted for age, IQ and menopausal symptoms. Between-group differences did in general not change after adjustment. Supplementary Figure 2 shows the differences per model.

Sequential and Exemestane group versus the control group

ITT analyses showed that the Sequential group had short-term and long-term decline on verbal memory (p = .01, ES: .26 and p = .003, ES: .34) and executive function (p = .007, ES: .27 and p = .002, ES: .38) compared to controls. The Exemestane group did not show decline on any cognitive domain compared to controls. A short-term improvement was found on information processing speed for the Exemestane group compared to controls (p = .02, ES: .22). The ITT results were confirmed in the PP analyses (see Table 3 and Supplementary Table 2 and 3 for all results). Figure 3 depicts changes over time for verbal memory and executive function (see Supplementary Figure 3 for changes over time for each cognitive domain).

Sequential group versus Exemestane group

The Sequential group showed a short-term decline on information processing speed (p = .01, ES: .33) and executive function (p = .01, ES: .32), and a long-term decline on verbal memory (p = .02, ES: .33) compared to the Exemestane group. The ITT results were confirmed in the PP analyses (see Supplementary Table 4 and 5).

Table 3 Results of the intention-to-treat analyses of the Sequential, the Exemestane and the control group

			Adjusted mean Z-scores					
			T0		T1		T2	
Cognitive domain/test	P-value overall group-by-time interaction	Group	Mean	95% CI	Mean	95% CI	Mean	95% CI
Verbal memory[b]	.02	Sequential	0.06	-0.11 to 0.22	0.03	-0.14 to 0.20	-0.32	-0.55 to -0.09
		Exemestane	-0.09	-0.24 to 0.05	0.03	-0.12 to 0.18	-0.16	-0.35 to 0.02
		Control	-0.09	-0.23 to 0.06	0.10	-0.04 to 0.24	-0.02	-0.22 to 0.18
Visual memory[b]	.12	Sequential	-0.38	-0.58 to -0.19	-0.36	-0.57 to -0.15	-0.37	-0.63 to -0.11
		Exemestane	-0.33	-0.51 to -0.15	-0.20	-0.38 to -0.02	-0.35	-0.56 to -0.13
		Control	-0.08	-0.25 to 0.09	-0.09	-0.26 to 0.09	0.12	-0.11 to 0.35
Information processing speed[b]	.04	Sequential	-0.24	-0.39 to -0.09	-0.28	-0.44 to -0.12	-0.44	-0.64 to -0.25
		Exemestane	-0.23	-0.36 to -0.09	-0.06	-0.20 to 0.08	-0.35	-0.51 to -0.19
		Control	-0.03	-0.16 to 0.10	-0.02	-0.15 to 0.11	-0.18	-0.35 to -0.01
Executive functioning[c]	.005	Sequential	-0.22	-0.41 to -0.04	-0.33	-0.52 to -0.14	-0.55	-0.77 to -0.32
		Exemestane	-0.35	-0.51 to -0.18	-0.21	-0.38 to -0.04	-0.41	-0.60 to -0.21
		Control	-0.05	-0.20 to 0.11	0.06	-0.10 to 0.22	0.04	-0.16 to 0.24
Motor speed[d]	.31	Sequential	-0.02	-0.18 to 0.14	-0.01	-0.17 to 0.16	-0.22	-0.41 to -0.02
		Exemestane	-0.05	-0.20 to 0.09	0.01	-0.14 to 0.16	-0.22	-0.39 to -0.06
		Control	-0.08	-0.21 to 0.06	0.02	-0.12 to 0.16	-0.07	-0.23 to 0.10
Verbal fluency[e]	.76	Sequential	-0.39	-0.53 to -0.25	-0.39	-0.54 to -0.25	-0.39	-0.57 to -0.21
		Exemestane	-0.42	-0.54 to -0.29	-0.37	-0.50 to -0.24	-0.48	-0.63 to -0.34
		Control	-0.09	-0.21 to 0.03	-0.07	-0.19 to 0.05	-0.16	-0.30 to -0.01
Reaction speed[f]	.523	Sequential	-0.27	-0.47 to -0.06	-0.40	-0.62 to -0.18	-0.27	-0.55 to 0.00
		Exemestane	-0.22	-0.40 to -0.03	-0.16	-0.36 to 0.03	-0.25	-0.49 to -0.02
		Control	0.02	-0.16 to 0.19	0.01	-0.17 to 0.19	-0.05	-0.30 to 0.20
Working memory[e]	.21	Sequential	-0.33	-0.51 to -0.15	-0.16	-0.35 to 0.03	-0.43	-0.70 to -0.15
		Exemestane	-0.30	-0.46 to -0.13	-0.20	-0.37 to -0.03	-0.45	-0.67 to -0.24
		Control	-0.06	-0.21 to 0.10	-0.01	-0.16 to 0.15	0.08	-0.14 to 0.31

Abbreviations: T0, baseline; T1; 1 year follow-up; T2, 5 year follow-up; CI, confidence interval; ES, effect size; AIC, Akaike information criterion; BIC, Bayesian information criterion.

a Effect sizes: .20 small effect, .50 moderate effect, .80 large effect.
b Adjusted for age, IQ, FACT-ES.
c Adjusted for age, IQ, HSCL, FACT-ES.
d Adjusted for age, IQ, HSCL.

Table continues

Sequential group and exemestane vs. control									
T0-T1				T0-T2					
Mean difference in change over time	95% CI	p	ES[a]	Mean difference in change over time	95% CI	p	ES[a]	AIC	BIC
-0.21	-0.38 to -0.05	.01	-0.26	-0.44	-0.72 to -0.15	.003	-0.34		
-0.06	-0.21 to 0.09	.42	-0.08	-0.13	-0.39 to 0.12	.31	-0.12	1375.19	1498.98
0.03	-0.18 to 0.23	.80	0.03	-0.19	-0.51 to 0.13	.24	-0.14		
0.13	-0.06 to 0.32	.17	0.14	-0.22	-0.51 to 0.07	.14	-0.18	1724.22	1848.96
-0.05	-0.19 to 0.09	.50	-0.07	-0.05	-0.29 to 0.18	.67	-0.05		
0.15	0.02 to 0.28	.02	0.22	0.02	-0.19 to 0.24	.82	0.03	1261.08	1385.96
-0.22	-0.37 to -0.06	.007	-0.27	-0.41	-0.68 to -0.15	.002	-0.38		
0.03	-0.11 to 0.17	.69	0.04	-0.15	-0.39 to 0.09	.22	-0.15	1475.47	1623.23
-0.08	-0.22 to 0.06	.25	-0.11	-0.21	-0.43 to 0.02	.07	-0.22		
-0.03	-0.16 to 0.10	.62	-0.05	-0.18	-0.39 to 0.02	.08	-0.22	1292.93	1417.82
-0.02	-0.15 to 0.10	.70	-0.04	0.06	-0.12 to 0.25	.49	0.09		
0.03	-0.09 to 0.14	.66	0.04	0.00	-0.16 to 0.16	>.99	0.00	1140.02	1242.04
-0.12	-0.35 to 0.11	.305	-0.10	0.05	-0.30 to 0.40	.76	0.04		
0.06	-0.15 to 0.27	.573	0.06	0.02	-0.30 to 0.35	.89	0.02	1813.94	1938.06
0.12	-0.14 to 0.38	.36	0.09	-0.23	-0.62 to 0.16	.24	-0.13		
0.04	-0.20 to 0.29	.72	0.03	-0.30	-0.65 to 0.05	.10	-0.19	1833.39	1935.06

e Adjusted for age, IQ.
f Adjusted for IQ, HSCL, FACT-ES.

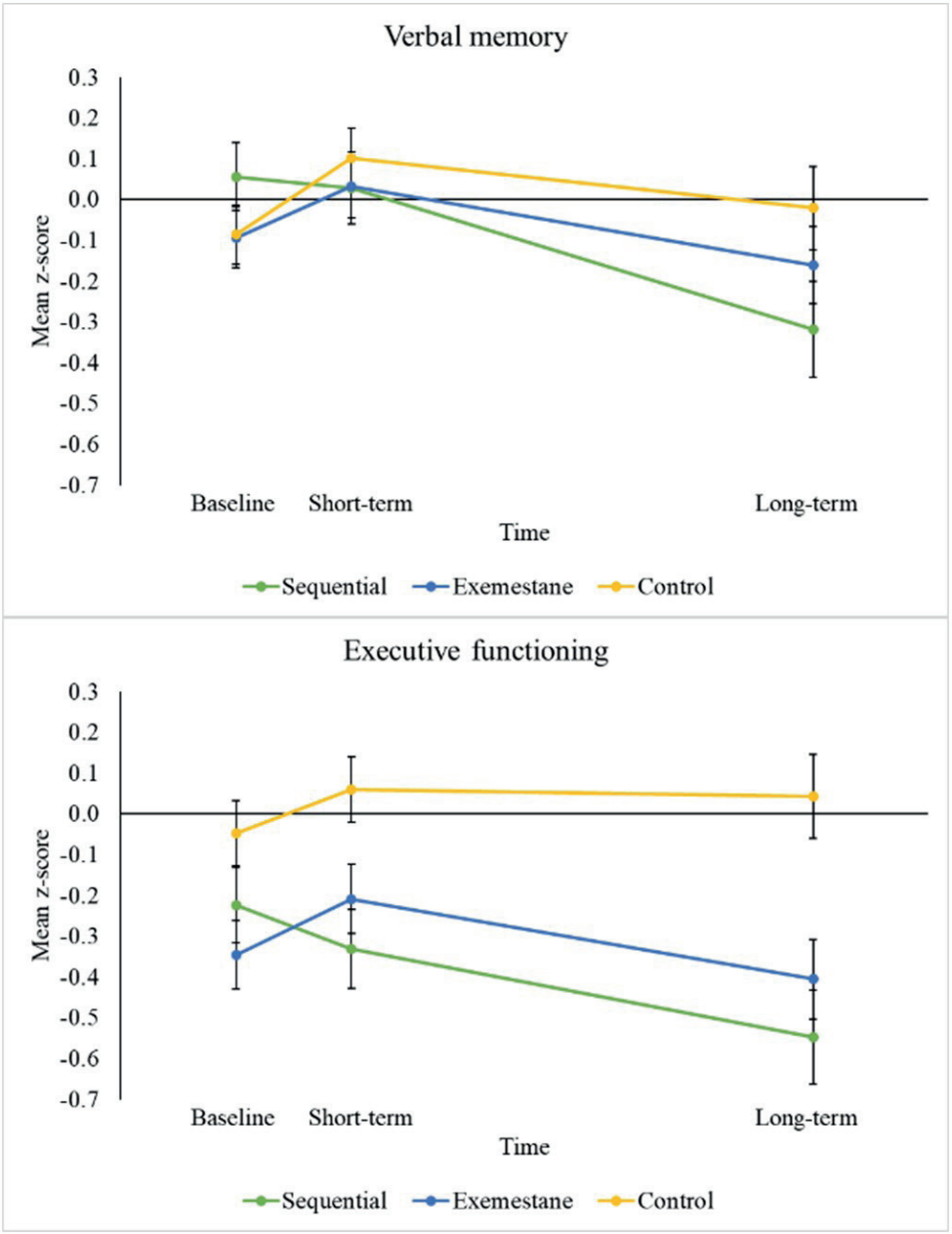


Figure 3. Adjusted standardized change over time on verbal memory and executive function between patients and controls

Mean z-scores are depicted per cognitive domain.

Discussion

Our earlier findings suggested cognitive adverse effects of tamoxifen and no effects of exemestane in a one-year follow-up study in postmenopausal early HR+ BC patients.[14] The current study augments these data by evaluating short-term and long-term cognitive effects of tamoxifen followed by exemestane and exemestane using all data and time points. After controlling for age, IQ, attrition patterns, menopausal symptoms, anxiety/depression, and/or fatigue, tamoxifen and tamoxifen followed by exemestane was associated with decline in verbal memory and executive functioning. Observed effects were of small magnitude. Treatment with exemestane only was not associated with cognitive decline. The ITT and PP analyses yielded comparable results. We found no differences between the patient groups and controls in changes over time in self-reported cognitive function and anxiety/depression. Differences in self-reported fatigue at baseline diminished over time in both patient groups compared to controls. The Exemestane group reported more menopausal symptoms over time than controls.

The observation of tamoxifen's small cognitive effects on verbal memory and executive functioning is in line with an emerging body of (predominantly cross-sectional) studies reporting adverse effects of tamoxifen.[38] Our observation of (small) adverse effects of tamoxifen alone and tamoxifen followed by exemestane, combined with the absence of any effect of exemestane mono-therapy, suggest a potential carry-over effect of tamoxifen. Cognitive effects of tamoxifen followed by another agent have not been examined previously. A small imaging study showed that tamoxifen was associated with structural brain changes, e.g., smaller hippocampal volumes,[39] which could partly explain long-term effects of tamoxifen.

The impact of ET on cognition and brain health is poorly studied and incompletely understood, both from a preclinical and clinical perspective.[10-11] The mixed findings in the literature do not give clear

direction for interpreting our results. Several important differences between tamoxifen and exemestane may have contributed to our observations, which could also provide guidance to future research initiatives: First, AI inactivate aromatase, thereby preventing conversion of androgens into estrogens. Tamoxifen, however, competitively binds to ERs. Tamoxifen has anti-estrogenic effects on breast tissue, but does not act as an anti-estrogen in all tissues.[40-42] Whether tamoxifen has an estrogenic or anti-estrogenic effect (or both) on the brain is unknown. The absence of cognitive effects of exemestane in our study may suggest that further downregulation of estrogen production in already postmenopausal women does not impact cognition. Also, ERα can be activated ligand-independently, without estrogens by growth factors such as insulin like growth factor 1.[43] ERs can still exert some transcriptional actions during exemestane use, in contrast to tamoxifen use. This might have contributed to the absence of cognitive changes following exemestane compared to tamoxifen. In addition, exemestane and its metabolites have a mild androgenic property that could be protective for cognition.[44] Our findings warrant further fundamental research to characterize the influence of tamoxifen as estrogenic, anti-estrogenic or maybe of a different character.

To better understand the cause of cognitive effects of tamoxifen, it could be useful to study the pharmacokinetics of tamoxifen in relation to cognition. Tamoxifen is a prodrug that exert its effects only after conversion to active metabolites mainly by the liver.[45] Therefore, focusing on the relation between tamoxifen's metabolites, e.g., endoxifen, and cognition can lead to a more direct examination of causality. Our research group is initiating a substudy of the TOTAM trial (Netherlands Trial Register NL6919/NTR7113) on dose- and serum-dependent cognitive effects of tamoxifen and its metabolites.

A limitation of this study is the small sample size at the long-term follow-up, reducing the statistical power of the long-term evaluation. Results of the long-term evaluation should be viewed as hypothesis-generating

and need to be confirmed in larger studies. The third cognitive assessment was not part of the original study protocol, which may have contributed to a lower accrual rate. Another limitation is that in both patient groups, vulnerable patients (of older age and with lower IQ) dropped out early, which could have biased the findings. The strengths of the current study include the prospective nature, the inclusion of chemotherapy-naïve patients only and a control group of women without a cancer history.

In conclusion, our results confirm our previous short-term findings and add to these by showing that sequential treatment with tamoxifen and exemestane was associated with short- and long-term decline on several tested cognitive functions while exemestane only was not. The modest adverse effects of tamoxifen and tamoxifen followed by exemestane occurred in absence of treatment-specific changes in self-reported cognitive symptoms. As cognitive test performance is associated with outcomes such as financial management, employability and medication management, adverse effects, even modest effects could be of clinical relevance.[46] Studies with additional measures are needed to investigate the impact on real-world performance. Given the large group of women receiving ET, the results underline the clinical need for well-controlled, prospective trials.

References

1. Early Breast Cancer Trialists' Collaborative Group. Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. *Lancet*. 2015;386(10001):1341-1352.
2. Norbury R, Cutter WJ, Compton J, et al. The neuroprotective effects of estrogen on the aging brain. *Exp Gerontol*. 2003;38(1-2):109-117.
3. Turgeon JL, Carr MC, Maki PM, et al. Complex actions of sex steroids in adipose tissue, the cardiovascular system, and brain: insights from basic science and clinical studies. *Endocr Rev*. 2006;27(6):575-605.
4. Maki PM, Dumas J. Mechanisms of action of estrogen in the brain: insights from human neuroimaging and psychopharmacologic studies. *Semin Reprod Med*. 2009;27(3):250-259.
5. Van Dyk K, Crespi CM, Bower JE, et al. The cognitive effects of endocrine therapy in survivors of breast cancer: a prospective longitudinal study up to 6 years after treatment. *Cancer*. 2019;125.5:681-689.
6. Jenkins V, Shilling V, Fallowfield L, et al. Does hormone therapy for the treatment of breast cancer have a detrimental effect on memory and cognition? A pilot study. *Psycho-Oncol*. 2004;13.1:61-66.
7. Phillips K-A, Ribi K, Sun Z, et al. Cognitive function in postmenopausal women receiving adjuvant letrozole or tamoxifen for breast cancer in the BIG 1-98 randomized trial. *Breast J*. 2010;19(5):388-395.
8. Le Ruhn E, Delbeuck X, Lefeuvre-Plesse C, et al. A phase III randomized multicenter trial evaluating cognition in post-menopausal breast cancer patients receiving adjuvant hormonotherapy. *Breast Cancer Res Treat*. 2015;152.3:569-580.
9. Jenkins VA, Ambroisine LM, Atkins L, et al. Effects of anastrozole on cognitive performance in postmenopausal women: a randomised, double-blind chemoprevention trial (IBIS II). *Lancet Oncol*. 2008;9(10):953-961.
10. Haggstrom LR, Vardy JL, Carson EK, et al. Effects of endocrine therapy on cognitive function in patients with breast cancer: a comprehensive review. *Cancers*. 2022;14(4):920.
11. Lee Meeuw Kjoe PR, van der Wall E, Schagen SB. Endocrine therapy

with or without CDK4/6 inhibitors in women with hormone-receptor positive breast cancer: what do we know about the effects on cognition? *Clin Breast Cancer*. 2022;22(3):191-199.

12. van de Velde CJH, Rea D, Seynaeve C, et al. Adjuvant tamoxifen and exemestane in early breast cancer (TEAM): a randomised phase 3 trial. *Lancet*. 2011;377(9762):321-331.
13. Schilder CM, Eggens PC, Seynaeve C, et al. Neuropsychological functioning in postmenopausal breast cancer patients treated with tamoxifen or exemestane after AC-chemotherapy: cross-sectional findings from the neuropsychological TEAM-side study. *Acta Oncol*. 2009;48(1):76-85.
14. Schilder CM, Seynaeve C, Beex LV, et al. Effects of tamoxifen and exemestane on cognitive functioning of postmenopausal patients with breast cancer: results from the neuropsychological side study of the tamoxifen and exemestane adjuvant multinational trial. *J Clin Oncol*. 2010;28(8):1294-1300.
15. Schilder CM, Seynaeve C, Linn SC, et al. Self-reported cognitive functioning in postmenopausal breast cancer patients before and during endocrine treatment: findings from the neuropsychological TEAM side-study. *Psycho-Oncol*. 2012;21.5:479-487.
16. Meulen EFJ, Schmand B, van Campen JP, et al. The seven minute screen: a neurocognitive screening test highly sensitive to various types of dementia. *J Neurol Neurosurg Psychiatry*. 2004;75(5):700-705.
17. van den Burg W, Saan RJ, Deelman BG. *15-Woordentest: Provisional Manual*. Groningen: University Hospital, Department of Neuropsychology;1985.
18. Lindeboom J, Schmand B, Tulner L, et al. Visual association test to detect early dementia of the Alzheimer type. *J Neurol Neurosurg Psychiatry*. 2002;73(2):126-133.
19. Wechsler D. *Wechsler memory scale-revised*. San Antonio, TX: Psychological Corporation; 1987.
20. Hammes JG. *De Stroop kleur-woord test*. Amsterdam: Harcourt Test Publ; 1978.
21. Reitan RM. Validity of the Trail Making Test as an indicator of organic

brain damage. *Percept Mot*. 1958;8(3):271-276.
22. Alpherts W, Aldenkamp AP. *FePsy: the iron psyche*. Heemstede: Instituut voor epilepsiebestrijding; 1994.
23. van der Elst WIM, van Boxtel MP, van Breukelen GJ, et al. Normative data for the Animal, Profession and Letter M Naming verbal fluency tests for Dutch speaking participants and the effects of age, education, and sex. *J Int Neuropsychol Soc*. 2006;12(1):80-89.
24. Lezak MD, Howieson DB, Loring DW, et al. *Neuropsychological assessment*. New York: Oxford University Press; 2004.
25. Wechsler D. *WAIS-III Nederlandstalige Bewerking. Technische Handleiding*. Lisse, the Netherlands: Swets & Zeitlinger; 2000.
26. Fröjdh K, Håkansson A, Karlsson I. The Hopkins Symptom Checklist-25 is a sensitive case-finder of clinically important depressive states in elderly people in primary care. *Int J Geriatr Psychiatry*. 2004;19(4):386-390.
27. Nettelbladt P, Hansson L, Stefansson CG, et al. Test characteristics of the Hopkins Symptom Check List-25 (HSCL-25) in Sweden, using the Present State Examination (PSE-9) as a caseness criterion. *Soc Psychiatry Psychiatr Epidemiol*. 1993;28(3):130-133.
28. Sandanger I, Moum T, Ingebrigtsen G, et al. The meaning and significance of caseness: the Hopkins Symptom Checklist-25 and the Composite International Diagnostic Interview II. *Soc Psychiatry Psychiatr Epidemiol*. 1999;34(1):53-59.
29. Fallowfield L, Leaity S, Howell A. Assessment of quality of life in women undergoing hormonal therapy for breast cancer: validation of an endocrine symptom subscale for the FACT-B. *Breast Cancer Res Treat*. 1999;55(2):187-197.
30. Cella D. *FACIT Manual: Manual of the Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System*. Evanston, IL: CORE; 1997.
31. Aaronson N, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*. 1993;85(5):365-376.

32. Fayers PM, Aaronson NK, Bjordal K, et al. *EORTC QLQ-C30 Scoring Manual*. 3rd ed. Brussels: EORTC; 2001.
33. Littell RC, Pendergast J, Natarajan R. Modelling covariance structure in the analysis of repeated measures data. *Stat Med*. 2000;19(13):1793-1819.
34. Schwarz G. Estimating the Dimension of a Model. *Ann Stat*. 1978;461-464.
35. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Caski F, eds. *Proceedings of the Second International Symposium on Information Theory*. Budapest: Akademiai Kiado; 1973:267-281.
36. Raftery A. Bayesian model selection in social research. *Sociol Methodol*. 1995;25:111-163.
37. Cohen J. *Statistical power analysis for the behavioural sciences*. revised ed. New York: Academic Press; 1977.
38. Jebahi F, Sharma S, Bloss JE, et al. Effects of tamoxifen on cognition and language in women with breast cancer: a systematic search and a scoping review. *Psycho-Oncol*. 2021;30(8):1262-1277.
39. Eberling JL, Wu C, Tong-Turnbeaugh R, et al. Estrogen- and tamoxifen-associated effects on brain structure and function. *Neuroimage*. 2004;21(1):364-371.
40. Newhouse P, Albert K, Astur R, et al. Tamoxifen improves cholinergically modulated cognitive performance in postmenopausal women. *Neuropsychopharmacology*. 2013;38(13):2632-2643.
41. Zhang Z, Park JW, Ahn IS, et al. Estrogen receptor alpha in the brain mediates tamoxifen-induced changes in physiology in mice. *Elife*. 2021;10:e63333.
42. Lichtenfels M, da Silva Dornelles A, dos Santos Petry F, et al. The anticancer estrogen receptor antagonist tamoxifen impairs consolidation of inhibitory avoidance memory through estrogen receptor alpha. *J Neural Transm*. 2017;124(11):1331-1339.
43. Baumgartner NE, Daniel JM. Estrogen receptor α: a critical role in successful female cognitive aging. *Climacteric*. 2021;24(4):333-339.

44. Hirshman E, Merritt P, Wang CC, et al. Evidence that androgenic and estrogenic metabolites contribute to the effects of dehydroepiandrosterone on cognition in postmenopausal women. *Horm Behav*. 2004;45(2), 144-155.
45. Jordan VC. Tamoxifen: a most unlikely pioneering medicine. *Nat Rev Drug Discov*. 2003;2(3):205-213.
46. Zwart W, Terra H, Linn SC, Schagen SB. Cognitive effects of endocrine therapy for breast cancer: keep calm and carry on? *Nat Rev Clin Oncol*. 2015;12(10):597-606.

Supplementary material

Supplementary Table 1

Drop-out patterns for the Sequential, Exemestane and control group

Drop-out pattern	Sequential group	Exemestane group	Control group
	N (%)	N (%)	N (%)
Completed Baseline only	11 (12)	14 (12)	4 (3)
Completed Baseline+FU1	45 (50)	42 (37)	57 (46)
Completed Baseline+FU1+FU2	35 (39)	57 (50)	62 (50)

Abbrevations: FU, follow-up.

Supplementary Table 2

Results of the intention-to-treat analyses of the Sequential, the Exemestane and the control group

			Adjusted mean Z-scores					
			To		T1		T2	
Cognitive domain/test	P-value overall group-by-time interaction	Group	Mean	95% CI	Mean	95% CI	Mean	95% CI
Verbal memory[b]	.02	Sequential	0.06	-0.11 to 0.22	0.03	-0.14 to 0.20	-0.32	-0.55 to -0.09
		Exemestane	-0.09	-0.24 to 0.05	0.03	-0.12 to 0.18	-0.16	-0.35 to 0.02
		Control	-0.09	-0.23 to 0.06	0.10	-0.04 to 0.24	-0.02	-0.22 to 0.18
RAVLT IR[c]	.04	Sequential	0.04	-0.15 to 0.24	0.03	-0.18 to 0.24	-0.27	-0.54 to -0.01
		Exemestane	-0.18	-0.36 to 0.00	-0.02	-0.20 to 0.16	-0.25	-0.47 to -0.03
		Control	-0.09	-0.25 to 0.08	0.04	-0.13 to 0.21	0.12	-0.12 to 0.36
RAVLT DR[c]	.12	Sequential	0.09	-0.11 to 0.28	0.03	-0.17 to 0.24	-0.08	-0.33 to 0.18
		Exemestane	-0.06	-0.23 to 0.12	0.06	-0.12 to 0.24	-0.01	-0.23 to 0.20
		Control	-0.08	-0.25 to 0.09	0.15	-0.02 to 0.32	0.08	-0.15 to 0.31
VAT[d]	.45	Sequential	0.07	-0.17 to 0.31	0.05	-0.21 to 0.31	-0.39	-0.76 to -0.02
		Exemestane	-0.12	-0.33 to 0.10	0.05	-0.17 to 0.27	-0.25	-0.54 to 0.04
		Control	-0.04	-0.24 to 0.17	0.17	-0.04 to 0.39	-0.13	-0.43 to 0.18
Visual memory[b]	.12	Sequential	-0.38	-0.58 to -0.19	-0.36	-0.57 to -0.15	-0.37	-0.63 to -0.11
		Exemestane	-0.33	-0.51 to -0.15	-0.20	-0.38 to -0.02	-0.35	-0.56 to -0.13
		Control	-0.08	-0.25 to 0.09	-0.09	-0.26 to 0.09	0.12	-0.11 to 0.35
WMS Visual Memory IR[b]	.73	Sequential	-0.35	-0.55 to -0.14	-0.40	-0.62 to -0.18	-0.38	-0.66 to -0.10
		Exemestane	-0.31	-0.49 to -0.12	-0.28	-0.47 to -0.09	-0.32	-0.55 to -0.09
		Control	-0.10	-0.28 to 0.08	-0.14	-0.32 to 0.04	0.03	-0.22 to 0.27
WMS Visual Memory DR[b]	.046	Sequential	-0.46	-0.67 to -0.25	-0.35	-0.58 to -0.13	-0.36	-0.65 to -0.08
		Exemestane	-0.36	-0.55 to -0.17	-0.12	-0.32 to 0.07	-0.35	-0.59 to -0.12
		Control	-0.07	-0.25 to 0.11	-0.04	-0.22 to 0.15	0.21	-0.04 to 0.47
Information processing speed[b]	.04	Sequential	-0.24	-0.39 to -0.09	-0.28	-0.44 to -0.12	-0.44	-0.64 to -0.25
		Exemestane	-0.23	-0.36 to -0.09	-0.06	-0.20 to 0.08	-0.35	-0.51 to -0.19
		Control	-0.03	-0.16 to 0.10	-0.02	-0.15 to 0.11	-0.18	-0.35 to -0.01
Stroop card 1[c]	.70	Sequential	-0.29	-0.50 to -0.08	-0.30	-0.52 to -0.07	-0.67	-0.95 to -0.39
		Exemestane	-0.21	-0.40 to -0.03	-0.19	-0.38 to 0.01	-0.52	-0.75 to -0.28
		Control	-0.05	-0.22 to 0.13	-0.16	-0.34 to 0.02	-0.40	-0.65 to -0.15
Stroop card 2[b]	.35	Sequential	-0.20	-0.40 to -0.01	-0.23	-0.43 to -0.02	-0.28	-0.52 to -0.03
		Exemestane	-0.17	-0.35 to 0.01	-0.02	-0.20 to 0.16	-0.22	-0.42 to -0.01
		Control	0.00	-0.17 to 0.17	0.05	-0.12 to 0.22	-0.09	-0.30 to 0.13
Trail Making Test A[c]	.006	Sequential	-0.21	-0.40 to -0.03	-0.26	-0.46 to -0.07	-0.36	-0.61 to -0.11
		Exemestane	-0.30	-0.47 to -0.14	0.02	-0.15 to 0.19	-0.34	-0.55 to -0.13
		Control	-0.05	-0.21 to 0.11	0.06	-0.10 to 0.22	0.03	-0.20 to 0.26
Executive functioning[c]	.005	Sequential	-0.22	-0.41 to -0.04	-0.33	-0.52 to -0.14	-0.55	-0.77 to -0.32
		Exemestane	-0.35	-0.51 to -0.18	-0.21	-0.38 to -0.04	-0.41	-0.60 to -0.21
		Control	-0.05	-0.20 to 0.11	0.06	-0.10 to 0.22	0.04	-0.16 to 0.24
Stroop card 3[c]	.07	Sequential	-0.24	-0.44 to -0.03	-0.35	-0.56 to -0.13	-0.53	-0.80 to -0.25
		Exemestane	-0.36	-0.55 to -0.18	-0.22	-0.41 to -0.03	-0.17	-0.40 to 0.06
		Control	-0.05	-0.22 to 0.13	0.07	-0.11 to 0.24	0.04	-0.20 to 0.29

Table continues

Sequential group and exemestane vs. control									
T0-T1				T0-T2					
Mean difference in change over time	95% CI	p	ES[a]	Mean difference in change over time	95% CI	p	ES[a]	AIC	BIC
-0.21	-0.38 to -0.05	.01	-0.26	-0.44	-0.72 to -0.15	.003	-0.34		
-0.06	-0.21 to 0.09	.42	-0.08	-0.13	-0.39 to 0.12	.31	-0.12	1375.19	1498.98
-0.13	-0.37 to 0.11	.28	-0.11	-0.52	-0.87 to -0.18	.003	-0.36		
0.04	-0.18 to 0.26	.73	0.03	-0.28	-0.60 to 0.04	.09	-0.21	1832.08	1979.97
-0.28	-0.50 to -0.06	.01	-0.25	-0.32	-0.66 to 0.01	.06	-0.22		
-0.11	-0.31 to 0.09	.27	-0.11	-0.12	-0.43 to 0.20	.47	-0.09	1767.07	1914.96
-0.22	-0.51 to 0.06	.12	-0.16	-0.36	-0.85 to 0.12	.14	-0.15		
-0.04	-0.30 to 0.22	.74	-0.03	-0.04	-0.47 to 0.39	.85	-0.02	2061.17	2184.96
0.03	-0.18 to 0.23	.80	0.03	-0.19	-0.51 to 0.13	.24	-0.14		
0.13	-0.06 to 0.32	.17	0.14	-0.22	-0.51 to 0.07	.14	-0.18	1724.22	1848.96
-0.01	-0.24 to 0.22	.94	-0.01	-0.16	-0.51 to 0.20	.38	-0.10		
0.06	-0.15 to 0.28	.57	0.06	-0.14	-0.46 to 0.19	.40	-0.10	1860.15	1985.00
0.08	-0.17 to 0.32	.53	0.06	-0.19	-0.55 to 0.18	.31	-0.12		
0.21	-0.02 to 0.43	.07	0.18	-0.27	-0.60 to 0.06	.10	-0.20	1902.66	2027.40
-0.05	-0.19 to 0.09	.50	-0.07	-0.05	-0.29 to 0.18	.67	-0.05		
0.15	0.02 to 0.28	.02	0.22	0.02	-0.19 to 0.24	.82	0.03	1261.08	1385.96
0.11	-0.13 to 0.35	.36	0.09	-0.03	-0.40 to 0.34	.89	-0.02		
0.14	-0.07 to 0.36	.19	0.13	0.05	-0.29 to 0.39	.78	0.03	1893.65	2041.67
-0.08	-0.26 to 0.10	.38	-0.08	0.01	-0.27 to 0.30	.92	0.01		
0.10	-0.06 to 0.27	.23	0.12	0.04	-0.22 to 0.31	.74	0.04	1645.12	1770.11
-0.17	-0.38 to 0.05	.13	-0.15	-0.23	-0.57 to 0.10	.17	-0.16		
0.21	0.01 to 0.41	.04	0.21	-0.12	-0.43 to 0.18	.43	-0.09	1724.82	1872.70
-0.22	-0.37 to -0.06	.007	-0.27	-0.41	-0.68 to -0.15	.002	-0.38		
0.03	-0.11 to 0.17	.69	0.04	-0.15	-0.39 to 0.09	.22	-0.15	1475.47	1623.23
-0.22	-0.45 to 0.00	.05	-0.19	-0.38	-0.75 to 0.00	.047	-0.17		
0.03	-0.18 to 0.24	.79	0.03	0.10	-0.24 to 0.45	.55	0.05	1843.34	1986.69

Table continues

Trail Making Test B[e]	.005	Sequential	-0.19	-0.45 to 0.08	-0.30	-0.53 to -0.07	-0.35	-0.68 to -0.01
		Exemestane	-0.47	-0.71 to -0.23	-0.18	-0.39 to 0.02	-0.52	-0.79 to -0.26
		Control	-0.03	-0.43 to 0.37	0.05	-0.14 to 0.24	0.02	-0.27 to 0.31
Motor speed[f]	.31	Sequential	-0.02	-0.18 to 0.14	-0.01	-0.17 to 0.16	-0.22	-0.41 to -0.02
		Exemestane	-0.05	-0.20 to 0.09	0.01	-0.14 to 0.16	-0.22	-0.39 to -0.06
		Control	-0.08	-0.21 to 0.06	0.02	-0.12 to 0.16	-0.07	-0.23 to 0.10
FePsy Fingertapping dominant hand[f]	.24	Sequential	-0.01	-0.17 to 0.15	0.03	-0.13 to 0.20	-0.18	-0.38 to 0.02
		Exemestane	0.02	-0.13 to 0.16	0.06	-0.09 to 0.21	-0.17	-0.34 to 0.00
		Control	-0.08	-0.22 to 0.06	0.01	-0.13 to 0.15	-0.03	-0.20 to 0.14
FePsy Fingertapping non-dominant hand[f]	.44	Sequential	-0.02	-0.19 to 0.16	-0.03	-0.22 to 0.15	-0.24	-0.45 to -0.02
		Exemestane	-0.13	-0.28 to 0.03	-0.04	-0.20 to 0.12	-0.29	-0.47 to -0.11
		Control	-0.07	-0.22 to 0.08	0.03	-0.12 to 0.18	-0.12	-0.30 to 0.06
Verbal fluency[g]	.76	Sequential	-0.39	-0.53 to -0.25	-0.39	-0.54 to -0.25	-0.39	-0.57 to -0.21
		Exemestane	-0.42	-0.54 to -0.29	-0.37	-0.50 to -0.24	-0.48	-0.63 to -0.34
		Control	-0.09	-0.21 to 0.03	-0.07	-0.19 to 0.05	-0.16	-0.30 to -0.01
Letterfluency[b]	.78	Sequential	-0.37	-0.55 to -0.19	-0.35	-0.54 to -0.16	-0.22	-0.44 to 0.01
		Exemestane	-0.41	-0.57 to -0.25	-0.31	-0.47 to -0.14	-0.27	-0.46 to -0.08
		Control	-0.07	-0.22 to 0.09	0.05	-0.11 to 0.20	0.07	-0.12 to 0.27
Category fluency -Animals[b]	.91	Sequential	-0.35	-0.54 to -0.16	-0.41	-0.61 to -0.21	-0.51	-0.77 to -0.25
		Exemestane	-0.34	-0.50 to -0.17	-0.36	-0.54 to -0.19	-0.52	-0.73 to -0.31
		Control	-0.06	-0.22 to 0.10	-0.06	-0.23 to 0.10	-0.11	-0.34 to 0.12
Category fluency - Professions[e]	.22	Sequential	-0.43	-0.64 to -0.22	-0.38	-0.57 to -0.20	-0.26	-0.53 to 0.00
		Exemestane	-0.52	-0.71 to -0.33	-0.40	-0.56 to -0.23	-0.53	-0.75 to -0.32
		Control	-0.13	-0.44 to 0.20	-0.15	-0.30 to 0.00	-0.17	-0.41 to 0.07
Reaction speed[h]	.523	Sequential	-0.27	-0.47 to -0.06	-0.40	-0.62 to -0.18	-0.27	-0.55 to 0.00
		Exemestane	-0.22	-0.40 to -0.03	-0.16	-0.36 to 0.03	-0.25	-0.49 to -0.02
		Control	0.02	-0.16 to 0.19	0.01	-0.17 to 0.19	-0.05	-0.30 to 0.20
FePsy Reaction Time dominant hand[h]	.04	Sequential	-0.17	-0.39 to 0.06	-0.47	-0.71 to -0.23	-0.22	-0.53 to 0.08
		Exemestane	-0.23	-0.44 to -0.03	-0.11	-0.32 to 0.10	-0.22	-0.48 to 0.04
		Control	0.02	-0.17 to 0.22	0.12	-0.08 to 0.31	0.07	-0.22 to 0.35
FePsy Reaction Times non-dominant hand[h]	.84	Sequential	-0.33	-0.56 to -0.11	-0.31	-0.55 to -0.08	-0.29	-0.58 to 0.01
		Exemestane	-0.20	-0.40 to 0.00	-0.22	-0.43 to -0.01	-0.28	-0.54 to -0.03
		Control	0.01	-0.18 to 0.19	-0.10	-0.30 to 0.09	-0.12	-0.39 to 0.16
Working memory[g]	.21	Sequential	-0.33	-0.51 to -0.15	-0.16	-0.35 to 0.03	-0.43	-0.70 to -0.15
		Exemestane	-0.30	-0.46 to -0.13	-0.20	-0.37 to -0.03	-0.45	-0.67 to -0.24
		Control	-0.06	-0.21 to 0.10	-0.01	-0.16 to 0.15	0.08	-0.14 to 0.31

Abbreviations: T0, baseline; T1; 1 year follow-up; T2, 5 year follow-up; CI, confidence interval; ES, effect size; RAVLT, Rey Auditory Verbal Learning Test; VAT, Visual Association Test; WMS, Wechsler Memory Scale; IR, immediate recall; DR, delayed recall; AIC, Akaike information criterion; BIC, Bayesian information criterion.

Table continues

-0.20	-0.41 to 0.00	.049	-0.20	-0.49	-0.84 to -0.13	.007	-0.24		
0.02	-0.17 to 0.20	.87	0.02	-0.43	-0.75 to -0.10	.01	-0.23	1800.92	1976.44
-0.08	-0.22 to 0.06	.25	-0.11	-0.21	-0.43 to 0.02	.07	-0.22		
-0.03	-0.16 to 0.10	.62	-0.05	-0.18	-0.39 to 0.02	.08	-0.22	1292.93	1417.82
-0.04	-0.19 to 0.11	.58	-0.05	-0.22	-0.46 to 0.03	.08	-0.21		
-0.04	-0.18 to 0.09	.52	-0.06	-0.23	-0.45 to -0.01	.04	-0.26	1357.80	1482.73
-0.12	-0.27 to 0.03	.13	-0.15	-0.17	-0.41 to 0.07	.16	-0.18		
-0.02	-0.16 to 0.12	.80	-0.03	-0.12	-0.34 to 0.10	.28	-0.14	1437.64	1562.53
-0.02	-0.15 to 0.10	.70	-0.04	0.06	-0.12 to 0.25	.49	0.09		
0.03	-0.09 to 0.14	.66	0.04	0.00	-0.16 to 0.16	>.99	0.00	1140.02	1242.04
-0.09	-0.26 to 0.08	.29	-0.11	0.02	-0.24 to 0.27	.90	0.02		
-0.01	-0.16 to 0.15	.94	-0.01	0.00	-0.23 to 0.24	.97	0.00	1490.45	1615.23
-0.06	-0.30 to 0.18	.62	-0.05	-0.11	-0.45 to 0.23	.52	-0.07		
-0.03	-0.25 to 0.19	.81	-0.02	-0.14	-0.45 to 0.17	.38	-0.10	1793.08	1917.96
0.10	-0.12 to 0.32	.38	-0.02	0.16	-0.21 to 0.53	.38	0.13		
0.18	-0.02 to 0.39	.08	-0.06	-0.01	-0.35 to 0.32	.93	-0.13	1684.64	1860.20
-0.12	-0.35 to 0.11	.305	-0.10	0.05	-0.30 to 0.40	.76	0.04		
0.06	-0.15 to 0.27	.573	0.06	0.02	-0.30 to 0.35	.89	0.02	1813.94	1938.06
-0.39	-0.68 to -0.10	.008	-0.28	-0.10	-0.50 to 0.30	.62	-0.07		
0.03	-0.24 to 0.30	.83	0.02	-0.03	-0.40 to 0.34	.87	-0.02	2026.97	2151.20
0.13	-0.14 to 0.40	.33	0.10	0.17	-0.23 to 0.57	.41	0.11		
0.09	-0.16 to 0.33	.49	0.07	0.04	-0.34 to 0.41	.84	0.03	1961.66	2085.78
0.12	-0.14 to 0.38	.36	0.09	-0.23	-0.62 to 0.16	.24	-0.13		
0.04	-0.20 to 0.29	.72	0.03	-0.30	-0.65 to 0.05	.10	-0.19	1833.39	1935.06

a Effect sizes: .20 small effect, .50 moderate effect, .80 large effect.
b Adjusted for age, IQ, FACT-ES.
c Adjusted for age, IQ, HSCL, FACT-ES.
d Adjusted for age, HSCL, FACT-ES.
e Adjusted for pattern missing, age, IQ, HSCL, FACT-ES.
f Adjusted for age, IQ, HSCL.
g Adjusted for age, IQ.
h Adjusted for IQ, HSCL, FACT-ES.

Supplementary Table 3

Results of the per-protocol analyses of the Sequential, the Exemestane and the control group

			Adjusted mean Z-scores					
			T0		T1		T2	
Cognitive domain/test	P-value overall group-by-time interaction	Group	Mean	95% CI	Mean	95% CI	Mean	95% CI
Verbal memory[b]	.046	Sequential	0.29	-0.26 to 0.84	0.18	-0.40 to 0.75	-0.61	-1.43 to 0.22
		Exemestane	-0.37	-0.85 to 0.12	0.12	-0.38 to 0.62	-0.25	-0.91 to 0.42
		Control	-0.22	-0.64 to 0.19	0.35	-0.07 to 0.77	-0.08	-0.67 to 0.52
RAVLT IR[b]	.049	Sequential	0.12	-0.10 to 0.34	0.07	-0.16 to 0.31	-0.23	-0.55 to 0.09
		Exemestane	-0.20	-0.40 to -0.01	-0.03	-0.24 to 0.17	-0.18	-0.45 to 0.10
		Control	-0.08	-0.24 to 0.09	0.05	-0.12 to 0.22	0.17	-0.07 to 0.41
RAVLT DR[c]	.08	Sequential	0.15	-0.08 to 0.37	0.06	-0.17 to 0.30	-0.08	-0.39 to 0.24
		Exemestane	-0.03	-0.22 to 0.17	0.08	-0.13 to 0.28	0.03	-0.23 to 0.29
		Control	-0.08	-0.25 to 0.09	0.16	-0.01 to 0.33	0.08	-0.15 to 0.31
VAT[d]	.61	Sequential	0.04	-0.23 to 0.31	0.04	-0.25 to 0.32	-0.16	-0.59 to 0.27
		Exemestane	-0.13	-0.37 to 0.11	0.09	-0.15 to 0.34	-0.25	-0.58 to 0.09
		Control	-0.03	-0.23 to 0.18	0.18	-0.03 to 0.38	-0.12	-0.40 to 0.17
Visual memory[c]	.39	Sequential	-0.67	-1.10 to -0.24	-0.66	-1.12 to -0.21	-0.58	-1.18 to 0.03
		Exemestane	-0.64	-1.02 to -0.26	-0.39	-0.78 to 0.01	-0.52	-1.03 to -0.02
		Control	-0.14	-0.46 to 0.19	-0.14	-0.47 to 0.19	0.23	-0.21 to 0.68
WMS Visual Memory IR[c]	.72	Sequential	-0.30	-0.53 to -0.06	-0.39	-0.64 to -0.14	-0.40	-0.74 to -0.06
		Exemestane	-0.31	-0.51 to -0.10	-0.26	-0.48 to -0.05	-0.16	-0.44 to 0.12
		Control	-0.09	-0.26 to 0.09	-0.12	-0.30 to 0.06	0.03	-0.22 to 0.27
WMS Visual Memory DR[c]	.09	Sequential	-0.39	-0.62 to -0.16	-0.33	-0.57 to -0.08	-0.17	-0.50 to 0.16
		Exemestane	-0.33	-0.53 to -0.13	-0.12	-0.32 to 0.09	-0.34	-0.61 to -0.06
		Control	-0.05	-0.22 to 0.12	-0.03	-0.20 to 0.15	0.20	-0.04 to 0.43
Information processing speed[c]	.002	Sequential	-0.21	-0.38 to -0.04	-0.27	-0.45 to -0.09	-0.37	-0.60 to -0.14
		Exemestane	-0.25	-0.40 to -0.10	-0.05	-0.20 to 0.11	-0.44	-0.63 to -0.25
		Control	-0.02	-0.14 to 0.11	0.00	-0.13 to 0.13	-0.18	-0.35 to -0.01
Stroop card 1[b]	.18	Sequential	-0.20	-0.43 to 0.03	-0.27	-0.52 to -0.03	-0.46	-0.79 to -0.14
		Exemestane	-0.25	-0.45 to -0.04	-0.15	-0.36 to 0.07	-0.65	-0.93 to -0.37
		Control	-0.02	-0.20 to 0.15	-0.14	-0.32 to 0.03	-0.41	-0.65 to -0.16
Stroop card 2[c]	.11	Sequential	-0.17	-0.39 to 0.05	-0.19	-0.42 to 0.04	-0.18	-0.47 to 0.11
		Exemestane	-0.19	-0.39 to 0.01	0.01	-0.19 to 0.21	-0.27	-0.52 to -0.03
		Control	0.01	-0.16 to 0.17	0.07	-0.10 to 0.24	-0.10	-0.31 to 0.11
Trail Making Test A[b]	.006	Sequential	-0.23	-0.44 to -0.01	-0.28	-0.51 to -0.06	-0.40	-0.71 to -0.09
		Exemestane	-0.32	-0.51 to -0.13	0.01	-0.18 to 0.20	-0.46	-0.72 to -0.20
		Control	-0.04	-0.19 to 0.12	0.08	-0.08 to 0.24	0.06	-0.17 to 0.29
Executive functioning[b]	.02	Sequential	-0.20	-0.41 to 0.00	-0.32	-0.53 to -0.10	-0.50	-0.77 to -0.23
		Exemestane	-0.35	-0.53 to -0.17	-0.22	-0.40 to -0.03	-0.34	-0.57 to -0.12
		Control	-0.03	-0.18 to 0.12	0.08	-0.07 to 0.24	0.06	-0.13 to 0.26
Stroop card 3[b]	.04	Sequential	-0.18	-0.41 to 0.05	-0.32	-0.57 to -0.08	-0.47	-0.79 to -0.15
		Exemestane	-0.39	-0.59 to -0.18	-0.23	-0.44 to -0.01	-0.10	-0.37 to 0.18
		Control	-0.03	-0.21 to 0.14	0.09	-0.08 to 0.27	0.07	-0.17 to 0.30
Trail Making Test B[b]	.02	Sequential	-0.23	-0.47 to 0.01	-0.31	-0.56 to -0.06	-0.47	-0.79 to -0.15
		Exemestane	-0.32	-0.53 to -0.11	-0.21	-0.43 to 0.00	-0.63	-0.90 to -0.37
		Control	-0.03	-0.21 to 0.15	0.07	-0.11 to 0.25	0.05	-0.18 to 0.29

Table continues

Sequential group and exemestane vs. control									
T0-T1				T0-T2					
Mean difference in change over time	95% CI	p	ES[a]	Mean difference in change over time	95% CI	p	ES[a]	AIC	BIC
-0.68	-1.18 to -0.18	.008	-0.29	-1.04	-1.97 to -0.11	.03	-0.28		
-0.08	-0.54 to 0.37	.72	-0.04	-0.03	-0.86 to 0.81	.95	-0.01	2541.72	2683.58
-0.17	-0.43 to 0.09	.19	-0.14	-0.59	-0.99 to -0.19	.004	-0.39		
0.04	-0.20 to 0.28	.73	0.04	-0.22	-0.58 to 0.14	.23	-0.16	1580.09	1722.95
-0.32	-0.56 to -0.08	.009	-0.28	-0.38	-0.76 to -0.01	.045	-0.26		
-0.13	-0.35 to 0.08	.23	-0.13	-0.11	-0.45 to 0.23	.54	-0.08	1537.96	1658.63
-0.20	-0.48 to 0.07	.15	-0.16	-0.11	-0.62 to 0.39	.66	-0.05		
0.02	-0.23 to 0.27	.88	0.02	-0.03	-0.46 to 0.41	.91	-0.01	1718.41	1816.00
0.02	-0.41 to 0.44	.94	0.01	-0.27	-0.98 to 0.43	.44	-0.10		
0.26	-0.13 to 0.64	.20	0.14	-0.26	-0.88 to 0.37	.42	-0.11	2350.82	2471.49
-0.06	-0.30 to 0.19	.65	-0.05	-0.22	-0.62 to 0.19	.29	-0.13		
0.08	-0.14 to 0.30	.49	0.08	0.03	-0.32 to 0.39	.85	0.02	1588.30	1708.93
0.04	-0.21 to 0.29	.770	0.03	-0.03	-0.41 to 0.35	.88	-0.02		
0.19	-0.04 to 0.42	.10	0.18	-0.25	-0.60 to 0.09	.14	-0.21	1572.20	1692.70
-0.08	-0.23 to 0.07	.31	-0.11	0.01	-0.26 to 0.27	.97	0.01		
0.19	0.05 to 0.32	.007	0.29	-0.03	-0.27 to 0.21	.81	-0.03	1077.25	1197.92
0.04	-0.19 to 0.28	.72	0.04	0.12	-0.29 to 0.52	.57	0.07		
0.22	0.00 to 0.44	.049	0.21	-0.03	-0.39 to 0.34	.89	-0.02	1583.39	1726.41
-0.09	-0.27 to 0.10	.37	-0.09	0.10	-0.22 to 0.42	.56	0.07		
0.14	-0.03 to 0.31	.12	0.17	0.03	-0.26 to 0.32	.85	0.02	1398.11	1518.90
-0.17	-0.41 to 0.07	.16	-0.16	-0.27	-0.65 to 0.11	.17	-0.17		
0.22	0.00 to 0.43	.050	0.22	-0.24	-0.58 to 0.11	.18	-0.17	1502.23	1645.10
-0.23	-0.39 to -0.07	.006	-0.30	-0.39	-0.69 to -0.09	.01	-0.25		
0.02	-0.13 to 0.17	.78	0.03	-0.09	-0.35 to 0.18	.52	-0.06	1241.25	1379.51
-0.26	-0.49 to -0.03	.02	-0.24	-0.39	-0.79 to 0.01	.06	-0.18		
0.04	-0.17 to 0.25	.72	0.04	0.19	-0.17 to 0.56	.30	0.10	1549.38	1687.88
-0.18	-0.39 to 0.03	.09	-0.20	-0.32	-0.66 to 0.02	.06	-0.32		
0.01	-0.18 to 0.20	.95	0.01	-0.40	-0.70 to -0.10	.01	-0.45	1500.90	1643.66

Table continues

Motor speed[e]	.15	Sequential	0.04	-0.15 to 0.21	0.02	-0.17 to 0.20	-0.24	-0.47 to -0.01
		Exemestane	-0.09	-0.25 to 0.07	0.04	-0.13 to 0.20	-0.15	-0.34 to 0.05
		Control	-0.06	-0.19 to 0.08	0.04	-0.10 to 0.17	-0.04	-0.20 to 0.12
FePsy Fingertapping dominant hand[f]	.32	Sequential	0.03	-0.15 to 0.21	0.05	-0.13 to 0.24	-0.19	-0.43 to 0.05
		Exemestane	-0.02	-0.17 to 0.15	0.09	-0.07 to 0.25	-0.08	-0.28 to 0.12
		Control	-0.06	-0.19 to 0.08	0.03	-0.11 to 0.16	-0.01	-0.18 to 0.15
FePsy Fingertapping non-dominant hand[e]	.12	Sequential	0.05	-0.15 to 0.24	-0.01	-0.21 to 0.20	-0.27	-0.53 to -0.02
		Exemestane	-0.16	-0.34 to 0.01	-0.02	-0.20 to 0.16	-0.21	-0.42 to 0.01
		Control	-0.05	-0.20 to 0.09	0.05	-0.10 to 0.19	-0.07	-0.25 to 0.11
Verbal fluency[c]	.82	Sequential	-1.16	-1.63 to -0.68	-1.11	-1.61 to -0.61	-1.08	-1.73 to -0.43
		Exemestane	-1.31	-1.73 to -0.88	-1.15	-1.58 to -0.71	-1.51	-2.04 to -0.98
		Control	-0.14	-0.50 to 0.22	-0.10	-0.47 to 0.26	-0.26	-0.72 to 0.21
Letterfluency[c]	.75	Sequential	-0.36	-0.56 to -0.16	-0.34	-0.55 to -0.13	-0.18	-0.45 to 0.09
		Exemestane	-0.45	-0.63 to -0.27	-0.33	-0.51 to -0.15	-0.30	-0.52 to -0.08
		Control	-0.05	-0.20 to 0.10	0.07	-0.08 to 0.22	0.14	-0.06 to 0.34
Category fluency - Animals[c]	.99	Sequential	-0.39	-0.60 to -0.18	-0.41	-0.63 to -0.18	-0.55	-0.87 to -0.23
		Exemestane	-0.36	-0.55 to -0.18	-0.42	-0.61 to -0.22	-0.53	-0.80 to -0.27
		Control	-0.04	-0.20 to 0.12	-0.05	-0.21 to 0.12	-0.12	-0.35 to 0.11
Category fluency - Professions[f]	.34	Sequential	-0.42	-0.64 to -0.19	-0.34	-0.56 to -0.13	-0.17	-0.50 to 0.16
		Exemestane	-0.48	-0.67 to -0.28	-0.40	-0.58 to -0.22	-0.57	-0.84 to -0.30
		Control	-0.10	-0.41 to 0.22	-0.14	-0.28 to 0.01	-0.13	-0.37 to 0.10
Reaction speed[c]	>.99	Sequential	0.28	0.02 to 0.55	0.27	-0.02 to 0.55	0.08	-0.29 to 0.45
		Exemestane	0.22	-0.01 to 0.46	0.17	-0.07 to 0.42	-0.02	-0.33 to 0.29
		Control	0.03	-0.17 to 0.22	-0.05	-0.25 to 0.15	-0.21	-0.49 to 0.08
FePsy Reaction Time dominant hand[b]	.84	Sequential	0.17	-0.12 to 0.47	0.25	-0.07 to 0.57	-0.15	-0.59 to 0.29
		Exemestane	0.24	-0.02 to 0.50	0.20	-0.07 to 0.48	-0.09	-0.46 to 0.29
		Control	0.05	-0.17 to 0.27	-0.10	-0.32 to 0.13	-0.35	-0.69 to 0.00
FePsy Reaction Times non-dominant hand[g]	.71	Sequential	0.59	0.31 to 0.87	0.42	0.11 to 0.72	0.41	0.01 to 0.82
		Exemestane	0.34	0.09 to 0.60	0.23	-0.04 to 0.49	0.12	-0.22 to 0.47
		Control	-0.03	-0.24 to 0.19	0.00	-0.22 to 0.22	-0.05	-0.34 to 0.25
Working memory[h]	.13	Sequential	-0.41	-0.61 to -0.20	-0.20	-0.42 to 0.02	-0.56	-0.89 to -0.22
		Exemestane	-0.30	-0.48 to -0.12	-0.19	-0.38 to 0.00	-0.45	-0.72 to -0.19
		Control	-0.03	-0.18 to 0.13	0.05	-0.11 to 0.20	0.15	-0.07 to 0.38

Abbreviations: T0, baseline; T1; 1 year follow-up; T2, 5 year follow-up; CI, confidence interval; ES, effect size; RAVLT, Rey Auditory Verbal Learning Test; IR, immediate recall; DR, delayed recall; VAT, Visual Association Test; WMS, Wechsler Memory Scale; AIC, Akaike information criterion; BIC, Bayesian information criterion.

a Effect sizes: .20 small effect, .50 moderate effect, .80 large effect.

b Adjusted for age, IQ, HSCL, FACT-ES.

c Adjusted for age, IQ, FACT-ES.

d Adjusted for age, FACT-ES.

e Adjusted for age, IQ, HSCL.

f Adjusted for pattern missing, age, IQ, HSCL, FACT-ES.

g Adjusted for FACT-ES.

h Adjusted for IQ.

Table continues

-0.11 0.03	-0.25 to 0.04 -0.10 to 0.16	.14 .63	-0.15 0.05	-0.29 -0.07	-0.54 to -0.04 -0.30 to 0.15	.02 .52	-0.22 -0.06	1089.61	1205.85
-0.06 0.02	-0.22 to 0.09 -0.12 to 0.16	.42 .79	-0.08 0.03	-0.27 -0.11	-0.53 to 0.00 -0.35 to 0.13	.05 .36	-0.18 -0.09	1139.89	1256.17
-0.16 0.04	-0.32 to 0.01 -0.10 to 0.19	.06 .55	-0.19 0.06	-0.31 -0.03	-0.59 to -0.02 -0.28 to 0.22	.03 .81	-0.20 -0.02	1224.24	1340.48
0.01 0.12	-0.42 to 0.44 -0.27 to 0.51	.97 .54	0.00 0.07	0.19 -0.09	-0.50 to 0.89 -0.70 to 0.52	.59 .76	0.07 -0.04	2409.68	2530.31
-0.10 0.00	-0.28 to 0.08 -0.16 to 0.16	.27 .96	-0.12 0.00	-0.01 -0.04	-0.31 to 0.28 -0.30 to 0.23	.93 .78	-0.01 -0.04	1283.05	1403.59
-0.01 -0.05	-0.27 to 0.25 -0.28 to 0.19	.93 .70	-0.01 -0.04	-0.08 -0.09	-0.47 to 0.31 -0.44 to 0.26	.70 .62	-0.05 -0.06	1551.23	1671.94
0.12 0.17	-0.12 to 0.37 -0.05 to 0.40	.32 .13	0.11 0.17	0.17 -0.05	-0.19 to 0.54 -0.38 to 0.27	.36 .74	0.13 -0.05	1445.47	1619.53
0.06 0.02	-0.24 to 0.36 -0.25 to 0.30	.70 .86	0.04 0.02	0.03 -0.01	-0.41 to 0.47 -0.41 to 0.39	.89 .95	0.01 -0.01	1727.05	1842.60
0.23 0.11	-0.15 to 0.60 -0.24 to 0.45	.24 .55	0.12 0.06	0.07 0.07	-0.48 to 0.63 -0.44 to 0.58	.80 .79	0.03 0.03	1944.56	2082.28
-0.20 -0.14	-0.53 to 0.13 -0.44 to 0.16	.23 .35	-0.13 -0.10	-0.16 -0.20	-0.63 to 0.31 -0.63 to 0.23	.50 .35	-0.07 -0.09	1846.77	1917.87
0.13 0.04	-0.15 to 0.40 -0.22 to 0.29	.36 .77	0.10 0.03	-0.33 -0.33	-0.77 to 0.10 -0.72 to 0.05	.13 .09	-0.17 -0.21	1578.58	1654.56

Supplementary Table 4

Results of the intention-to-treat analyses of the Sequential and the Exemestane group

			Adjusted mean Z-scores					
			T0		T1		T2	
Cognitive domain/test	P-value overall group-by-time interaction	Group	Mean	95% CI	Mean	95% CI	Mean	95% CI
Verbal memory[b]	.050	Sequential	-0.01	-0.18 to 0.16	-0.04	-0.21 to 0.14	-0.42	-0.66 to -0.18
		Exemestane	-0.20	-0.35 to -0.05	-0.08	-0.24 to 0.07	-0.29	-0.48 to -0.10
RAVLT IR[c]	.25	Sequential	-0.06	-0.26 to 0.14	-0.07	-0.28 to 0.14	-0.41	-0.67 to -0.14
		Exemestane	-0.26	-0.44 to -0.08	-0.12	-0.30 to 0.07	-0.37	-0.59 to -0.15
RAVLT DR[d]	.23	Sequential	0.03	-0.17 to 0.22	-0.04	-0.25 to 0.16	-0.18	-0.44 to 0.07
		Exemestane	-0.11	-0.29 to 0.07	-0.01	-0.19 to 0.17	-0.12	-0.33 to 0.10
VAT[e]	.37	Sequential	0.05	-0.21 to 0.31	0.02	-0.25 to 0.30	-0.40	-0.81 to 0.00
		Exemestane	-0.18	-0.42 to 0.05	-0.02	-0.26 to 0.23	-0.30	-0.62 to 0.02
Visual memory[d]	.50	Sequential	-0.50	-0.70 to -0.29	-0.46	-0.67 to -0.25	-0.51	-0.77 to -0.24
		Exemestane	-0.47	-0.65 to -0.28	-0.34	-0.53 to -0.15	-0.52	-0.74 to -0.30
WMS Visual Memory IR[d]	.87	Sequential	-0.45	-0.66 to -0.24	-0.49	-0.71 to -0.26	-0.51	-0.80 to -0.23
		Exemestane	-0.44	-0.63 to -0.25	-0.41	-0.61 to -0.21	-0.48	-0.72 to -0.25
WMS Visual Memory DR[d]	.36	Sequential	-0.58	-0.80 to -0.35	-0.46	-0.70 to -0.23	-0.50	-0.80 to -0.21
		Exemestane	-0.50	-0.70 to -0.29	-0.26	-0.47 to -0.05	-0.52	-0.77 to -0.28
Information processing speed[d]	.03	Sequential	-0.34	-0.49 to -0.19	-0.38	-0.54 to -0.22	-0.54	-0.73 to -0.34
		Exemestane	-0.35	-0.49 to -0.21	-0.19	-0.33 to -0.05	-0.48	-0.64 to -0.31
Stroop card 1[d]	.93	Sequential	-0.41	-0.62 to -0.20	-0.43	-0.65 to -0.21	-0.77	-1.06 to -0.49
		Exemestane	-0.30	-0.49 to -0.11	-0.28	-0.48 to -0.09	-0.59	-0.83 to -0.35
Stroop card 2[d]	.16	Sequential	-0.30	-0.49 to -0.11	-0.34	-0.53 to -0.14	-0.36	-0.60 to -0.11
		Exemestane	-0.30	-0.47 to -0.12	-0.16	-0.34 to 0.02	-0.34	-0.54 to -0.13
Trail Making Test A[d]	.003	Sequential	-0.30	-0.50 to -0.11	-0.38	-0.58 to -0.17	-0.47	-0.74 to -0.20
		Exemestane	-0.44	-0.62 to -0.26	-0.12	-0.30 to 0.07	-0.49	-0.72 to -0.27
Executive functioning[d]	.04	Sequential	-0.34	-0.55 to -0.14	-0.46	-0.67 to -0.24	-0.67	-0.93 to -0.40
		Exemestane	-0.52	-0.70 to -0.33	-0.39	-0.58 to -0.19	-0.59	-0.81 to -0.37
Stroop card 3[d]	.06	Sequential	-0.36	-0.59 to -0.13	-0.46	-0.70 to -0.22	-0.62	-0.93 to -0.30
		Exemestane	-0.53	-0.74 to -0.32	-0.39	-0.60 to -0.18	-0.32	-0.58 to -0.06
Trail Making Test B[f]	.10	Sequential	-0.31	-0.60 to -0.01	-0.43	-0.69 to -0.17	-0.51	-0.90 to -0.12
		Exemestane	-0.65	-0.92 to -0.38	-0.35	-0.59 to -0.12	-0.73	-1.03 to -0.42
Motor speed[c]	.87	Sequential	-0.11	-0.27 to 0.06	-0.08	-0.25 to 0.09	-0.33	-0.53 to -0.12
		Exemestane	-0.11	-0.26 to 0.03	-0.05	-0.20 to 0.10	-0.28	-0.45 to -0.11
FePsy Fingertapping dominant hand[g]	>.99	Sequential	-0.10	-0.26 to 0.06	-0.05	-0.21 to 0.11	-0.27	-0.47 to -0.06
		Exemestane	-0.06	-0.21 to 0.08	-0.01	-0.16 to 0.13	-0.24	-0.41 to -0.07
FePsy Fingertapping non-dominant hand[g]	.54	Sequential	-0.11	-0.29 to 0.07	-0.12	-0.30 to 0.06	-0.33	-0.55 to -0.11
		Exemestane	-0.19	-0.35 to -0.03	-0.11	-0.27 to 0.06	-0.36	-0.55 to -0.18
Verbal fluency[b]	.40	Sequential	-0.48	-0.61 to -0.34	-0.48	-0.62 to -0.34	-0.49	-0.67 to -0.32
		Exemestane	-0.51	-0.63 to -0.38	-0.46	-0.58 to -0.33	-0.59	-0.73 to -0.44
Letterfluency[b]	.66	Sequential	-0.49	-0.67 to -0.32	-0.48	-0.66 to -0.30	-0.39	-0.61 to -0.17
		Exemestane	-0.51	-0.67 to -0.35	-0.42	-0.58 to -0.26	-0.39	-0.57 to -0.21

Table continues

Sequential group vs. exemestane									
To-T1				To-T2					
Mean difference in change over time	95% CI	p	ES[a]	Mean difference in change over time	95% CI	p	ES[a]	AIC	BIC
-0.15	-0.31 to 0.01	.07	-0.23	-0.33	-0.61 to -0.05	.02	-0.33	905.35	975.35
-0.16	-0.40 to 0.09	.20	-0.17	-0.24	-0.55 to 0.08	.14	-0.26	1144.52	1247.75
-0.17	-0.38 to 0.04	.12	-0.20	-0.20	-0.51 to 0.10	.19	-0.23	1086.51	1173.31
-0.19	-0.52 to 0.14	.25	-0.15	-0.33	-0.87 to 0.20	.22	-0.17	1366.12	1452.09
-0.10	-0.31 to 0.11	.35	-0.12	0.04	-0.25 to 0.34	.78	0.04	1091.35	1178.02
-0.06	-0.30 to 0.18	.60	-0.07	-0.02	-0.36 to 0.32	.91	-0.02	1181.97	1268.77
-0.12	-0.38 to 0.13	.33	-0.13	0.10	-0.25 to 0.45	.57	0.09	1215.52	1302.19
-0.20	-0.35 to -0.05	.01	-0.33	-0.06	-0.30 to 0.17	.59	-0.08	823.23	910.03
-0.03	-0.28 to 0.22	.82	-0.03	-0.07	-0.43 to 0.29	.72	-0.05	1194.85	1281.69
-0.17	-0.36 to 0.02	.08	-0.22	-0.02	-0.30 to 0.27	.90	-0.02	1033.89	1120.74
-0.39	-0.63 to -0.16	.001	-0.42	-0.11	-0.46 to 0.23	.53	-0.10	1142.78	1229.58
-0.24	-0.43 to -0.05	.01	-0.32	-0.25	-0.55 to 0.05	.10	-0.27	1063.40	1150.11
-0.24	-0.51 to 0.03	.08	-0.22	-0.48	-0.89 to -0.06	.03	-0.24	1274.26	1356.97
-0.23	-0.46 to 0.00	.050	-0.25	-0.04	-0.42 to 0.33	.82	-0.03	1219.47	1322.69
-0.03	-0.18 to 0.11	.65	-0.06	-0.06	-0.29 to 0.18	.64	-0.05	809.02	911.81
0.00	-0.15 to 0.15	.97	0.01	0.02	-0.23 to 0.26	.90	-0.08	834.95	921.66
-0.09	-0.25 to 0.07	.27	-0.14	-0.05	-0.29 to 0.20	.71	-0.06	913.06	999.73
-0.05	-0.18 to 0.08	.45	-0.10	0.07	-0.12 to 0.26	.48	0.11	704.76	775.13
-0.07	-0.24 to 0.09	.37	-0.11	-0.02	-0.25 to 0.21	.85	-0.03	928.05	998.54

Table continues

Category fluency - Animals[d]	.90	Sequential	-0.43	-0.61 to -0.26	-0.50	-0.68 to -0.31	-0.59	-0.84 to -0.34
		Exemestane	-0.43	-0.59 to -0.27	-0.46	-0.63 to -0.30	-0.62	-0.83 to -0.42
Category fluency - Professions[d]	.16	Sequential	-0.54	-0.69 to -0.38	-0.51	-0.68 to -0.34	-0.44	-0.67 to -0.21
		Exemestane	-0.57	-0.72 to -0.43	-0.46	-0.61 to -0.30	-0.66	-0.85 to -0.47
Reaction speed[h]	.24	Sequential	-0.33	-0.55 to -0.11	-0.43	-0.66 to -0.20	-0.32	-0.61 to -0.02
		Exemestane	-0.25	-0.46 to -0.05	-0.17	-0.38 to 0.03	-0.27	-0.52 to -0.02
FePsy Reaction Time dominant hand[h]	.03	Sequential	-0.22	-0.46 to 0.02	-0.50	-0.76 to -0.25	-0.25	-0.57 to 0.08
		Exemestane	-0.26	-0.48 to -0.04	-0.12	-0.34 to 0.11	-0.23	-0.51 to 0.04
FePsy Reaction Times non-dominant hand[h]	.83	Sequential	-0.41	-0.64 to -0.17	-0.35	-0.59 to -0.10	-0.32	-0.64 to 0.00
		Exemestane	-0.25	-0.46 to -0.03	-0.23	-0.45 to -0.01	-0.29	-0.56 to -0.02
Working memory[b]	.87	Sequential	-0.43	-0.61 to -0.25	-0.27	-0.46 to -0.08	-0.56	-0.84 to -0.28
		Exemestane	-0.39	-0.55 to -0.23	-0.30	-0.47 to -0.13	-0.58	-0.79 to -0.36

Abbreviations: T0, baseline; T1; 1 year follow-up; T2, 5 year follow-up; CI, confidence interval; ES, effect size; RAVLT, Rey Auditory Verbal Learning Test; IR, immediate recall; DR, delayed recall; VAT, Visual Association Test; WMS, Wechsler Memory Scale; AIC, Akaike information criterion; BIC, Bayesian information criterion.

a Effect sizes: .20 small effect, .50 moderate effect, .80 large effect.
b Adjusted for age, IQ.
c Adjusted for age, IQ, HSCL, FACT-ES.
d Adjusted for age, IQ, FACT-ES.
e Adjusted for age, HSCL, FACT-ES.
f Adjusted for pattern missing, age, IQ, FACT-ES.
g Adjusted for age, IQ, HSCL.
h Adjusted for IQ, HSCL, FACT-ES.

Table continues

	-0.03	-0.27 to 0.20	.78	-0.04	0.03	-0.29 to 0.36	.83	0.03	1070.50	1157.21
	-0.09	-0.31 to 0.12	.41	-0.10	0.18	-0.14 to 0.51	.27	0.11	993.84	1076.46
	-0.18	-0.44 to 0.07	.15	-0.19	0.03	-0.33 to 0.39	.87	0.03	1178.79	1265.04
	-0.43	-0.76 to -0.10	.01	-0.35	-0.06	-0.46 to 0.34	.78	-0.05	1326.08	1412.47
	0.04	-0.24 to 0.33	.76	0.04	0.13	-0.29 to 0.55	.55	0.10	1266.74	1352.99
	0.07	-0.20 to 0.34	.61	0.06	0.06	-0.34 to 0.45	.78	0.04	1133.66	1203.85

Supplementary Table 5

Results of the per-protocol analyses of the Sequential and the Exemestane group

			Adjusted mean Z-scores					
			T0		T1		T2	
Cognitive domain/test	P-value overall group-by-time interaction	Group	Mean	95% CI	Mean	95% CI	Mean	95% CI
Verbal memory[b]	.03	Sequential	0.03	-0.54 to 0.59	-0.09	-0.68 to 0.49	-1.16	-2.02 to -0.30
		Exemestane	-0.66	-1.17 to -0.14	-0.19	-0.72 to 0.34	-0.65	-1.33 to 0.03
RAVLT IR[c]	.12	Sequential	0.01	-0.21 to 0.23	-0.06	-0.29 to 0.18	-0.38	-0.71 to -0.05
		Exemestane	-0.33	-0.53 to -0.14	-0.17	-0.38 to 0.03	-0.35	-0.62 to -0.07
RAVLT DR[d]	.19	Sequential	0.07	-0.16 to 0.30	-0.02	-0.27 to 0.22	-0.18	-0.49 to 0.13
		Exemestane	-0.13	-0.34 to 0.08	-0.02	-0.24 to 0.20	-0.14	-0.40 to 0.13
VAT[d]	.40	Sequential	0.01	-0.28 to 0.30	0.00	-0.31 to 0.31	-0.19	-0.67 to 0.29
		Exemestane	-0.19	-0.46 to 0.07	0.02	-0.25 to 0.29	-0.32	-0.70 to 0.05
Visual memory[b]	.45	Sequential	-0.94	-1.38 to -0.50	-0.90	-1.36 to -0.44	-0.93	-1.52 to -0.33
		Exemestane	-0.97	-1.37 to -0.57	-0.72	-1.13 to -0.31	-1.08	-1.60 to -0.57
WMS Visual Memory IR[e]	.55	Sequential	-0.42	-0.66 to -0.18	-0.49	-0.74 to -0.24	-0.54	-0.89 to -0.20
		Exemestane	-0.46	-0.68 to -0.24	-0.41	-0.64 to -0.19	-0.38	-0.67 to -0.10
WMS Visual Memory DR[e]	.06	Sequential	-0.53	-0.77 to -0.29	-0.46	-0.71 to -0.20	-0.33	-0.65 to 0.00
		Exemestane	-0.49	-0.70 to -0.27	-0.28	-0.50 to -0.06	-0.60	-0.87 to -0.33
Information processing speed[e]	<.001	Sequential	-0.32	-0.49 to -0.15	-0.39	-0.57 to -0.21	-0.47	-0.70 to -0.23
		Exemestane	-0.38	-0.53 to -0.22	-0.18	-0.34 to -0.02	-0.59	-0.79 to -0.39
Stroop card 1[e]	.21	Sequential	-0.34	-0.57 to -0.11	-0.42	-0.67 to -0.18	-0.62	-0.95 to -0.29
		Exemestane	-0.33	-0.54 to -0.12	-0.24	-0.46 to -0.03	-0.70	-0.98 to -0.42
Stroop card 2[e]	.04	Sequential	-0.28	-0.49 to -0.06	-0.30	-0.52 to -0.08	-0.27	-0.56 to 0.02
		Exemestane	-0.34	-0.53 to -0.15	-0.15	-0.34 to 0.05	-0.41	-0.65 to -0.16
Trail Making Test A[d]	.005	Sequential	-0.34	-0.58 to -0.10	-0.43	-0.68 to -0.17	-0.49	-0.84 to -0.13
		Exemestane	-0.48	-0.70 to -0.26	-0.15	-0.38 to 0.07	-0.64	-0.94 to -0.35
Executive functioning[e]	.08	Sequential	-0.35	-0.59 to -0.11	-0.46	-0.71 to -0.21	-0.65	-0.97 to -0.32
		Exemestane	-0.55	-0.76 to -0.33	-0.42	-0.64 to -0.20	-0.54	-0.81 to -0.27
Stroop card 3[b]	.050	Sequential	-0.33	-0.59 to -0.07	-0.45	-0.73 to -0.17	-0.59	-0.97 to -0.20
		Exemestane	-0.60	-0.83 to -0.36	-0.43	-0.68 to -0.18	-0.26	-0.59 to 0.07
Trail Making Test B[e]	.20	Sequential	-0.37	-0.65 to -0.10	-0.47	-0.75 to -0.18	-0.66	-1.02 to -0.30
		Exemestane	-0.50	-0.75 to -0.25	-0.42	-0.68 to -0.17	-0.90	-1.20 to -0.60
Motor speed[c]	.17	Sequential	-0.07	-0.25 to 0.10	-0.08	-0.27 to 0.10	-0.34	-0.58 to -0.11
		Exemestane	-0.17	-0.33 to -0.01	-0.05	-0.21 to 0.12	-0.23	-0.42 to -0.03
FePsy Fingertapping dominant hand[c]	.47	Sequential	-0.07	-0.25 to 0.10	-0.04	-0.22 to 0.14	-0.29	-0.52 to -0.06
		Exemestane	-0.11	-0.26 to 0.05	0.01	-0.15 to 0.16	-0.17	-0.37 to 0.02
FePsy Fingertapping non-dominant hand[f]	.10	Sequential	-0.08	-0.28 to 0.12	-0.13	-0.34 to 0.08	-0.38	-0.65 to -0.12
		Exemestane	-0.24	-0.43 to -0.06	-0.10	-0.29 to 0.08	-0.29	-0.51 to -0.06
Verbal fluency[e]	.43	Sequential	-1.46	-1.91 to -1.01	-1.43	-1.90 to -0.95	-1.39	-2.04 to -0.74
		Exemestane	-1.68	-2.09 to -1.27	-1.53	-1.95 to -1.11	-1.96	-2.50 to -1.43
Letterfluency[g]	.49	Sequential	-0.50	-0.69 to -0.30	-0.50	-0.70 to -0.29	-0.37	-0.63 to -0.11
		Exemestane	-0.57	-0.74 to -0.39	-0.46	-0.64 to -0.28	-0.45	-0.67 to -0.24

Table continues

Sequential group vs. exemestane									
To-T1				To-T2					
Mean difference in change over time	95% CI	p	ES[a]	Mean difference in change over time	95% CI	p	ES[a]	AIC	BIC
-0.59	-1.13 to -0.05	.03	-0.32	-1.19	-2.20 to -0.19	.02	-0.42	1448.41	1544.35
-0.22	-0.50 to 0.05	.11	-0.23	-0.38	-0.79 to 0.03	.07	-0.36	903.49	984.86
-0.20	-0.45 to 0.04	.11	-0.24	-0.24	-0.59 to 0.11	.17	-0.31	859.85	925.53
-0.23	-0.57 to 0.11	.19	-0.20	-0.07	-0.67 to 0.54	.83	-0.04	1037.77	1102.97
-0.22	-0.65 to 0.21	.32	-0.15	0.12	-0.55 to 0.80	.72	0.07	1303.54	1400.06
-0.11	-0.37 to 0.14	.38	-0.13	-0.20	-0.61 to 0.21	.34	-0.16	908.40	989.54
-0.13	-0.39 to 0.13	.33	-0.15	0.32	-0.04 to 0.68	.08	0.37	882.07	963.02
-0.27	-0.44 to -0.10	.002	-0.46	0.06	-0.23 to 0.35	.69	0.05	638.50	715.77
-0.17	-0.42 to 0.08	.18	-0.19	0.10	-0.33 to 0.52	.66	0.07	889.67	970.87
-0.22	-0.43 to -0.01	.04	-0.30	0.08	-0.28 to 0.43	.67	0.09	791.12	872.31
-0.41	-0.69 to -0.14	.003	-0.44	0.02	-0.43 to 0.46	.94	0.01	922.64	988.32
-0.24	-0.45 to -0.03	.03	-0.32	-0.30	-0.68 to 0.07	.11	-0.22	834.05	911.21
-0.29	-0.59 to 0.01	.06	-0.28	-0.59	-1.10 to -0.08	.02	-0.29	993.04	1085.70
-0.17	-0.42 to 0.08	.18	-0.21	0.12	-0.26 to 0.50	.54	0.15	931.29	1012.31
-0.13	-0.29 to 0.02	.09	-0.25	-0.21	-0.48 to 0.06	.12	-0.21	610.00	687.16
-0.08	-0.24 to 0.07	.30	-0.15	-0.15	-0.43 to 0.12	.28	-0.14	616.09	693.30
-0.19	-0.36 to -0.01	.04	-0.29	-0.26	-0.57 to 0.05	.10	-0.22	707.65	769.38
-0.12	-0.59 to 0.35	.62	-0.07	0.35	-0.43 to 1.14	.37	0.15	1338.65	1419.72
-0.10	-0.28 to 0.08	.28	-0.16	0.01	-0.27 to 0.30	.93	0.02	721.58	787.45

Table continues

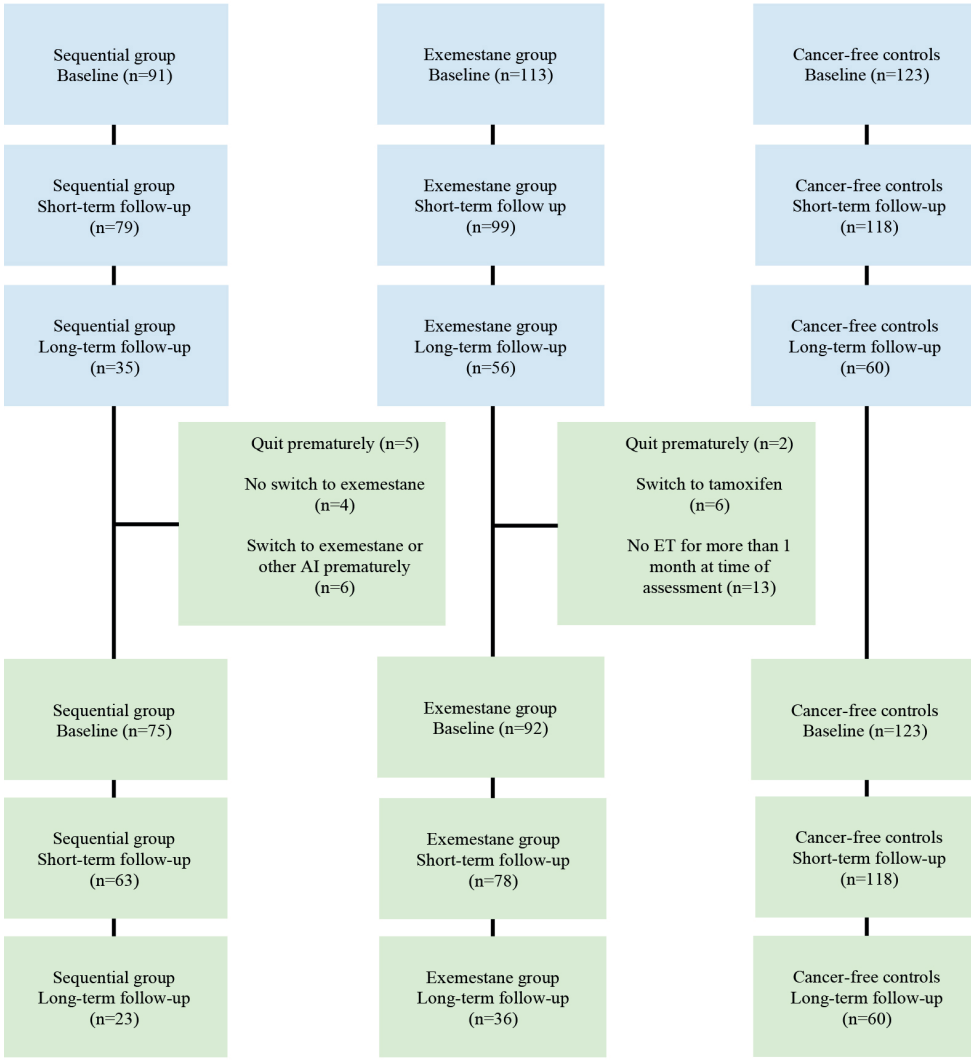
Category fluency - Animals[e]	.96	Sequential	-0.48	-0.67 to -0.29	-0.51	-0.71 to -0.30	-0.66	-0.97 to -0.36
		Exemestane	-0.49	-0.66 to -0.32	-0.55	-0.73 to -0.37	-0.70	-0.95 to -0.45
Category fluency - Professions[e]	.44	Sequential	-0.55	-0.73 to -0.38	-0.51	-0.70 to -0.32	-0.45	-0.73 to -0.16
		Exemestane	-0.61	-0.77 to -0.45	-0.51	-0.68 to -0.35	-0.69	-0.93 to -0.46
Reaction speed[e]	.92	Sequential	0.45	0.14 to 0.75	0.40	0.07 to 0.73	0.22	-0.15 to 0.60
		Exemestane	0.32	0.05 to 0.60	0.24	-0.05 to 0.52	0.03	-0.29 to 0.35
FePsy Reaction Time dominant hand[b]	.89	Sequential	0.36	0.02 to 0.69	0.39	0.03 to 0.75	-0.05	-0.56 to 0.47
		Exemestane	0.35	0.04 to 0.66	0.27	-0.05 to 0.59	-0.04	-0.49 to 0.41
FePsy Reaction Times non-dominant hand[h]	.60	Sequential	0.53	0.21 to 0.86	0.37	0.02 to 0.72	0.51	0.11 to 0.91
		Exemestane	0.31	0.02 to 0.60	0.22	-0.08 to 0.53	0.12	-0.23 to 0.47
Working memory[i]	.77	Sequential	-0.51	-0.71 to -0.31	-0.31	-0.53 to -0.09	-0.69	-1.02 to -0.35
		Exemestane	-0.39	-0.57 to -0.21	-0.28	-0.47 to -0.09	-0.54	-0.80 to -0.27

Abbreviations: T0, baseline; T1; 1 year follow-up; T2, 5 year follow-up; CI, confidence interval; ES, effect size; RAVLT, Rey Auditory Verbal Learning Test; IR, immediate recall; DR, delayed recall; VAT, Visual Association Test; WMS, Wechsler Memory Scale; AIC, Akaike information criteria; BIC, Bayesian information criteria.

a Effect sizes: .20 small effect, .50 moderate effect, .80 large effect.
b Adjusted for age, IQ, HSCL, FACT-ES.
c Adjusted for age, IQ, HSCL.
d Adjusted for age, FACT-ES.
e Adjusted for age, IQ, FACT-ES.
f Adjusted for age, HSCL.
g Adjusted for age, IQ.
h Adjusted for age.
i Adjusted for IQ.

Table continues

0.04	-0.23 to 0.31	.78	0.04	0.03	-0.38 to 0.43	.89	0.02	826.66	907.73
-0.05	-0.30 to 0.20	.70	-0.06	0.19	-0.19 to 0.57	.32	0.18	769.99	851.07
0.04	-0.37 to 0.45	.86	0.03	0.07	-0.28 to 0.42	.69	0.11	1053.04	1133.69
0.11	-0.37 to 0.59	.65	0.06	-0.01	-0.69 to 0.67	.97	-0.01	1198.88	1291.06
-0.07	-0.50 to 0.36	.74	-0.05	0.17	-0.22 to 0.56	.39	0.24	1122.75	1173.01
0.09	-0.21 to 0.39	.56	0.08	-0.03	-0.50 to 0.45	.91	-0.02	875.68	925.94



Supplementary Figure 1. Flow-chart inclusion per-protocol analyses

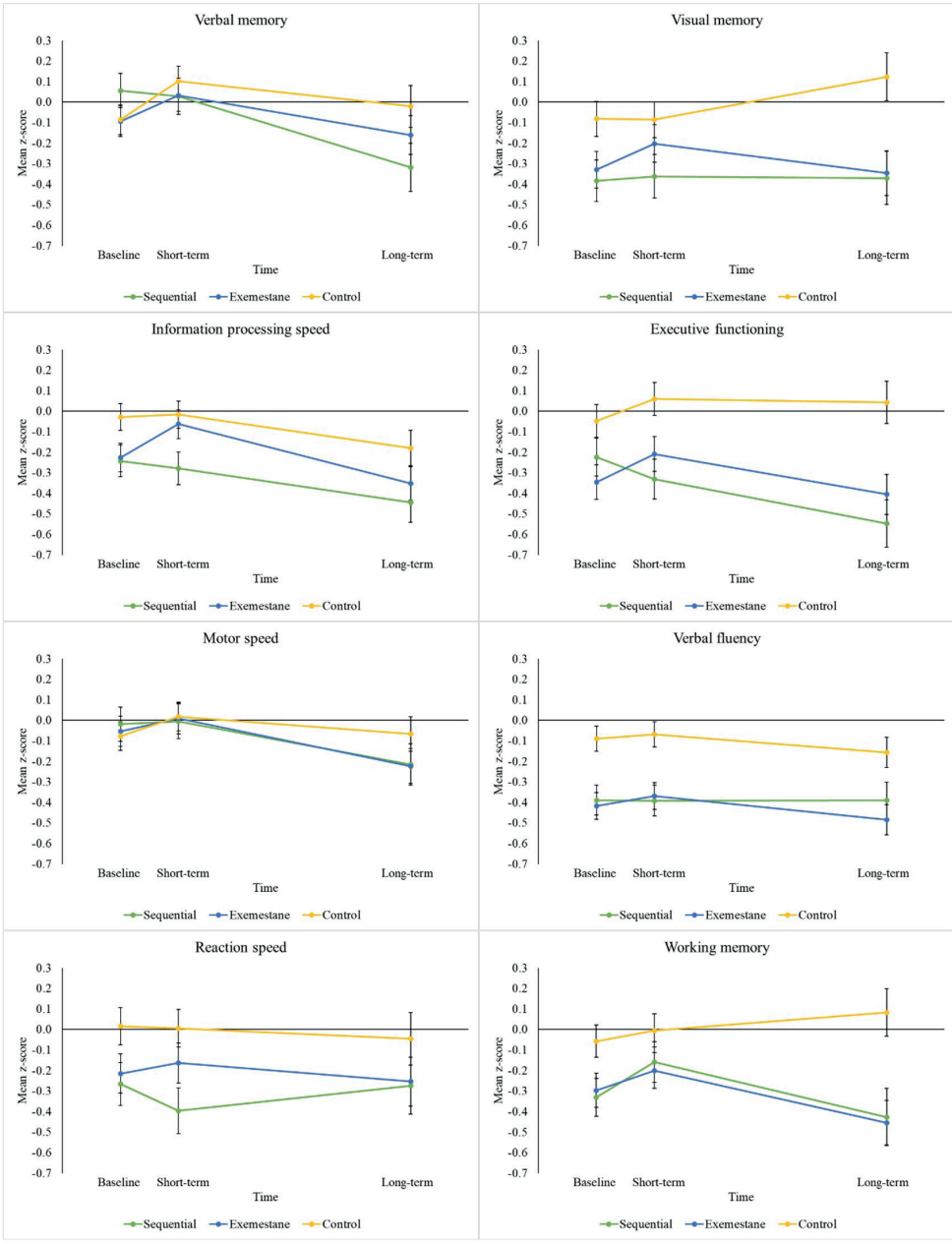
Abbreviations: ET, endocrine therapy; T1, 1 year follow-up assessment; T2, 5 year follow-up assessment.

	Sequential and Exemestane vs. control group				Sequential vs. Exemestane group			
	ITT		PP		ITT		PP	
	ST	LT	ST	LT	ST	LT	ST	LT
Domains								
Processing Speed	Exe vs. con							
Executive Functioning								
Motor Speed		Exe vs. con						
Tests								
RAVLT IR		Exe vs. con						
RAVLT DR		Seq vs. con		Seq vs. con				
Stroop Card 2								
TMT A	Exe vs. con	Seq vs. con						
Stroop Card 3		Seq vs. con	Seq vs. con					
TMT B				Seq vs. con				
Fepsy Fingertapping (dom)		Exe vs. con						
Fepsy Fingertapping (ndom)								

Supplementary Figure 2. Summary of results that differed before and after correction

Abbreviations: ITT, Intention-to-treat analyses; PP, Per-protocol analyses; ST, short-term; LT, long-term; Seq, Sequential group; Exe, Exemestane group; RAVLT, Rey Auditory Verbal Learning Test; IR, Immediate recall; DR, Delayed recall; TMT, Trail Making Test; ndom, non-dominant hand; dom, dominant hand.

NOTE. Blue is effect that arose after correction. Yellow is effect that disappeared after correction. Green remained the same. All other domains, and all other tests have been omitted from this table, as results were identical before and after correction.



Supplementary Figure 3. Adjusted standardized change over time between patients and controls for each cognitive domain

Mean z-scores are depicted per cognitive domain.

Chapter 7:
Long-term effects of premenopausal risk-reducing salpingo-oophorectomy on cognition in women with high familial risk of ovarian cancer: A cross-sectional study

Lara Terra*, Philippe R. Lee Meeuw Kjoe*, Joost A. Agelink Van Rentergem, Maarten J. Beekman, Bernadette A.M. Heemskerk-Gerritsen, Marc van Beurden, Jeanine E. Roeters van Lennep, Helena C. van Doorn, Joanna A. De Hullu, Marian J.E. Mourits, Eleonora B.L. van Dorst, Constantijne H. Mom, Brigitte F.M. Slangen, Katja N. Gaarenstroom, Lizet E. van der Kolk, J. Margriet Collée, Marijke R. Wevers, Margreet G.E.M. Ausems, Klaartje van Engelen, Irma van de Beek, Lieke P.V. Berger, Christi J. van Asperen, Encarna B. Gomez Garcia, Angela H.E.M. Maas, Maartje J. Hooning, Elsken van der Wall, Flora E. van Leeuwen**, Sanne B. Schagen**.

*Shared first authorship.
**Shared last authorship.

BJOG: An International Journal of Obstetrics and Gynaecology. 2023 [published online ahead of print January 30, 2023].

Abstract

Objective

To examine the effect of a premenopausal risk-reducing salpingo-oophorectomy (RRSO) in women at increased risk of ovarian cancer on objective and subjective cognition at least 10 years after RRSO.

Design

A cross-sectional study with prospective follow-up, nested in a nationwide cohort.

Setting

Multicenter in the Netherlands.

Population or Sample

641 women (66% *BRCA1/2* pathogenic variant carriers) who underwent either a premenopausal RRSO ≤ age 45 (n=436) or a postmenopausal RRSO ≥ age 54 (n=205). All participants were older than 55 years at recruitment.

Methods

Participants completed an online cognitive test battery and a questionnaire on subjective cognition. We used multivariable regression analyses, adjusting for age, education, breast cancer, hormone replacement therapy, cardiovascular risk factors and depression.

Main Outcome Measures

The influence of RRSO on objective and subjective cognition of women with a premenopausal RRSO compared with women with a postmenopausal RRSO.

Results

After adjustment, women with a premenopausal RRSO (mean time since RRSO 18.2 years) performed similarly on objective cognitive tests as women with a postmenopausal RRSO (mean time since RRSO 11.9 years). However, they more frequently reported problems with reasoning (odds ratio (OR) 1.8 (95% confidence interval (95%CI) 1.1-3.1)) and multitasking (OR 1.9 (95%CI 1.1-3.4)) than women with a postmenopausal RRSO. This difference between groups disappeared in an analysis restricted to women of comparable ages (60-70 years).

Conclusions

Reassuringly, approximately 18 years after RRSO, we found no association between premenopausal RRSO and objective cognition.

Introduction

Women carrying a *BRCA1/2* germline pathogenic variant (*BRCA1/2pv*) have an increased life-time risk of breast and/or ovarian cancer.[1] To prevent ovarian cancer, guidelines recommend risk-reducing salpingo-oophorectomy (RRSO) after completing childbirth, for *BRCA1pv* carriers between ages 35-40, and for *BRCA2pv* carriers between ages 40-45.[2-3] Although this preventive surgery reduces the risk of ovarian cancer, it induces immediate menopause long before natural menopause occurs at an average age of 51 years.[4] This may adversely impact cognition, as studies have shown neuroprotective effects of estrogens.[5] The uptake of RRSO in *BRCA1/2pv* carriers is high (86-91%).[6] Therefore, knowledge about long-term health consequences of RRSO, such as cognitive effects, is important for effective counseling.

Several studies reported conflicting findings on the association between (bilateral) oophorectomy before natural menopause and cognition later in life. Some studies showed that an oophorectomy is associated with cognitive impairment and long-term increased risk of dementia.[7-10] This association is argued to be dependent on age at oophorectomy. A recent study observed that women with a bilateral oophorectomy before age 46 had increased risk of cognitive impairment 20 years later compared with women without bilateral oophorectomy.[9] However, in *BRCA1/2pv* carriers who underwent RRSO before age 46 compared with carriers with RRSO at later ages, such an association between age at oophorectomy and cognitive decline was not found after approximately nine years.[9,11]

It is difficult to compare results across studies as these had several methodological limitations.[12-13] First, studies did not always account for the indication for oophorectomy in the exposure group (e.g., ovarian cancer or a benign ovarian condition), nor were data provided on whether a bilateral oophorectomy, unilateral oophorectomy or a hysterectomy

without oophorectomy was performed. Second, studies use different comparison groups (e.g., general population, premenopausal women without oophorectomy). Third, adjustment for confounding (e.g., hormone replacement therapy (HRT)) was done inconsistently.[14] Also, many studies did not adjust for comorbidities like depression and hypertension which are known risk factors for dementia,[15] nor for cancer, which can also negatively affect cognition.[16] Last, no studies examined the relationship between RRSO and subjective cognition. As such, while there is some evidence for an association between premenopausal RRSO and cognitive impairment at later ages, many questions remain unanswered.

Our aim was to examine the effect of a premenopausal RRSO on long-term cognitive functioning. We compared objective and subjective cognitive functioning at least 10 years after RRSO between women at high familial risk of ovarian cancer with a premenopausal RRSO ≤ age 45 and women with a postmenopausal RRSO ≥ age 54.

Methods

Participants

Participants were women participating in the HARMOny study (ClinicalTrials.gov NCT03835793); a multicenter cross-sectional study assessing long-term effects of RRSO on (sub)clinical cardiovascular disease, bone health, cognition and quality of life by comparing women with a premenopausal RRSO and women with a postmenopausal RRSO.

Study design and procedures have been described previously.[17] Briefly, between 2018 and 2021, we invited 1271 women from the well-established HEBON cohort study of Dutch families with a high familial risk of breast/ovarian cancer to participate in the HARMOny study (see also Supplementary Methods I).[18] Women were eligible if they had a RRSO ≤ age 45 and were currently aged ≥55 years, resulting in at least 10 years

since RRSO. We compared them with women currently aged ≥55 years with a RRSO ≥ age 54, aiming to frequency-match on age. We chose the cut-off of 45 years based on clinical recommendations for *BRCApv* carriers.

Exclusion criteria were ovarian carcinoma, metastatic disease, early-onset dementia, and insufficient understanding of the Dutch language. These criteria were checked via medical files and questionnaires. A history of breast cancer was no exclusion criterion. This study was approved by the Institutional Review Board of the Netherlands Cancer Institute. All participants provided written informed consent.

Study assessments

Participants completed, after inclusion to the HARMOny study, an online questionnaire on sociodemographic data, general health, cancer-specific outcomes, cardiovascular health, and medical treatments, including cancer treatment and use of HRT.

To study objective cognition, participants completed the Amsterdam Cognition Scan (ACS): an online neuropsychological assessment that is completed by computer at home without supervision.[19] The ACS consists of seven cognitive tests, based on traditional neuropsychological tests covering four domains: verbal memory, attention, executive functioning, and processing speed (10 outcome variables; see Table S1). The ACS takes approximately one hour to complete, and is tailored to detect cognitive dysfunction associated with cancer (treatment). The ACS has shown concurrent validity and test-retest reliability, and normative data have been collected.[20]

Subjective cognition was assessed by the Medical Outcomes Study cognitive functioning scale (MOS-cog), measuring the frequency of self-reported cognitive problems in daily life.[21] The MOS-cog is a validated questionnaire consisting of six questions on reasoning, memory, attention,

concentration, multitasking and thinking speed (score: 0-5). Higher scores indicate a higher frequency of cognitive complaints.

Outliers objective cognition

Outliers were removed in two steps. For each cognitive outcome, non-legitimate test scores (i.e., impossible scores due to computer malfunction, non-adherence to test instructions or low motivation) were removed using pre-defined cut-offs based on test instructions and clinical expertise. The cut-offs per test are depicted in Table S8. After removing impossible scores, we excluded outliers on the speed-based tests with the median absolute deviation (MAD) method.[22] The MAD was applied separately in participants below age 60 and above to adequately take age into account.

Age-corrected cognitive domain scores

Based on the demographically adjusted normative dataset of the ACS (women aged ≥55 only, n=157),[19] we calculated age-corrected z-scores for test outcomes. Performance on the online versions of the Trail Making Test A (TMT-A) and TMT-B, Visual Reaction Time, Tower of London and Grooved Pegboard were reversed (*z*-score times -1), so that higher scores indicated better performance.

We calculated cognitive domain scores by averaging the means of the age-corrected z-scores of the subtests belonging to the same domain. This led to four cognitive domain scores: verbal memory, executive functioning, processing speed and attention (Table S1).

Statistical analyses

With a two-sided α of .05, with 200 women in the study we have 94% power to detect a difference in z-score of 0.5 between the two groups. With 750 women in the study, we have 98% power to detect a difference in z-score of 0.3, and 78% power to detect a difference in z-score of 0.2.

Patient characteristics were compared using independent samples t-tests, $\chi 2$ tests or Fisher exact tests. To examine the effect of a premenopausal RRSO on long-term objective cognition, age-corrected z-scores for the four cognitive domains were compared between the premenopausal and postmenopausal RRSO groups, and between the RRSO groups and the normative population using independent samples t-tests. In addition, we performed multivariable linear regression analyses with the four cognitive domains as dependent variables and RRSO (premenopausal or postmenopausal) as the independent variable, adjusting for current age, level of education, breast cancer, HRT, depression and cardiovascular risk factors (i.e., hypertension). BMI, diabetes and smoking were not confounding factors, and were therefore omitted from our analyses.

To examine the effect of a premenopausal RRSO on long-term subjective cognition, independent sample t-tests were used to compare the six subjective cognition outcomes (score 0-5) between the premenopausal and postmenopausal RRSO groups. In addition, we performed ordered logistic regression analyses, adjusting for current age, level of education, breast cancer, HRT, depression and cardiovascular risk factors, yielding odds ratios (OR).

We performed several subgroup analyses for objective and subjective cognition. Because of the clinical recommendations for *BRCA1/2pv* carriers, we compared cognition between women with RRSO ≤ age 40 and between ages 41 and 45. Furthermore, because of the age difference between the pre- and postmenopausal RRSO groups, we compared cognition in participants

whose ages overlapped (60-70 years). Additionally, to account for potential confounding effects of breast cancer history and HRT, we performed stratified analyses according to breast cancer history; within the premenopausal RRSO group we performed stratified analyses according to HRT-use. Due to collinearity between premenopausal or postmenopausal RRSO and 'time since RRSO', we did not add 'time since RRSO' to the regression analyses. Subsequently, we performed sensitivity analyses with 'time since RRSO' as a continuous variable.

Proportion cognitively affected

Participants who scored ≥1 standard deviation (SD) below the age-corrected normative mean on two tests from different cognitive domains were classified as cognitively affected. This criterion is used in studies in the field of cancer (treatment) and cognition.[23-25] The proportion of cognitively affected participants was compared between the premenopausal and postmenopausal RRSO groups using a χ^2 test, and compared with the expected proportion (30%) of cognitively affected based on the probability curves provided by Ingraham and Aiken.[26] To analyse the effect of a premenopausal RRSO on the proportion cognitively affected, we used multivariable logistic regression analyses, adjusting for current age, breast cancer history, HRT, depression, education, and cardiovascular disease.

For all statistical analyses, Stata, version 15.0 (StataCorp) was used. An alpha of .05 was used as the criterion for statistical significance.

Results

Participation

In total, 758 women who met the eligibility criteria gave written informed consent (response rate 59.6%) of whom 505 were in the premenopausal RRSO group (RRSO ≤ age 45) and 253 in the postmenopausal RRSO group (RRSO ≥ age 54) (Figure S1).

The ACS was completed by 641 women; 436 with a premenopausal RRSO and 205 with a postmenopausal RRSO. Forty-one participants were unable to complete the ACS because they did not have access to a laptop or computer. Compared to ACS completers, noncompleters were older (mean age: 63.1 years (SD 6.1) versus 67.2 years (SD 6.1), p <.01). They did not report cognitive complaints more frequently (p-values ranged from .08 to .31).

Participant characteristics

The premenopausal and postmenopausal RRSO groups differed on several characterstics, partly due to the inclusion criteria (Table 1). The premenopausal RRSO group was younger at study participation than the postmenopausal RRSO group (59.9 versus 70.1 years; p<.001) and had a longer time since RRSO (18.1 versus 11.7 years; p<.001). Compared to the postmenopausal RRSO group, the premenopausal RRSO group more often completed at least middle level education (65.6% versus 51.2%; p<.001), more often received chemotherapy for breast cancer (76.9% versus 53.5%; p=.01), (had) more often used HRT (24.8% versus 10.3%; p<.001) and less often had cardiovascular risk factors (35.8% versus 56.2%; p=.001). There was no difference between the groups in occurrence of breast cancer and treatments other than chemotherapy, depression, or *BRCA1/2pv* status (67% in the premenopausal RRSO group and 64.6% in the postmenopausal RRSO group). The premenopausal RRSO group reported more frequent computer use than the postmenopausal RRSO group.

Table 1

Demographics of study participants that completed the online Amsterdam Cognition Scan

	Premenopausal RRSO (n = 436)	Postmenopausal RRSO (n = 205)	p-value[V]
Age (mean, sd)	59.9 (3.5)	70.1 (4.4)	<.001
Age at RRSO (mean, sd)	41.8 (2.7)	58.5 (3.7)	<.001
Time since RRSO (mean, sd)	18.1 (4.2)	11.7 (3.0)	<.001
Age at menopause (mean, sd)	41.8 (2.8)	50.3 (5.0)	<.001
Pathogenic genetic variants[I]			.41
BRCA1 germline mutation	209 (47.9%)	64 (31.2%)	
BRCA2 germline mutation	83 (19.0%)	68 (33.2%)	
Non carrier BRCA1/2	144 (33.0%)	73 (35.6%)	
Breast cancer (yes)	247 (56.7%)	127 (62.0%)	.19
Treatment of breast cancer			
Surgery	243 (97.9%)	120 (94.5%)	.93
Chemotherapy	190 (76.9%)	68 (53.5%)	.01
Radiotherapy	155 (62.5%)	65 (53.3%)	.09
Endocrine therapy	93 (37.5%)	35 (28.7%)	.09
HRT use			<.001
Current user	23 (5.3%)	2 (1.0%)	
Past user	85 (19.5%)	19 (9.3%)	
Never user	287 (65.8%)	165 (80.5%)	
Unknown	41 (9.4%)	37 (18.0%)	
HRT duration in years (mean (sd))	1.6 (5.1)	.4 (1.7)	.13
Type HRT			
Tibolone	25 (23.1%)	2 (9.5%)	
Estradiol/progestogen	19 (17.6%)	0 (0.0%)	
Estradiol only	6 (5.6%)	1 (4.8%)	
Unknown	58 (53.7%)	18 (85.7%)	
BMI (mean, sd)	26.2 (5.0)	25.7 (4.4)	.29
Educational level			<.001
Primary school/lower level high school	111 (25.5%)	77 (37.6%)	
Middle level high school	143 (32.8%)	38 (18.5%)	
Advanced vocational/university	143 (32.8%)	67 (32.7%)	
Missing	39 (8.9%)	23 (11.2%)	
Hours of computer use per week	12.8 (12.3)	6.7 (7.6)	<.001
Depression (yes)	63 (14.4%)	20 (9.8%)	.10
Cardiovascular risk[II]	156 (35.8%)	123 (56.2%)	<.001
Cardiovascular disease[III]	78 (19.1%)	51 (25.4%)	.08
Hysterectomy (Yes)[IV]	60 (13.8%)	42 (20.5%)	.002

I: All participants have a high familial risk of ovarian cancer. All women are tested for germline mutations, not all have a BRCA1/2 mutation. Among the established non-carriers there are for example CHEK2 mutation carriers.
II: Cardiovascular risk factors: hypertension, hypercholesterolemia and/or type 2 diabetes.
III: Cardiovascular disease: myocardial infarction, angina, heart failure, arrhythmia, cardiac valve disorder, transient ischemic attack and/or cerebro vascular accident.
IV: In the Netherlands a hysterectomy is not standard of care when performing RRSO.
V: Groups compared using independent samples t-test or Fischer exact test when appropriate.
Abbreviations: RRSO: risk-reducing salpingo-oophorectomy; SD: standard deviation; BMI: body mass index; HRT: hormone replacement therapy.

With regard to breast cancer, we had missing data for 18 women (2.8%). With regard to depression, we had missing data in 22 cases (3.4%). For HRT, we had 78 women (12.2%) who did not remember if they took HRT. We took this into account in our analyses by performing sensitivity analyses. With regard to the other missing confounders: as this was less than 5%, we did not perform multiple imputation.

Outlier removal

In total, 3.0% of test scores was excluded from the analyses, resulting in different numbers per subtest available for analyses. Based on extreme value detection, 1.5% of test scores (106 scores) were excluded from analyses, 1.2% of test scores (83 outliers) were from the verbal memory test. Some participants had difficulty understanding how to enter their answers or completed the recall phase longer than one hour after the learning phase. Participants with outliers on this subtest less frequently used the computer, were on average two years older and were lower educated than participants without outliers. Using MADs, we removed an additional 103 outliers (27 from TMT-B, 32 from Visual Reaction Time, 19 from Tower of London, 7 from

Corsi Block and 18 from Grooved Pegboard) due to improbably low test scores that were likely to reflect technical glitches or a moment of inattention.

Objective cognition

Based on independent samples t-tests, the premenopausal RRSO group performed better than the postmenopausal RRSO group on executive functioning (mean age-corrected z-score .20 and -.01, respectively, p=.005), processing speed (mean age-corrected z-score .33 and -.02, respectively, p<.001) and attention (mean age-corrected z-score .11 and -.05, respectively, p=.01), (Figure 1; Table S2). However, after adjusting for confounders, a premenopausal RRSO was not associated with any of the cognitive domains (β-coefficient and 95%CI for a premenopausal RRSO on verbal memory: 0.07 (-0.26;0.39), processing speed: 0.05 (-0.29;0.38), executive functioning: 0.01 (-0.26;0.28), and attention: 0.14 (-0.10;0.38), see also Table S7). A higher level of education was associated with better cognitive performance. Longer time since RRSO did not influence the results.

Compared with the normative group, the premenopausal RRSO group performed better and the postmenopausal RRSO group similarly on attention, executive functioning and processing speed. Both groups performed slightly worse on verbal memory (Table S2).

Proportion cognitively affected

In the postmenopausal RRSO group, the proportion of cognitively affected women (44.5%) was higher compared with the premenopausal RRSO group (33.3%; p=.01) and compared with 30% expected based on the probability curves.[26] After correcting for confounders, this difference in proportion of cognitively affected between the premenopausal and postmenopausal RRSO disappeared.

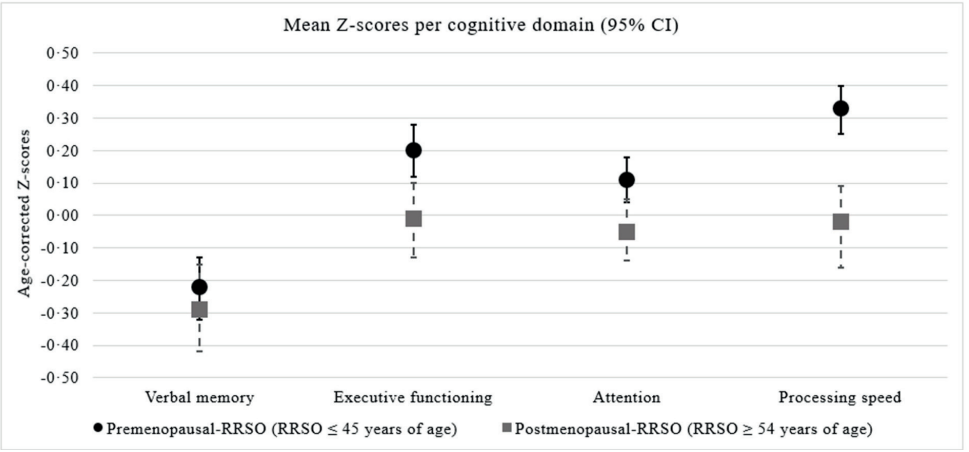


Figure 1. Differences in age-corrected z-scores per cognitive domain and the corresponding 95% confidence interval between the premenopausal and postmenopausal RRSO groups

The Z-scores are age-corrected and describe the score's relationship to the mean in a group of scores, with 0 being the mean of the group and 1 or -1 being 1 standard deviation above or below the mean, respectively.
Abbreviations: RRSO; risk-recucing salpingo-oophorectomy, CI; confidence interval.

Subgroup analysis objective cognition

Based on independent samples t-tests, the premenopausal RRSO ≤40 years group performed similarly on verbal memory, processing speed and attention, but better on executive functioning compared to the premenopausal RRSO at ages 41-45 years group (Figure 2; Table S3). After adjusting for confounders, RRSO ≤40 compared to RRSO at ages 41-45 years was not associated with any cognitive outcome (Table S7).

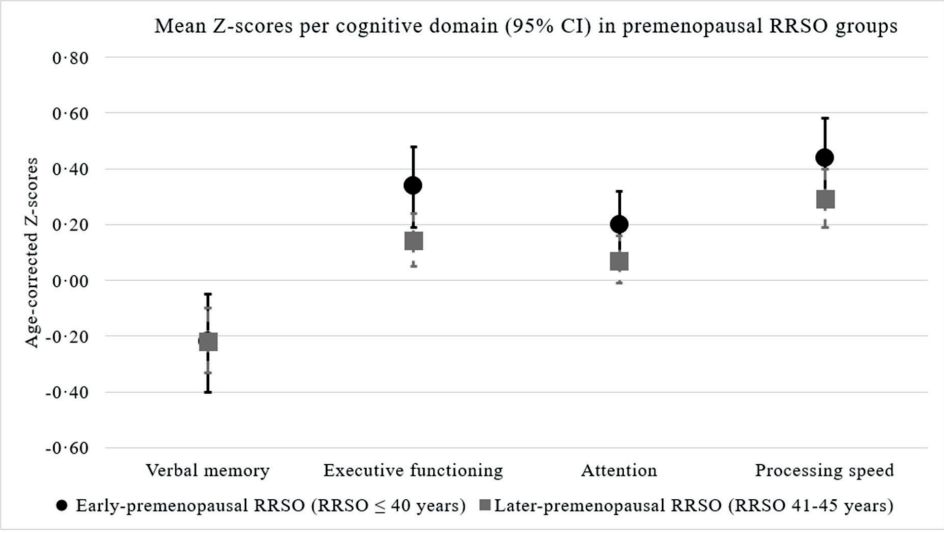


Figure 2. Comparison performance cognitive domains between early-premenopausal RRSO (RRSO ≤ 40 years) versus later-premenopausal RRSO (RRSO 41-45 years)

The Z-scores are age-corrected and describe the score's relationship to the mean in a group of scores, with 0 being the mean of the group and 1 or -1 being 1 standard deviation above or below the mean, respectively.
Abbreviations: RRSO; risk-recucing salpingo-oophorectomy, CI; confidence interval.

The proportion of cognitively affected women with a premenopausal RRSO at 41-45 years was 36.4%, and 26.5% in women with a premenopausal RRSO ≤40 years. This difference was not statistically significant (p=.06). Also, both proportions do not differ significantly from the expected 30% cognitively impaired in a healthy population given the number of tests administered.[26]

When comparing the premenopausal and postmenopausal RRSO groups in the overlapping age range of 60-70 years at study inclusion (n=222), no differences were found in verbal memory, executive functioning, attention

and processing speed (see Table S4). After adjusting for confounders, we found no effect of RRSO on any cognitive domain (Table S7).

When stratifying by history of breast cancer, we also did not find differences between the premenopausal RRSO group and the postmenopausal RRSO group (Table S7). In addition, within the premenopausal RRSO group, there were no differences in cognitive test performance between women with and without breast cancer nor between women with and without HRT use (Table S5).

Subjective cognition

After adjustment for confounders, the premenopausal RRSO group more frequently reported problems with reasoning (OR 1.8, 95% CI 1.1-3.1) and with multitasking (OR 1.9 95% CI 1.1-3.4), both borderline significant (p=.03) (Figure 3). For all six questions on subjective cognition, a depression diagnosis was associated with a higher frequency of complaints (OR between 1.7 – 3.1). A sensitivity analysis showed that time since RRSO did not change the results (Table S6). In the overlapping age category (ages 60-70 years), we did not find differences in subjective cognition.

Subgroup analysis subjective cognition (Table S6)

After adjusting for confounders, a premenopausal RRSO at ages 41-45 was associated with a higher frequency of problems with reasoning (OR 1.66, 95%CI 1.07;2.57) and slow thinking (OR 1.79, 95%CI 1.13;2.82) compared to a premenopausal RRSO before 41 years of age. Sensitivity analyses showed that longer time since RRSO was associated with less frequent cognitive complaints on all six questions (ORs varied between 0.90-0.93; 95%CI's between 0.83;0.99).

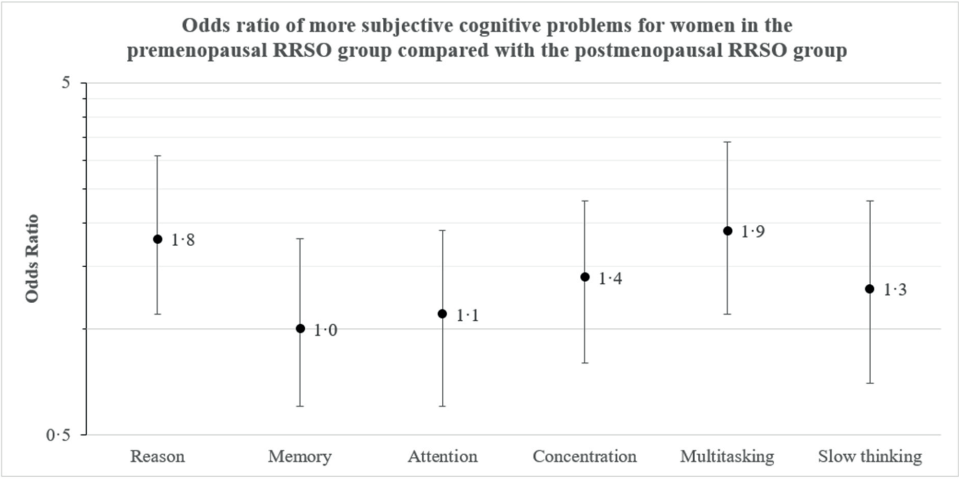


Figure 3. Odds ratios and their 95% confidence interval for more subjective cognitive complaints in women with a premenopausal RRSO as compared to women with a postmenopausal RRSO Corrected for age, breast cancer, HRT, education, depression and cardiovascular risk factors.

In women with a history of breast cancer, no differences were found between the premenopausal and postmenopausal RRSO group. In women without a history of breast cancer, the premenopausal RRSO group more frequently reported problems with reasoning (p=.031, OR:2.5, 95%CI 1.1-5.7). Within women with a premenopausal RRSO, HRT users more frequently reported forgetfulness compared to non-users (p=.03, OR: 1.2, 95%CI 1.1-2.8).

Discussion

After adjusting for age, education, breast cancer, HRT, depression and cardiovascular risk factors among women at high familial risk of breast/ovarian cancer who have had a RRSO, we found no association between timing of RRSO (i.e., premenopausal or postmenopausal) and long-term objective cognition. Women with a premenopausal RRSO before age 46 did not perform worse on cognitive tests than women with a postmenopausal RRSO after age 54. Moreover, women with a premenopausal RRSO before age 41 did not perform worse on cognitive tests than women with a premenopausal RRSO at ages 41-45 years. We found no protective effect of HRT (mean duration, 1.6 years) on cognitive functioning. We showed that the differences in objective cognition between the premenopausal and postmenopausal RRSO groups in univariate analyses could be explained by confounding factors. Regarding subjective cognition, after confounder adjustment, we observed that women with a premenopausal RRSO more frequently reported cognitive complaints about reasoning and multitasking compared to women with a postmenopausal RRSO. Unexpectedly, women with a RRSO at ages 41 – 45 years more frequently reported cognitive complaints about reasoning and thinking speed compared to women with a premenopausal RRSO before age 41.

In general, our results showed no association between premenopausal RRSO and long-term cognitive functioning. This observation is consistent with one recent study[11] in which women with a *BRCA1/2pv* and a mean age at oophorectomy of 46 years completed an online cognitive screening instrument at a mean age of 54. The authors found no differences in test performance between women with a RRSO before and after age 45, nor between women with a RRSO compared to women without a RRSO. Our study confirms and adds to these results by examining the cognitive effects of a premenopausal RRSO, measured by a more extensive cognitive test battery as well as a self-report questionnaire, with substantially longer time since premenopausal RRSO.

However, our results are in contrast with several earlier studies on the effects of RRSO on cognition.[7-10] A possible explanation for the inconsistent findings lies in the adjustment for confounding factors. We showed that adjusting for these factors influenced the results. Previous studies did not account for cancer, HRT, hypertension and depression. It is therefore possible that the previously observed relationship between RRSO and long-term cognition was due to confounding. Also, previous studies included women who underwent bilateral oophorectomy for different indications, i.e., an oophorectomy for cancer treatment, and usually compared with a non-oophorectomy group. This may have caused confounding by indication. The association between RRSO and cognition might have been caused by shared genetic (e.g., pathogenic variant in estrogen receptor) and/or non-genetic factors (e.g., education) that (directly and/or indirectly) increase the likelihood to undergo oophorectomy as well as the risk of developing cognitive impairment.[27-28] In contrast, in our study, all women underwent RRSO because of a high familial risk of ovarian cancer.

Women with a premenopausal RRSO reported more cognitive complaints than women with a postmenopausal RRSO, also after adjustment for age and education. An explanation could be that the age adjustment we applied in the regression analyses was insufficient due to the large age differences between the two groups, as among women aged 60-70, we found no differences in subjective cognition between women with a pre- or postmenopausal RRSO. Another explanation might be that subjective cognition does not only reflect cognitive ability, but also psychosocial factors such as expectations and coping style. Women in the premenopausal RRSO group were more often employed than women in the postmenopausal RRSO group and possibly had higher expectations of their own functioning. In addition, women with a premenopausal RRSO may have been more alert to their cognitive problems because they were aware of possible cognitive consequences of premature menopause.

Unexpectedly, within the premenopausal RRSO, women with RRSO between ages 41-45 reported somewhat more cognitive complaints than women with a premenopausal RRSO ≤ age 40 despite similar stage of life, education and cognitive test performance. The difference was small, and was not in line with our other findings. Future studies could focus on cognitive complaints after a premenopausal RRSO in relation to stage of life and whether they progress over time.

It is noteworthy that, in our study, verbal memory was the only domain where both groups scored lower than the normative population. Verbal memory has frequently been shown to be affected after oophorectomy[9] and is associated with brain regions that are rich in estrogen receptors (i.e., hippocampus and frontal lobe).[29] On other domains, the premenopausal RRSO group performed better than the normative population. Participants visiting clinical geneticists are generally higher educated, have greater awareness of health issues and genetic risk factors for cancer, and are less socially deprived, more affluent than the general population.[28,30] All these characteristics have been associated with better cognition.[31] Moreover, women from *BRCA1/2pv* families tend to have healthier lifestyles than their peers, especially after RRSO.[32] This healthy lifestyle may protect against cognitive impairment.[33]

A limitation of this study is the difference in inclusion rate between the premenopausal and postmenopausal RRSO groups: The inclusion rate in the postmenopausal RRSO group was relatively low, possibly since the HARMOny study was focused on long-term effects after a premenopausal RRSO. This could mean that we overestimated the cognitive ability in the postmenopausal RRSO group as women with lower functioning may not have participated. However, if this is the case, the premenopausal group would have performed even better compared to the postmenopausal RRSO group, providing even more evidence against the earlier hypothesis. Another limitation is possible misclassification bias; 10.1% of the women in

our study could not remember if they ever used HRT. However, we performed sensitivity analyses with patients with missing values included and the results did not differ from the complete case analyses. A last limitation was that we did not correct our analyses for computer use due to multicollinearity between computer use and age, and timing of RRSO. An earlier study of our group showed that more frequent computer use was associated with better performance on the online cognitive tests.[34] Therefore, we may have overestimated the test performance of the premenopausal RRSO group who more often used the computer than the postmenopausal RRSO group.

Conclusion

After adjustment for confounders (age, education, breast cancer, HRT, depression and cardiovascular risk factors), timing of RRSO was not associated with long-term objective cognition. We found no difference in cognitive test performance between women with a premenopausal or postmenopausal RRSO. Women with a premenopausal RRSO did report more complaints, but this may have been due to the large age difference between the groups and/or awareness of potential cognitive consequences of premature menopause in the premenopausal RRSO group. Future studies should longitudinally examine objective and subjective cognition to see whether cognitive changes arise at later ages in women with a premenopausal RRSO. If our results are confirmed by other studies, a clinical implication could be that the age at which women undergo RRSO does not make a difference in long-term cognitive effects. In view of the clinical guidelines for *BRCA1/2* mutation carriers recommending a premenopausal RRSO to reduce ovarian cancer risk and the high uptake of RRSO, our findings regarding cognition are reassuring.

References

1. Kuchenbaecker KB, Hopper JL, Barnes DR, et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *Jama*. 2017;317(23):2402-2416.
2. Meeuwissen PA, Seynaeve C, Brekelmans CT, Meijers-Heijboer HJ, Klijn JG, Burger CW. Outcome of surveillance and prophylactic salpingo-oophorectomy in asymptomatic women at high risk for ovarian cancer. *Gynecol Oncol*. 2005;97(2):476-482.
3. Liu YL, Breen K, Catchings A, et al. Risk-reducing bilateral salpingo-oophorectomy for ovarian cancer: a review and clinical guide for hereditary predisposition genes. *JCO Oncol Pract*. 2022;18(3):201-209.
4. Daan NM, Fauser BC. Menopause prediction and potential implications. *Maturitas*. 2015;82(3):257-265.
5. Scott E, Zhang QG, Wang R, Vadlamudi R, Brann D. Estrogen neuroprotection and the critical period hypothesis. *Front Neuroendocrinol*. 2012;33(1):85-104.
6. van Driel CM, de Bock GH, Arts HJ, et al. Stopping ovarian cancer screening in BRCA1/2 mutation carriers: effects on risk management decisions & outcome of risk-reducing salpingo-oophorectomy specimens. *Maturitas*. 2015;80(3):318-322.
7. Rocca W, Bower J, Maraganore D, et al. Increased risk of cognitive impairment or dementia in women who underwent oophorectomy before menopause. *Neurology*. 2007;69(11):1074-1083.
8. Bove R, Secor E, Chibnik LB, et al. Age at surgical menopause influences cognitive decline and Alzheimer pathology in older women. *Neurology*. 2014;82(3):222-229.
9. Rocca WA, Lohse CM, Smith CY, Fields JA, Machulda MM, Mielke MM. Association of premenopausal bilateral oophorectomy with cognitive performance and risk of mild cognitive impairment. *JAMA Netw Open*. 2021;4(11):e2131448.

10. Ryan J, Scali J, Carriere I, et al. Impact of a premature menopause on cognitive function in later life. *BJOG*. 2014;121(13):1729-1739.
11. Kotsopoulos J, Kim SJ, Armel S, et al. An evaluation of memory and attention in BRCA mutation carriers using an online cognitive assessment tool. *Cancer*. 2021;127(17):3183-3193.
12. Georgakis MK, Beskou-Kontou T, Theodoridis I, Skalkidou A, Petridou ET. Surgical menopause in association with cognitive function and risk of dementia: a systematic review and meta-analysis. *Psychoneuroendocrinology*. 2019;106:9-19.
13. Georgakis MK, Petridou ET. Long-term risk of cognitive impairment and dementia following bilateral oophorectomy in premenopausal women-time to rethink policies? *JAMA Netw Open*. 2021;4(11):e2133016.
14. Gervais NJ, Au A, Almey A, et al. Cognitive markers of dementia risk in middle-aged women with bilateral salpingo-oophorectomy prior to menopause. *Neurobiol Aging*. 2020;94:1-6.
15. Livingston G, Sommerlad A, Orgeta V, et al. Dementia prevention, intervention, and care. *Lancet*. 2017;390(10113):2673-2734.
16. Koppelmans V, Breteler MM, Boogerd W, Seynaeve C, Schagen SB. Late effects of adjuvant chemotherapy for adult onset non-CNS cancer; cognitive impairment, brain structure and risk of dementia. *Crit Rev Oncol Hematol*. 2013;88(1):87-101.
17. Terra L, Hooning MJ, Heemskerk-Gerritsen BAM, et al. Long-term morbidity and health after early menopause due to oophorectomy in women at increased risk of ovarian cancer: protocol for a nationwide cross-sectional study with prospective follow-up (HARMOny study). *JMIR Res Protoc*. 2021;10(1):e24414.
18. Schrijver LH, Olsson H, Phillips KA, et al. Oral contraceptive use and breast cancer risk: retrospective and prospective analyses from a BRCA1 and BRCA2 mutation carrier cohort study. *JNCI Cancer Spectr*. 2018;2(2):pky023.
19. Feenstra HEM, Vermeulen IE, Murre JMJ, Schagen SB. Online self-administered cognitive testing using the Amsterdam Cognition Scan: establishing psychometric properties and normative data. *J Med Internet Res*. 2018;20(5):e192.

20. Feenstra HEM, Murre JMJ, Vermeulen IE, Kieffer JM, Schagen SB. Reliability and validity of a self-administered tool for online neuropsychological testing: the Amsterdam Cognition Scan. *J Clin Exp Neuropsychol*. 2018;40(3):253-273.
21. Stewart AL, Ware, JE. *Measuring functioning and well-being: the medical outcomes study approach*. Durham, NC: Duke University Press; 1992.
22. Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J Exp Soc Psychol*. 2013;49(4):764-766.
23. Klaver KM, Duijts SFA, Geusgens CAV, et al. Internet-based cognitive rehabilitation for WORking Cancer survivors (i-WORC): study protocol of a randomized controlled trial. *Trials*. 2020;21(1):664.
24. Witlox L, Schagen SB, de Ruiter MB, et al. Effect of physical exercise on cognitive function and brain measures after chemotherapy in patients with breast cancer (PAM study): protocol of a randomised controlled trial. *BMJ open*. 2019;9(6):e028117.
25. Gehring K, Sitskoorn MM, Gundy CM, et al. Cognitive rehabilitation in patients with gliomas: a randomized, controlled trial. *J Clin Oncol*. 2009;27(22):3712-3722.
26. Ingraham LJ, Aiken CB. An empirical approach to determining criteria for abnormality in test batteries with multiple measures. *Neuropsychology*. 1996;10(1):120.
27. Rocca WA, Grossardt BR, Shuster LT. Oophorectomy, estrogen, and dementia: a 2014 update. *Mol Cell Endocrinol*. 2014;389(1-2):7-12.
28. van Riel E, van Dulmen S, Ausems MG. Who is being referred to cancer genetic counseling? Characteristics of counselees and their referral. *J Community Genet*. 2012;3(4):265-274.
29. Deroo BJ, Korach KS. Estrogen receptors and human disease. *J Clin Investig*. 2006;116(3):561-570.
30. Holloway SM, Bernhard B, Campbell H, Cetnarskyj R, Lam WW. Inequality of use of cancer genetics services by members of breast, ovarian and colorectal cancer families in South East Scotland. *Fam cancer*. 2008;7(3):259-264.

31. Batty GD, Der G, Macintyre S, Deary IJ. Does IQ explain socioeconomic inequalities in health? Evidence from a population based cohort study in the west of Scotland. *BMJ*. 2006;332(7541):580-584.
32. Michelsen TM, Tonstad S, Pripp AH, Trope CG, Dorum A. Coronary heart disease risk profile in women who underwent salpingo-oophorectomy to prevent hereditary breast ovarian cancer. *Int J Gynecol Cancer*. 2010;20(2):233-239.
33. Fratiglioni L, Marseglia A, Dekhtyar S. Ageing without dementia: can stimulating psychosocial and lifestyle experiences make a difference? *Lancet Neurol*. 2020;19(6):533-543.
34. Lee Meeuw Kjoe PR, Agelink van Rentergem JA, Vermeulen IE, Schagen SB. How to Correct for Computer Experience in Online Cognitive Testing? *Assessment*. 2021;28(5):1247-1255.

Supplementary material

Supplementary Methods I

In the Netherlands, women are tested for a pathogenic variant in a breast cancer gene if they are diagnosed with breast cancer before the age of 40 years, are diagnosed with bilateral breast cancer with the first tumor before the age of 50 years, are diagnosed with multiple primary tumors in one breast, with the first tumor before the age of 50 years, if they are diagnosed with triple negative breast cancer before the age of 60 years, if they are diagnosed with ovarian cancer, or if they are diagnosed with breast cancer and also have a first or second degree relative with breast cancer before age 50, prostate cancer before the age of 60 years, or ovarian cancer. All but five (0.7%) women in the current study were tested for a *BRCA1/2pv*. All women who participated in the study are either probands of *BRCA1/2pv* families, or family members of the *BRCA1/2pv* proband. Date of RRSO was determined via linkage with the Dutch Nationwide Pathology Databank.

Sample size calculation

The sample size was calculated to detect clinical differences in coronary calcium scores in the HARMOny study. However, a previous retrospective study[7] on effects of oophorectomy on cognitive impairment and dementia reported significant findings with 427 women with bilateral oophorectomy before age 49, and a median follow-up of 25 years. With 641 participants and a two-tailed p-value set at .05, we expected this study has more than sufficient power to detect differences between groups and perform subgroup analyses.

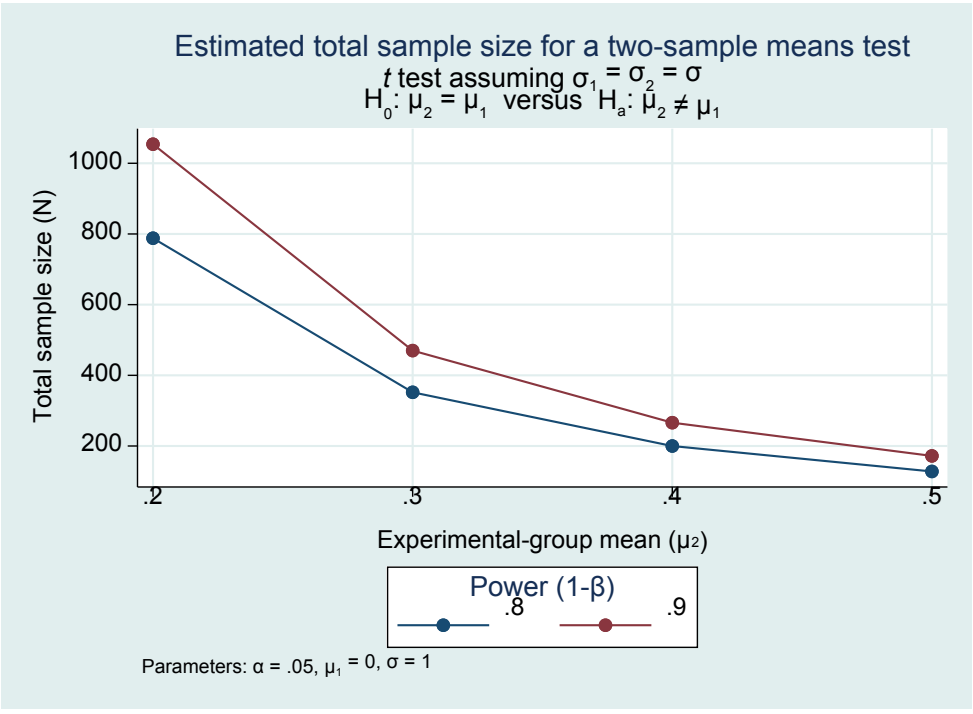

Estimated total sample size for a two-sample means test
t test assuming σ1 = σ2 = σ
H0: μ2 = μ1 versus Ha: μ2 ≠ μ1
Total sample size (N)
1000
800
600
400
200
.2
.3
.4
.5
Experimental-group mean (μ2)
Power (1-β)
.8
.9
Parameters: α = .05, μ1 = 0, σ = 1

Table S1
Tests of the Amsterdam Cognition Scan and their equivalent traditional tests

Cognitive domain	Amsterdam Cognition Scan subtest	Traditional equivalent	Outcome variable
Verbal memory	Word List Learning	RAVLT Immediate Recall	Total number of correct words;
	Word List Delayed Recall	RAVLT Delayed Recall	Total number of free recall
Processing speed	Connect the Dots I	Trail Making Test A	Completion time in seconds
	Reaction Speed	FePsy Visual Reaction Time	Completion time in seconds
	Fill the Grid	Grooved Pegboard	Completion time in seconds
Executive functioning	Connect the Dots II	Trail Making Test B	Completion time in seconds
	Place the Beads	Tower of London	Total number of extra moves
	Digit Sequences II	WAIS III Digit Span backward	Total number of correctly repeated sequences
Attention	Box Tapping	CorsiBlock	Total number of correctly repeated sequences
	Digit Sequences I	WAIS III Digit Span forward	Total number of correctly repeated sequences

Abbreviations: RAVLT; Rey Auditory Verbal Learning Test, WAIS; Wechsler Adult Intelligence Scale.

Table S2
Differences in age-corrected z-scores per cognitive domain between the premenopausal RRSO and postmenopausal RRSO groups

Cognitive domain	Premenopausal RRSO		Postmenopausal RRSO		Effect size[α]	p-value[β]
	n	Mean scores (standard deviation)	n	Mean scores (standard deviation)		
Verbal memory	400	-.22 (1.0)	175	-.29 (.9)	.07	.46
Executive functioning	409	.20 (.8)	172	-.01 (.8)	.26	.005
Processing Speed	407	.33 (.8)	181	-.02 (1.0)	.40	<.001
Attention	432	.11 (.8)	191	-.05 (.7)	.22	.01

α Effect sizes represent Cohen's d.
β Independent samples t-tests used for group comparisons.
NOTE. Comparison between premenopausal and postmenopausal RRSO was done using independent samples t-tests. Due to the different numbers of outliers per cognitive subtest, the number of participants per cognitive domain varies.

Table S3

Comparison performance cognitive domains between early-premenopausal-RRSO (RRSO ≤ 40 years) versus later-premenopausal-RRSO (RRSO 41-45 years)

Cognitive domain	Early-premenopausal-RRSO (RRSO ≤ 40 years)		Later-premenopausal-RRSO (RRSO 41-45 years)		Effect size[α]	p-value[β]
	n	Mean scores (standard deviation)	n	Mean scores (standard deviation)		
Verbal memory	122	-.22 (.98)	276	-.22 (.97)	.00	.969
Executive functioning	124	.34 (.80)	283	.14 (.81)	.24	.029
Processing Speed	124	.44 (.72)	281	.29 (.80)	.20	.071
Attention	131	.20 (.72)	299	.07 (.77)	.17	.114

α Effect sizes represent Cohen's d.
β Independent samples t-tests used for group comparisons.
NOTE. Comparison between premenopausal and postmenopausal RRSO was done using independent samples t-tests. Due to the different numbers of outliers per cognitive subtest, the number of participants per cognitive domain varies.

Table S4

Differences in age-corrected z-scores per cognitive domain between the premenopausal RRSO and postmenopausal RRSO groups aged 60-70 years

Cognitive domain	Premenopausal-RRSO		Postmenopausal-RRSO		Effect size[α]	p-value[β]
	n	Mean Z-scores (standard deviation)	n	Mean Z-scores (standard deviation)		
Verbal memory	151	-.17 (.94)	95	-.23 (.88)	.06	.662
Executive functioning	157	.18 (.89)	97	-.02 (.90)	.22	.087
Processing Speed	161	.30 (.78)	95	.10 (.81)	.26	.048
Attention	171	.10 (.79)	100	.01 (.69)	.12	.34

α Effect sizes represent Cohen's d.
β Independent samples t-tests used for group comparisons.
NOTE. Comparison between premenopausal and postmenopausal RRSO was done using independent samples t-tests. Due to the different numbers of outliers per cognitive subtest, the number of participants per cognitive domain varies.

Table S5

Mean age-corrected z-scores per cognitive domain (SD) per analyses done: Premenopausal RRSO group compared with the postmenopausal RRSO group in: (1) complete group, (2) women aged 60-70 years at questionnaire completion, (3) women with a history of breast cancer, and (4) women without a history of breast cancer. Within the premenopausal RRSO group we compared (5) early premenopausal RRSO with later premenopausal RRSO, and (6) Current or former HRT-users with never HRT users.

		Number	Verbal memory	Executive functioning	Processing speed	Attention
1	Premenopausal RRSO (RRSO ≤ 45 years)	432	-.22 (.97)	.20 (.81)	.33 (.78)	.11 (.76)
	Postmenopausal RRSO (RRSO ≥ 54 years)	191	-.29 (.91)	-.01 (.83)	-.02 (.99)	-.05 (67)
2	Premenopausal RRSO & ages 60-70 years at questionnaire completion	171	-.17 (.88)	.18 (.89)	.30 (.78)	.10 (.79)
	Postmenopausal RRSO & ages 60-70 years at questionnaire completion	100	-.23 (.94)	-.02 (.90)	.10 (.81)	.01 (.69)
3	Premenopausal RRSO & history of breast cancer (YES)	244	-.28 (.95)	.19 (.81)	.36 (.77)	.11 (.74)
	Postmenopausal RRSO & history of breast cancer (YES)	117	-.34 (.92)	-.01 (.84)	-.13 (1.10)	-.02 (.68)
4	Premenopausal RRSO & history of breast cancer (NO)	177	-.17 (1.01)	.24 (.81)	.32 (.76)	.12 (.79)
	Postmenopausal RRSO & history of breast cancer (NO)	69	-.23 (.84)	.01 (.83)	.15 (.77)	-.05 (.63)
	Premenopausal RRSO (RRSO ≤ 45 years)					
5	Early premenopausal RRSO (RRSO ≤ 40 years)	131	-.22 (.98)	.34 (.80)	.44 (.72)	.20 (.72)
	Later premenopausal RRSO (RRSO 41-45 years)	301	-.23 (.97)	.14 (.81)	.28 (.80)	.07 (.77)
6	Premenopausal RRSO (RRSO ≤ 45 years)					
	Current or former HRT-users	108	-.29 (.99)	.19 (.80)	.38 (.69)	.13 (.76)
	Never HRT-users	284	-.22 (.97)	.22 (.82)	.34 (.77)	.05 (.74)

Table S6

Mean scores per subjective cognitive question (95%CI) per analyses done: Premenopausal RRSO group compared with the postmenopausal RRSO group in: (1) complete group, (2) women aged 60-70 years at questionnaire completion, (3) women with a history of breast cancer, and (4) women without a history of breast cancer. Within the premenopausal RRSO group we compared (5) early premenopausal RRSO with later premenopausal RRSO, and (6) Current or former HRT-users with never HRT users.

		Number	Reasoning	Forgetful	Attention	Concentration	Multitasking	Slow thinking
1	Premenopausal RRSO (RRSO ≤ 45 years)	436	1.20 (1.86;2.13)	2.51 (2.41;2.61)	2.33 (2.23;2.43)	2.35 (2.25;2.45)	1.85 (1.76;1.95)	1.81 (1.73;1.89)
	Postmenopausal RRSO (RRSO ≥ 54 years)	215	2.24 (2.13;2.33)	2.21 (2.09;2.34)	1.99 (1.86;2.12)	2.01 (1.87;2.15)	1.52 (1.41;1.63)	1.57 (1.46;1.67)
2	Premenopausal RRSO & ages 60-70 years at questionnaire completion	174	2.17 (2.02;2.32)	2.38 (2.23;2.53)	2.21 (2.05;2.37)	2.21 (2.06;2.36)	1.77 (1.62;1.92)	1.72 (1.59;1.84)
	Postmenopausal RRSO & ages 60-70 years at questionnaire completion	111	1.97 (1.78;2.17)	2.18 (1.99;2.37)	2.03 (1.84;2.22)	1.99 (1.79;2.19)	1.51 (1.35;1.67)	1.59 (1.42;1.75)
3	Premenopausal RRSO & history of breast cancer (YES)	253	2.35 (2.10;2.37)	2.50 (2.37;2.63)	2.39 (2.25;2.52)	2.40 (2.27;2.53)	1.91 (1.77;2.05)	1.81 (1.70;1.93)
	Postmenopausal RRSO & history of breast cancer (YES)	137	2.01 (1.85;2.17)	2.16 (2.00;2.32)	1.99 (1.82;2.15)	2.03 (1.85;2.21)	1.55 (1.41;1.69)	1.59 (1.46;1.72)
4	Premenopausal RRSO & history of breast cancer (NO)	183	2.23 (2.09;2.38)	2.52 (2.37;2.67)	2.26 (2.10;2.42)	2.27 (2.12;2.43)	1.77 (1.64;1.91)	1.81 (1.68;1.93)
	Postmenopausal RRSO & history of breast cancer (NO)	78	1.97 (1.73;2.22)	2.31 (2.09;2.53)	2.00 (1.79;2.21)	1.97 (1.75;2.20)	1.47 (1.30;1.63)	1.52 (1.33;1.71)
5	Premenopausal RRSO (RRSO ≤ 45 years)							
	Early premenopausal RRSO (RRSO ≤ 40 years)	133	2.17 (1.97;2.36)	2.49 (2.30;2.68)	2.31 (2.12;2.50)	2.30 (2.11;2.49)	1.70 (1.53;1.87)	1.69 (1.54;1.85)
	Later premenopausal RRSO (RRSO 41-45 years)	303	2.27 (2.15;2.38)	2.51 (2.40;2.63)	2.34 (2.22;2.46)	2.37 (2.25;2.49)	1.92 (1.80;2.04)	1.86 (1.76;1.96)
6	Premenopausal RRSO (RRSO ≤ 45 years)							
	Current or former HRT-users	114	2.33 (2.13;2.53)	2.68 (2.48;2.89)	2.42 (2.20;2.63)	2.43 (2.24;2.63)	1.85 (1.65;2.05)	1.88 (1.71;2.04)
	Never HRT-users	292	2.17 (2.06;2.29)	2.41 (2.29;2.53)	2.29 (2.16;2.41)	2.30 (2.18;2.43)	1.84 (1.72;1.96)	1.76 (1.66;1.87)

Table S7

B-coefficient and its corresponding 95% confidence interval per cognitive domain (95%CI) per analyses done: Premenopausal RRSO group compared with the postmenopausal RRSO group in: (1) complete group, (2) women aged 60-70 years at questionnaire completion, (3) women with a history of breast cancer, and (4) women without a history of breast cancer. Within the premenopausal RRSO group we compared, and (5) early premenopausal RRSO with later premenopausal RRSO. Adjustment was done for the following confounders: age at questionnaire completion, breast cancer (yes/no), hormone replacement therapy (yes/no), depression (yes/no), educational level, cardiovascular disease.

1	Number (%)	Verbal memory	Number (%)
Postmenopausal RRSO (RRSO ≥ 54 years)	175 (30.4%)	0.00 (REF)	181 (30.8%)
Premenopausal RRSO (RRSO ≤ 45 years)	400 (69.6%)	0.07 (-0.26;0.39)	407 (69.2%)
Age 55-59 years	243 (42.3%)	0.00 (REF)	240 (40.8%)
Age 60-64 years	145 (25.2%)	0.07 (-0.20;0.35)	152 (25.9%)
Age 65-69 years	101 (17.6%)	0.10 (-0.31;0.512)	104 (17.7%)
Age 70-74 years	63 (11.0%)	0.19 (-0.33;0.71)	67 (11.4%)
Age 75 + years	23 (4.0%)	0.23 (-0.45;0.90)	25 (4.3%)
Breast cancer (No)	220 (39.4%)	0.00 (REF)	232 (40.3%)
Breast cancer (Yes)	339 (60.6%)	-0.16 (-0.40;0.09)	344 (59.7%)
HRT (No, never)	405 (78.2%)	0.00 (REF)	418 (77.7%)
HRT (Yes, current or former)	113 (21.8%)	-0.09 (-0.40;0.21)	120 (22.3%)
Primary school/lower level high school	162 (30.7%)	0.00 (REF)	166 (30.8%)
Middle level high school	168 (31.9%)	0.27 (-0.002;0.55)	172 (31.9%)
Advanced vocational/university	197 (37.4%)	0.48 (0.21;0.75)	201 (37.3%)
Depression (No)	482 (86.9%)	0.00 (REF)	495 (86.5%)
Depression (Yes)	73 (13.2%)	-0.02 (-0.38;0.33)	77 (13.5%)
Cardiovascular disease (No)	267 (46.4%)	0.00 (REF)	273 (46.4%)
Cardiovascular disease (Yes)	308 (53.6%)	-0.24 (0.47;-0.02)	315 (53.6%)
Constant		-0.25 (-1.03;0.52)	

Table continues

Processing speed	Number (%)	Executive functioning	Number (%)	Attention
0.00 (REF)	172 (29.6%)	0.00 (REF)	191 (30.7%)	0.00 (REF)
0.05 (-0.29;0.38)	409 (70.4%)	0.01 (-0.26;0.28)	432 (69.3%)	0.14 (-0.10;0.38)
0.00 (REF)	246 (42.3%)	0.00 (REF)	255 (40.9%)	0.00 (REF)
0.03 (-0.20;0.26)	150 (25.8%)	0.03 (-0.20;0.27)	162 (26.0%)	-0.07 (-0.23;0.09)
-0.22 (-0.56;0.13)	104 (17.9%)	-0.20 (-0.55;0.14)	109 (17.5%)	0.05 (-0.21;0.30)
-0.04 (-0.48;0.39)	61 (10.5%)	-0.15 (-0.59;0.29)	73 (11.7%)	0.10 (-0.20;0.41)
-0.67 (-1.22;-0.13)	20 (3.4%)	-0.08 (-0.67;0.51)	24 (3.9%)	-0.24 (-0.64;0.17)
0.00 (REF)	227 (39.8%)	0.00 (REF)	246 (40.5%)	0.00 (REF)
-0.07 (-0.27;0.12)	343 (60.2%)	-0.06 (-0.26;0.14)	361 (59.5%)	-0.02 (-0.16;0.12)
0.00 (REF)	415 (77.9%)	0.00 (REF)	439 (77.4%)	0.00 (REF)
0.08 (-0.16;0.32)	118 (22.1%)	0.01 (-0.23;0.26)	128 (22.6%)	-0.05 (-0.22;0.12)
0.00 (REF)	163 (30.8%)	0.00 (REF)	181 (32.1%)	0.00 (REF)
0.29 (0.06;0.52)	167 (31.5%)	0.30 (0.07;0.54)	177 (31.4%)	0.23 (0.06;0.39)
0.39 (0.17;0.61)	200 (37.7%)	0.50 (0.27;0.73)	206 (36.5%)	0.25 (0.10;0.40)
0.00 (REF)	490 (86.6%)	0.00 (REF)	520 (86.2%)	0.00 (REF)
-0.09 (-0.37;0.19)	76 (13.4%)	-0.16 (-0.45;0.12)	83 (13.8%)	-0.05 (-0.25;0.14)
0.00 (REF)	276 (47.5%)	0.00 (REF)	285 (45.8%)	0.00 (REF)
-0.13 (-0.32;0.06)	305 (52.5%)	-0.18 (-0.37;0.01)	338 (54.3%)	-0.11 (-0.23;0.02)
0.20 (-0.45;0.85)		-0.14 (-0.78;0.51)		-0.09 (-0.38;0.20)

2	Number (%)	Verbal memory	Number (%)
Postmenopausal RRSO & ages 60-70 years at questionnaire completion	173 (52.1%)	0.00 (REF)	179 (51.4%)
Premenopausal RRSO & ages 60-70 years at questionnaire completion	159 (47.9%)	0.04 (-0.25;0.34)	169 (48.6%)
Age (per 1 year)		-0.001 (-0.03;0.03)	
Breast cancer (No)	127 (39.2%)	0.00 (REF)	138 (40.4%)
Breast cancer (Yes)	197 (60.8%)	0.07 (-0.16;0.29)	204 (59.7%)
HRT (No, never)	233 (79.0%)	0.00 (REF)	250 (79.6%)
HRT (Yes, current or former)	62 (21.0%)	0.10 (-0.20;0.39)	64 (20.4%)
Primary school/lower level high school	103 (33.7%)	0.00 (REF)	105 (33.2%)
Middle level high school	81 (26.5%)	0.11 (-0.16;0.38)	86 (27.2%)
Advanced vocational/university	122 (39.9%)	0.51 (0.26;0.75)	125 (39.6%)
Depression (No)	283 (87.9%)	0.00 (REF)	296 (87.1%)
Depression (Yes)	39 (12.1%)	0.03 (-0.33;0.38)	44 (12.9%)
Cardiovascular disease (No)	133 (40.1%)	0.00 (REF)	142 (40.8%)
Cardiovascular disease (Yes)	199 (59.9%)	-0.18 (-0.40;0.04)	206 (59.2%)
Constant		-0.40 (-2.39;1.59)	

3	Number (%)	Verbal memory	Number (%)
Postmenopausal RRSO & history of breast cancer (YES)	110 (32.5%)	0.00 (REF)	112 (32.6%)
Premenopausal RRSO & history of breast cancer (YES)	229 (67.6%)	0.32 (-0.16;0.80)	232 (67.4%)
Age 55-59 years	142 (41.9%)	0.00 (REF)	140 (40.7%)
Age 60-64 years	86 (25.4%)	0.12 (-0.15;0.39)	87 (25.3%)
Age 65-69 years	53 (15.6%)	0.34 (-0.17;0.85)	55 (16.0%)
Age 70-74 years	43 (12.7%)	0.32 (-0.26;0.90)	47 (13.7%)
Age 75 + years	13 (4.4%)	0.46 (-0.23;1.15)	15 (4.4%)
HRT (No, never)	286 (84.4%)	0.00 (REF)	286 (83.1%)
HRT (Yes, current or former)	25 (7.4%)	0.28 (-.14;0.70)	31 (9.0%)
Primary school/lower level high school	107 (33.4%)	0.00 (REF)	104 (32.2%)
Middle level high school	94 (29.4%)	0.18 (-0.09;0.45)	94 (29.1%)
Advanced vocational/university	119 (37.2%)	0.51 (0.25;0.76)	15 (38.9%)
Depression (No)	295 (87.8%)	0.00 (REF)	297 (87.1%)
Depression (Yes)	41 (12.2%)	-0.41 (-0.77;-0.06)	44 (12.9%)
Cardiovascular disease (No)	163 (48.1%)	0.00 (REF)	165 (48.0%)
Cardiovascular disease (Yes)	176 (51.9%)	-0.13 (-0.34;0.09)	179 (52.0%)
Constant		-0.79 (-1.31;-0.26)	

Processing speed	Number (%)	Executive functioning	Number (%)	Attention
0.00 (REF)	170 (50.8%)	0.00 (REF)	189 (51.4%)	0.00 (REF)
-0.02 (-0.31;0.27)	165 (49.3%)	0.03 (-0.26;0.31)	179 (48.6%)	0.02 (-0.22;0.26)
-0.04 (-0.07;-0.01)		-0.01 (-0.04;0.02)		-0.005 (-0.03;0.02)
0.00 (REF)	129 (39.2%)	0.00 (REF)	145 (40.3%)	0.00 (REF)
0.001 (-0.22;0.22)	200 (60.8%)	-0.03 (-0.24;0.18)	215 (59.7%)	0.02 (-0.16;0.20)
0.00 (REF)	242 (79.6%)	0.00 (REF)	262 (78.9%)	0.00 (REF)
0.09 (-0.20;0.37)	62 (20.4%)	0.14 (-0.13;0.41)	70 (21.1%)	0.16 (-0.06;0.39)
0.00 (REF)	101 (33.2%)	0.00 (REF)	115 (34.6%)	0.00 (REF)
0.25 (-0.02;0.51)	81 (26.6%)	0.21 (-0.05;0.47)	88 (26.5%)	0.21 (-0.01;0.42)
0.37 (0.13;0.62)	122 (40.1%)	0.37 (0.13;0.60)	129 (38.9%)	0.18 (-0.01;0.38)
0.00 (REF)	286 (87.5%)	0.00 (REF)	312 (87.2%)	0.00 (REF)
0.16 (-0.17;0.49)	41 (12.5%)	-0.07 (-0.39;0.25)	46 (12.9%)	-0.08 (-0.35;0.19)
0.00 (REF)	141 (42.1%)	0.00 (REF)	148 (40.2%)	0.00 (REF)
-0.20 (-0.42;0.01)	194 (57.9%)	-0.24 (-0.45;-0.04)	220 (59.8%)	-0.10 (-0.27;0.08)
2.85 (0.89;4.81)		0.71 (-1.24;2.65)		0.22 (-1.39;1.83)

Processing speed	Number (%)	Executive functioning	Number (%)	Attention
0.00 (REF)	106 (30.9%)	0.00 (REF)	117 (32.4%)	0.00 (REF)
-0.01 (-0.48;0.46)	237 (69.1%)	-0.21 (-0.63;0.20)	244 (67.6%)	-0.18 (-0.53;0.18)
0.00 (REF)	143 (41.7%)	0.00 (REF)	146 (44.4%)	0.00 (REF)
0.18 (-0.08;0.44)	89 (26.0%)	-0.07 (-0.30;0.17)	94 (26.0%)	-0.10 (-0.30;0.11)
-0.26 (-0.75;0.23)	58 (16.9%)	-0.36 (-0.78;0.07)	57 (15.8%)	-0.21 (-0.58;0.11)
-0.48 (-1.05;0.08)	41 (12.0%)	-0.53 (-1.04;-0.03)	49 (13.6%)	-0.19 (-0.63;0.24)
-0.81 (-1.48;-0.13)	12 (3.5%)	-0.32 (-0.95;0.32)	15 (4.2%)	-0.68 (-1.20;-0.15)
0.00 (REF)	285 (83.1%)	0.00 (REF)	301 (90.1%)	0.00 (REF)
0.11 (-0.25;0.48)	31 (9.0%)	-0.04 (-0.36;0.29)	33 (9.9%)	-0.004 (-0.28;0.27)
0.00 (REF)	104 (32.5%)	0.00 (REF)	115 (34.1%)	0.00 (REF)
0.14 (-0.13;0.41)	91 (28.4%)	0.26 (0.02;0.50)	97 (28.8%)	0.24 (0.04;0.45)
0.36 (0.12;0.61)	125 (39.1%)	0.49 (0.27;0.71)	125 (37.1%)	0.26 (0.07;0.45)
0.00 (REF)	295 (86.8%)	0.00 (REF)	310 (86.6%)	0.00 (REF)
-0.12 (-0.45;0.21)	45 (13.2%)	-0.21 (-0.51;0.09)	48 (13.4%)	-0.14 (-0.39;0.11)
0.00 (REF)	168 (49.0%)	0.00 (REF)	172 (47.7%)	0.00 (REF)
-0.16 (-0.37;0.05)	175 (51.0%)	-0.13 (-0.31;0.06)		-0.13 (-0.29;0.03)
0.21 (-0.31;0.72)		0.28 (-0.17;0.72)		0.22 (-0.17;0.60)

4	Number (%)	Verbal memory	Number (%)
Postmenopausal RRSO & history of breast cancer (NO)	60 (27.3%)	0.00 (REF)	66 (28.5%)
Premenopausal RRSO & history of breast cancer (NO)	160 (72.7%)	0.09 (-0.40;0.59)	166 (71.6%)
Age 55-59 years	93 (42.3%)	0.00 (REF)	94 (40.5%)
Age 60-64 years	56 (25.5%)	-0.06 (-0.42;0.30)	62 (26.7%)
Age 65-69 years	44 (20.0%)	-0.17 (-0.67;0.33)	46 (19.8%)
Age 70-74 years	19 (8.6%)	-0.25 (-0.87;0.38)	20 (8.6%)
Age 75 + years	8 (3.6%)	-0.09 (-1.00;0.83)	10 (4.3%)
HRT (No, never)	119 (54.1%)	0.00 (REF)	132 (56.9%)
HRT (Yes, current or former)	88 (40.0%)	-0.24 (-0.56;0.07)	89 (38.4%)
Primary school/lower level high school	55 (26.6%)	0.00 (REF)	62 (28.7%)
Middle level high school	74 (35.8%)	0.14 (0.22;0.50)	78 (36.1%)
Advanced vocational/university	78 (37.7%)	0.49 (0.13;0.85)	76 (35.2%)
Depression (No)	187 (85.4%)	0.00 (REF)	198 (85.7%)
Depression (Yes)	32 (14.6%)	0.11 (-0.34;0.55)	33 (14.3%)
Cardiovascular disease (No)	104 (47.3%)	0.00 (REF)	108 (46.6%)
Cardiovascular disease (Yes)	116 (52.7%)	-0.24 (-0.53;0.05)	124 (53.5%)
Constant		-0.20 (-0.79;0.39)	
Premenopausal RRSO (RRSO ≤ 45 years)			
5	**Number (%)**	**Verbal memory**	**Number (%)**
Later premenopausal RRSO (RRSO 41-45 years)	278 (69.5%)	0.00 (REF)	283 (69.5%)
Early premenopausal RRSO (RRSO ≤ 40 years)	122 (30.5%)	0.07 (-0.17;0.31)	124 (30.5%)
Age (per 1 year)		-0.00 (-0.03;0.03)	
Breast cancer (No)	160 (41.1%)	0.00 (REF)	166 (41.7%)
Breast cancer (Yes)	229 (58.9%)	-0.18 (-0.43;0.07)	232 (58.3%)
HRT (No, never)	265 (66.3%)	0.00 (REF)	270 (66.3%)
HRT (Yes, current or former)	96 (24.0%)	-0.18 (-0.47;0.12)	102 (25.1%)
Primary school/lower level high school	100 (27.3%)	0.00 (REF)	101 (26.9%)
Middle level high school	131 (35.8%)	0.20 (-0.06;0.47)	137 (36.5%)
Advanced vocational/university	135 (36.9%)	0.41 (0.15;0.68)	137 (36.5%)
Depression (No)	331 (85.5%)	0.00 (REF)	337 (85.1%)
Depression (Yes)	56 (14.5%)	-0.12 (-0.45;0.21)	59 (14.9%)
Cardiovascular disease (No)	205 (51.3%)	0.00 (REF)	205 (50.4%)
Cardiovascular disease (Yes)	195 (48.8%)	-0.18 (-0.39;0.03)	202 (49.6%)
Constant		-0.19 (-2.12;1.73)	

Processing speed	Number (%)	Executive functioning	Number (%)	Attention
0.00 (REF)	63 (27.8%)	0.00 (REF)	69 (28.1%)	0.00 (REF)
-0.01 (-0.39;0.36)	164 (72.3%)	0.22 (-0.17;0.62)	177 (72.0%)	0.50 (0.13;0.87)
0.00 (REF)	98 (43.2%)	0.00 (REF)	101 (41.1%)	0.00 (REF)
-0.22 (-0.48;0.04)	58 (25.6%)	0.02 (-0.26;0.30)	65 (26.4%)	-0.03 (-0.29;0.23)
-0.21 (-0.59;0.17)	43 (18.9%)	-0.04 (-0.44;0.37)	48 (19.5%)	0.29 (-0.08;0.66)
-0.28 (-0.75;0.19)	20 (8.8%)	0.11 (-0.39;0.62)	23 (9.4%)	0.37 (-0.09;0.83)
-0.47 (-1.11;0.17)	8 (3.5%)	-0.4 (-0.78;0.71)	9 (3.7%)	0.37 (-0.30;1.04)
0.00 (REF)	130 (57.3%)	0.00 (REF)	138 (56.1%)	0.00 (REF)
0.06 (-0.18;0.30)	87 (38.3%)	-0.08 (-0.33;0.17)	95 (38.6%)	-0.17 (-0.41;0.06)
0.00 (REF)	59 (28.1%)	0.00 (REF)	66 (29.1%)	0.00 (REF)
0.27 (0.01;0.53)	76 (36.2%)	0.23 (-0.05;0.51)	80 (35.2%)	0.22 (-0.04;0.48)
0.39 (0.12;0.66)	75 (35.7%)	0.38 (0.09;0.67)	81 (35.7%)	0.29 (0.03;0.56)
0.00 (REF)	195 (86.3%)	0.00 (REF)	210 (85.7%)	0.00 (REF)
0.10 (-0.22;0.43)	31 (13.7%)	0.03 (-0.33;0.39)	35 (14.3%)	0.18 (-0.14;0.51)
0.00 (REF)	108 (48.6%)	0.00 (REF)	113 (45.9%)	0.00 (REF)
-0.9 (-0.31;0.13)	119 (52.4%)	-0.17 (-0.40;0.07)	133 (54.1%)	-0.09 (-0.31;0.12)
0.21 (-0.22;0.65)		-0.07 (-0.54;0.39)		-0.45 (-0.88;-0.03)

Processing speed	Number (%)	Executive functioning	Number (%)	Attention
0.00 (REF)	285 (69.7%)	0.00 (REF)	301 (69.7%)	0.00 (REF)
0.13 (-0.05;0.31)	124 (30.3%)	0.18 (-0.01;0.38)	131 (30.3%)	0.13 (-0.04;0.31)
-0.02 (-0.04;0.01)		-0.01 (-0.03;0.02)		0.004 (-0.02;0.03)
0.00 (REF)	164 (40.9%)	0.00 (REF)	199 (42.0%)	0.00 (REF)
0.02 (-0.17;0.20)	237 (59.1%)	-0.11 (-0.31;0.09)	244 (58.0%)	-0.11 (-0.29;0.07)
0.00 (REF)	276 (67.5%)	0.00 (REF)	284 (65.7%)	0.00 (REF)
0.05 (-0.17;0.26)	99 (24.2%)	-0.15 (-0.31;0.09)	108 (25.0%)	-0.18 (-0.39;0.03)
0.00 (REF)	103 (27.5%)	0.00 (REF)	110 (27.9%)	0.00 (REF)
0.00 (-0.20;0.20)	134 (35.7%)	0.25 (0.03;0.46)	142 (36.0%)	0.21 (0.02;0.41)
0.16 (-0.04;0.37)	138 (36.8%)	0.40 (0.19;0.62)	142 (36.0%)	0.28 (0.08;0.47)
0.00 (REF)	340 (85.2%)	0.00 (REF)	356 (85.0%)	0.00 (REF)
-0.19 (-0.43;0.05)	59 (14.8%)	-0.10 (-0.36;0.15)	63 (15.0%)	-0.01 (-0.25;0.23)
0.00 (REF)	208 (50.9%)	0.00 (REF)	215 (49.8%)	0.00 (REF)
-0.08 (-0.24;0.08)	201 (49.1%)	-0.07 (-0.24;0.10)	217 (50.2%)	-0.09 (-0.25;0.06)
1.19 (-0.26;2.64)		0.61 (-0.92;2.14)		-0.16 (-1.57;1.26)

Table S8

ACS absurd values detection

Subtest ACS	Scores identified as outliers
TMT A	<10 sec, >180 sec
TMT B	<10 sec, >240 sec
RAVLT IR	<5, >75
RAVLT DR	<1, >15 > 1 hour between end RAVLT IR and start RAVLT DR
Visual Reaction Time	<130 ms, >3950 ms
ToL	First, correction per trial: time to complete trial >2 minutes → score=20 and maximal score per trial = 20 <0, >200
Corsi block	<1, >20
Pegboard	<10 sec, >200 sec
DS forward	<1, >16
DS reverse	<1, >14

Figure S1. Participant flowchart. All women that were invited to participate in the HARMOny study were included in the HEBON study cohort. Number of participant screenings, inclusions, non-responders and decline of participation

Abbreviations: RRSO; risk-reducing salpingo-oophorectomy.

Chapter 8:
Summary and General Discussion

The relevance of online cognitive assessment in oncology: the Amsterdam Cognition Scan

The first aim of this thesis was to further develop the Amsterdam Cognition Scan (ACS). The second aim was to implement cognitive tests in studies on cognitive effects of endocrine therapy (ET).

In this chapter, the results of the research described in this thesis will shortly be summarized, put into broader perspective, and methodological considerations, implications and future directions will be provided. This is done separately for both aims.

Part I: Further development of the Amsterdam Cognition Scan

Summary of findings

In **Chapter 2,** a study is described in which it is investigated how to correct for computer experience when analyzing online cognitive test performance. Computer experience was assessed by a performance-based and a self-report measure, and the influence of computer experience was examined on both the ACS and its equivalent traditional paper-and-pencil tests. It was found that after correction for demographics, better performance-based computer experience was associated with better performance on both the ACS and traditional, predominantly speed-based, tests. Better self-reported computer experience was associated with better performance on speed-based ACS tests but not with performance on most traditional tests. It was concluded that correcting for computer experience is best done using a self-report measure. Because computer use can be influenced by disorders, it is recommended to correct for computer use prior to diagnosis rather than current use.

In **Chapter 3,** a study is described on cross-lingual word criteria for new parallel Rey Auditory Verbal Learning Test (RAVLT)-based word lists in the ACS. Thirteen word selection criteria were identified, two new American-English and one Dutch word list(s) were developed and the criteria were validated by comparing online performance on the new word lists to performance on the original word lists. It was found that application of the word selection criteria led to two highly comparable new American-English word lists that both had lower trial scores compared to the original American-English list, possibly indicating that the criteria helped to develop parallel lists with fewer unwanted word associations. The new Dutch word list was highly comparable to the original Dutch version on all outcomes. It was concluded that application of the word selection criteria can guide development of new parallel ACS RAVLT word lists, including in new language areas.

In **Chapter 4**, a study is described that examined the existence of cognitive subgroups in breast cancer patients who had undergone chemotherapy using a data driven approach of ACS data. Subgrouping results were compared to traditional normative comparisons results. The results consistently showed two subgroups using various clustering methods: One had cognitive normal scores, while the other had lower scores on all cognitive tests. The group with lower scores consisted of 45% of the patients. Results of the subgroup clustering model matched those of traditional normative comparison method requiring ≥ two test scores of ≥ one SD below the normative mean, with the additional requirement that the two test scores are from different cognitive domains. The data-driven subgrouping method provided support for the existence of two subgroups, and proved to be useful for identification of cognitively affected and unaffected patients using ACS data.

Discussion of findings

Computer experience

The results of chapter 2 confirm and add to our earlier findings that more computer experience was associated with better ACS test performance.[1] It was concluded that correcting online cognitive test performance for computer experience is best done using a self-report measure of computer use. The finding of a self-report measure of computer experience that correlates with online test performance only is a useful addition to the literature as currently no consensus exists on the operationalization of the construct of computer experience.

Although it is expected that in future generations virtually everyone will use computers, the influence of computer experience on ACS test performance will likely remain relevant. People will most probably continue to differ in frequency of computer use at work (c.f., a IT specialist and a pet sitter) and in personal life (c.f., a gamer and a hiker). Since digital cognitive performance is potentially associated with technology use, correction for (previous) technology use remains relevant.

Standardization of test development

The finding that the newly developed word criteria helped to create parallel word lists for the verbal memory subtest is in line with research showing that many word properties affect the memorability of words[2] and is a valuable contribution to the field of verbal memory assessment as currently no extensive test development guidelines exist. It was concluded that application of the word criteria can guide development of new parallel ACS RAVLT word lists across languages. The new word criteria can also be used for word lists of verbal memory tests outside of the ACS and therefore, its development is in line with the current call for international harmonization of cognitive tests and measures.

Additionally, standardization of ACS test development would be useful not only for the verbal memory subtest but also for the remaining six subtests. New versions will be needed for other countries, as well as parallel versions. Development of new versions for other countries potentially requires country-specific adaptations. Development of parallel versions requires standardized test design. Test development procedures including these aspects can facilitate the creation of equivalent ACS subtests for use within and between countries.

Identifying affected patients

The data-driven subgrouping method provided support for the existence of two subgroups and proved to be useful for identification of cognitively affected and unaffected patients using ACS data. The existence of cognitive subgroups is not studied before and is highly relevant as reliable identification of cognitively affected patients is a point of concern in the field.[3] The percentage of cognitively affected patients —45%— was higher than typically found based on consensus-based cutoffs. By approaching the problem of subgroups from a new angle that does not rely on consensus-based cutoffs, we prevent the application of a binary classification strategy that is unreliable. Although our data-driven method is more cumbersome to apply, we strongly advocate the use of this analytic approach to reliably determine the existence of subgroups as this will also enable the investigation of risk factor profiles.

Methodological considerations

When interpreting the results of this first part of the thesis, two methodological considerations should be taken into account.

First, the research samples consisted of cancer patients and controls without cancer who presumably have a positive attitude toward computers

since they volunteered to undergo the ACS. Also, participants were predominantly highly educated. Based on these samples, the results cannot be generalized to people who have a negative attitude toward computers nor to those with a lower educational background. It should be noted however that the ACS is designed to minimize demands of computer skills and education level by elaborate instructional videos, practice rounds with feedback, simplification of user responses and inclusion of mostly non-verbal subtests.

Second, due to the unsupervised administration of the ACS, it cannot be ruled out with certainty if participants performed their utmost best and/or made use of external aids such as a notepad or had help of others. It should be noted however that noncredible performance (i.e., performance that is either worse or better than a patient is capable of) is uncommon in cancer patients and controls without cancer who participate in cognitive research.[4] Also, there was no clear incentive for making use of external aids or receiving help of others since the level of cognitive test performance was not associated with personal benefits. Furthermore, there were in general no indications for unreliable data such as unexpectedly low or high scores and unusual scores were removed from the analyses based on prespecified cutoff values and outlier detection methods.

Implications

The findings of the current thesis can help to improve ACS test interpretation and development. Interpretation of ACS test performance can be improved by correcting ACS scores for the self-report measure of frequency of computer use per week. Additionally, the normative database of the ACS should include information on the people's level of computer use to make the demographically-corrected normative comparisons even more accurate. Subtyping methods can be used to identify cognitive subgroups in patient groups. In terms of ACS test development, the new word criteria enable new language versions of the ACS to be made.

The ACS is ready for widespread implementation in studies. As of the beginning of 2023, the ACS is implemented in approximately 30 clinical studies. Given its broad spectrum of cognitive tests, the ACS can be implemented in settings outside of oncology as well, such as the field of dementia and traumatic brain injury. As an example, the ACS is currently being used in a study examining long-term cognitive effects of Coronavirus disease 2019 (COVID-19).

Future research

To increase the (international) applicability of the ACS in oncological studies and outside oncology, several aspects of the ACS can be further developed, including new language versions with normative databases, extension of subtests and parallel tests. As this thesis focuses on oncology, several additional ideas will shortly be discussed in relation to the ACS that can help to achieve a better understanding of the incidence, prevalence and nature of cognitive problems in cancer patients.

It may be useful to develop a mobile (tablet or smartphone) version of the ACS. Mobile cognitive assessments have some advantages over computer assessments including frequent longitudinal testing in real time and easy combination with data from wearables and self-report data.[5] Smartphone and tablet use is rapidly increasing across countries. As of 2021, the proportion of Americans that own a smartphone (85%) is higher than the proportion that owns a desktop/laptop (77%).[6] Since the response input of the ACS is now based on the computer mouse and keyboard, development of a tablet or smartphone version would entail substantial adaptations of all subtests to make input by touch screen possible. Additionally, psychometric properties such as reliability, validity and feasibility should be examined again and new normative data should be collected.

To further improve ACS test score interpretation, future research could examine whether underlying cognitive processes can be disentangled from traditional outcome measures. A previous study of our group[7] showed that computational modelling of data from the ACS version of the Trail Making Test[8] indicated that cancer patients after systemic therapy were cognitively slower and more indecisive than people without cancer when a test was complex, while they did not differ from controls in motor speed. Traditional data only indicated that cancer patients treated with systemic therapy were slower than controls. Disentangling underlying cognitive processes in other ACS subtests can help to identify (the nature of the) cognitive impairment more precisely.

Effort measures could be developed for the ACS. Performance on the ACS is only informative when patients put effort and motivation in completing the tests. Effort measures consist of either stand-alone or embedded performance validity measures.[9] Stand-alone tests are separate tests that are solely administered to measure effort. Embedded measures are derived from data of the existing cognitive subtests. To minimize ACS completion time, embedded performance validity measures would be preferred.

Ultimately, the clinical care of patients confronted with cognitive problems should be improved. Regular implementation of the ACS can fuel our research, but can also be instrumental for daily oncology practice. Based on the ACS tests, a (computer-adaptive) screening tool can be developed for the use in clinical care. Patients could complete the screening test at home. And with this instrument, patients can reliably and quickly be referred to further diagnostics and/or the right follow-up care. Such an innovative procedure will modernize oncology practice in such a way that we are future-proof and ready to offer adequate care to the growing group of individuals, like patient A, in our society in need for accurate cognitive diagnostics, referral and treatment.

Part II: Implementation of cognitive tests in studies on cognitive effects of endocrine therapy

Summary of findings

In **Chapter 5,** an overview is given of potential cognitive adverse effects of endocrine therapy (ET) and cyclin-dependent kinases 4 and 6 (CDK4/6) inhibitors in patients with hormone-receptor positive breast cancer (HR+ BC). First, mechanisms underlying potential cognitive effects of aforementioned treatments are described. Both ET and CDK4/6 inhibitors may affect cognition by direct (e.g., increased cytokine secretion) and/or indirect effects (e.g., increased fatigue). Subsequently, increasing evidence was found for cognitive adverse effects of ET, especially on verbal memory. No studies were found on cognitive effects of CDK4/6 inhibitors. It was concluded that 1) the literature indicates potential cognitive adverse effects of ET in pre-, peri- and postmenopausal women with HR+ BC, 2) cognitive effects might differ between ET agents, with potentially less favorable outcomes with tamoxifen use compared to AI use), and 3) longitudinal studies on cognitive effects of the combined ET-CDK4/6 inhibitors are highly needed to properly inform patients about potential short-term and long-term cognitive side effects.

In **Chapter 6**, building on a previous study of our group, a study is described in which short- and long-term cognitive effects of tamoxifen and exemestane are investigated in women with HR+ BC using data from the Tamoxifen and Exemestane Adjuvant Multinational (TEAM) trial: a prospective phase III randomized clinical trial. Cognition was assessed with traditional cognitive tests. It was found that women who underwent sequential treatment with tamoxifen and exemestane demonstrated a modest short- and long-term decline on several tested cognitive functions while women who underwent exemestane monotherapy did not. The cognitive adverse effects of tamoxifen and tamoxifen followed by

exemestane were found in absence of changes in self-reported cognitive complaints. It was concluded that tamoxifen alone and after switching to exemestane was associated with cognitive adverse effects, which is indicative of a carry-over effect on cognition.

In **Chapter 7,** a study is described on the long-term effect of risk-reducing salphingo-oophorectomy (RRSO) on cognitive functioning. This study was performed in women at high familial risk for ovarian cancer who had undergone RRSO in their pre- or postmenopause to reduce risk of ovarian cancer. The ACS and a questionnaire were administered to assess tested and self-reported cognition. No differences were found in tested cognition between women with a premenopausal RRSO ≤ age 45 or a postmenopausal RRSO ≥ age 54, nor between women with a RRSO between ages 41-45 or before age 40. Although more cognitive complaints were reported by women with a premenopausal RRSO compared to women with a postmenopausal RRSO, this difference disappeared in analyses including only women of comparable ages (60-70 years). It was concluded that approximately 18 years after RRSO, no association was found between premenopausal RRSO and cognition which is reassuring for women at high familial risk for ovarian cancer.

Discussion of the findings

Literature review

The literature overview on cognitive effects of endocrine therapy (ET) illustrates the need for further research. It was concluded that the literature indicates tested and self-reported cognitive adverse effects of ET, most consistently on verbal memory, in women with hormone receptor positive breast cancer (HR+ BC) and differential cognitive effects between ET agents (potentially less favorable outcomes with tamoxifen use compared to AI use). The conclusions are in line with earlier meta-analyses[e.g.,10] and reviews,[11] and

support the biological plausibility of effects of ET on cognition. However, it was also concluded that many earlier studies had suboptimal study designs. Prospective studies, preferably RCTs, are needed to investigate cognitive effects of ET with tested and self-reported cognition measurements pre- and post-treatment.

Cognitive effects of tamoxifen and exemestane

In line with the literature, tamoxifen was shown to have cognitive adverse effects, on verbal memory and executive function, while exemestane was not. Differential cognitive effects are not surprising given the different mechanisms of action. The finding of tamoxifen followed by exemestane, combined with the absence of any effect of exemestane mono-therapy, suggest a potential carry-over effect of tamoxifen, and is an addition to the literature which has not been investigated before. Given that clinical guidelines permit the choice between different ET regimens in both the adjuvant[12] and metastatic setting,[13] this is a clinically relevant finding that needs to be investigated further.

Long-term cognitive effects of RRSO

The finding that timing of surgical menopause and hormone replacement therapy is not associated with long-term cognitive functioning is in direct contrast with a subset of studies showing a long-term increased risk of cognitive impairment long after a premenopausal RRSO.[e.g.,14] As discussed in chapter 7, the difference in findings could potentially be explained by the lack of correction for confounders such as cancer history, hormone replacement therapy and depression, and inclusion of women with a RRSO for different indications in earlier studies. By correcting for the aforementioned factors the association between a premenopausal RRSO and long-term cognitive function disappeared. If replicated by future studies, this finding could be reassuring for women at familial risk of ovarian cancer who are recommended to undergo a RRSO.

Methodological considerations

When interpreting the results of this second part of the thesis, two methodological considerations should be taken into account.

In the TEAM study (chapter 6), it should be noted that the sample size at the long-term follow-up assessment was small. This reduced the statistical power of the long-term evaluation. A possible explanation for the low accrual rate was that the third cognitive assessment was not part of the original study protocol, resulting in a lower response rate. However, important to recognize is the fact that the sample size at baseline and short-term follow-up was adequate.

It should also be noted that the HARMOny study (chapter 7) was cross-sectional of nature. As such, this study did not have a baseline cognitive assessment before RRSO and did not measure cognitive functioning over time. To prove causal effects of any intervention, longitudinal studies are the preferred study design. It is recommended that future research assesses patients prospectively.

Implications

At this point, it is not possible to make clinical recommendations for ET use with regards to cognitive side effects. However, based on current results, clinicians including oncologists, occupational physicians, general practitioners and psychologists should be aware of cognitive problems in patients treated with ET, especially tamoxifen, and actively ask patients about cognitive functioning. An online survey in about 2400 participants showed that around 60% of women undergoing ET in the early BC and metastatic setting reported cognitive side effects but that one third of patients did not feel that side effects including cognitive problems and the impact on quality of life were taken seriously by clinicians.[15] This underscores that there is ample room for improvement.

In case of cognitive problems, patients can be referred to a clinical neuropsychologist who can administer cognitive testing and self-report measures for various patient-reported outcomes to evaluate the presence of cognitive disorders and pattern of cognitive strengths and weaknesses. Based on this information, patients can be given advice on how to deal with such complaints and cognitive rehabilitation interventions can be given.

It is clinically relevant that patients and caregivers as well as professionals have knowledge on the possibility of cognitive effects after ET. The Cognition Group at the Netherlands Cancer Institute has made a series of informational videos regarding this topic for patients as well as professionals, see: https://www.avl.nl/voorbereiding-afspraak/afdelingen-en-centra/centrum-voor-kwaliteit-van-leven-ondersteunende-zorg/cognitieve-problemen-bij-kanker-en-kankerbehandeling/.

Future research

The incidence, severity and causes of cognitive effects of ET should be further researched, also given their broad prescription in not only breast cancer but also in other cancers like prostate cancer. Given the different working mechanisms of ET agents, it is recommended to investigate cognitive effects per ET agents separately. Implementation of the ACS in these studies can facilitate large-scale data-collection. In chapter five, several recommendations for such studies are given. These studies should not only involve pre-and on-study cognitive testing but should also include cognitive data after completion of ET, as such data is currently practically absent. What's more, they should be of sufficient size given the fact that in general, long-term adherence to ET is relatively low due to treatment-related side effects.[16]

ET agents are likely to differ in cognitive effects because of differences in working mechanisms and in background characteristics of patient groups to which agents are primarily prescribed (for example, pre-

or postmenopausal women). Switching from one ET agent to another is common practice in adjuvant and metastatic settings, making this relevant to investigate as cognitive decline might remain present, diminish, or even worsen after switching. In addition, as ET is increasingly combined with other non-ET agents, these combination treatments should be included in future analyses. Basic scientific studies are needed to clarify causal mechanisms and origins of differences in cognitive effects between ET agents. Given the current finding of cognitive adverse effects of tamoxifen, which is in line with the existing literature, tamoxifen might be the first candidate to examine further in terms of underlying pharmacokinetics in relation to cognition.

As mentioned in chapter 6, the ACS is implemented in the Selecting the Optimal position of CDK4/6 Inhibitors in HR+ Advanced breast cancer (SONIA) study. In a side study, Evaluation of cognitive Functioning in patients with metastatic breast cancer treated with Endocrine or Combined Therapy (SONIA-EfFECT), it is investigated whether there is a difference in cognitive functioning over time in patients with metastatic HR+ BC who will undergo first-line treatment with either a non-steroidal AI combined with CDK4/6 inhibition (arm A SONIA study) or with a single agent non-steroidal AI (arm B SONIA study). In addition, the prevalence of cognitive impairment is examined in this population prior to treatment for metastatic disease and its relation to prior treatments received in the adjuvant setting. As of January 2023, 200 patients and 175 controls have completed the ACS twice.

As mentioned in chapter 7, the ACS is implemented in the TOTAM study to examine whether there is a difference between tested and self-reported cognitive functioning between BC patients who underwent tamoxifen for two years and controls without cancer. Additionally, the association is examined between on the one hand tamoxifen dose, plasma concentrations of tamoxifen and endoxifen, and on the other hand tested and self-reported cognition. As of January 2023, 99 patients have completed the ACS.

The data on ET in this thesis focused on female patients. As ET is also a commonly used treatment in men with prostate cancer and may exert adverse effects on cognitive function in this population as well,[17] further research should also focus on these patients as, despite the difference in gender, underlying disease, treatment and type of ET, it cannot be ruled out that some effects of ET on cognition could be comparable.

Data on cognitive effects of ET can propel forward research on how to intervene. Interventions using cognitive rehabilitation have shown to be effective in helping patients to better manage cognitive problems and increase quality of life. These therapies focus on learning compensatory skills and internal metacognitive strategies and have been shown to improve real-world functioning.[18] Other interventions that are currently under study focus on pharmacological or life style interventions. Exercise is potentially associated with improved cognitive functioning but more research is needed.[19-20] When it is known how to intervene in case of cognitive problems, patients at risk for these problems, such as patient A., could not only be informed but also better supported to address their complaints.

Ultimately, availability of a large collection of cognitive data collected across studies using the ACS might give the opportunity to aggregate data over several studies, facilitating large-scale analyses and subgroup and item-level analyses. Given the complexity of cognitive effects of ET due to additional influences of other cancer therapies and psychological effects of diagnosis and treatment, large datasets are needed. A 'biobank' that includes cognitive data, sociodemographic and clinical data, but also guidelines for the use of tests, items and parallel versions, can help harmonize assessments and facilitate the use of (international) studies.

Conclusion

Online cognitive assessments, such as the ACS, can play an important role in studies on cognitive functioning in oncology. The ACS has several advantages over other online/computerized cognitive assessments. One advantage is the high similarity of test design compared to traditional paper-and-pencil tests. This is in contrast to other assessments that often opted for a new test paradigm. Another advantage is the self-administrative nature of the ACS which aids in cost-effectiveness and patient-friendliness. Additionally, the ACS sets itself apart by the presence of usability, validity, reliability and normative data, and availability in several languages.

In this thesis, the ACS is further developed and cognitive tests are implemented in studies on cognitive effects of ET. Given its research and clinical opportunities, the ACS can strongly move us forward in the field of cancer and other diseases, and cognition. This will help us in our ultimate goal: Providing patients, such as Patient A. with our best possible care.

References

1. Feenstra HEM, Vermeulen IE, Murre JMJ, Schagen SB. Online self-administered cognitive testing using the Amsterdam Cognition Scan: establishing psychometric properties and normative data. *J Med Internet Res*. 2018;20(5):e192.
2. Madan CR. Exploring word memorability: how well do different word properties explain item free-recall probability? *Psychon Bull Rev*. 2021;28(2):583-595.
3. Bernstein LJ, McCreath GA, Komeylian Z, et al. Cognitive impairment in breast cancer survivors treated with chemotherapy depends on control group type and cognitive domains assessed: a multilevel meta-analysis. *Neurosci Biobehav Rev*. 2017;83:417-428.
4. Wefel JS, Kornet RL, Schagen SB. Systemically treated breast cancer patients and controls: An evaluation of the presence of noncredible performance. *J Int Neuropsychol Soc*. 2014;20(4):357-369.
5. Areán PA, Ly KH, Andersson G. Mobile technology for mental health assessment. *Dialogues Clin Neurosci*. 2016;18(2):163-169.
6. Pew Research Center. Mobile Fact Sheet. [Internet]. Available at: https://www.pewresearch.org/internet/fact-sheet/mobile/. [Accessed October 29th, 2022].
7. Agelink van Rentergem JA, Vermeulen IE, Lee Meeuw Kjoe PR, Schagen SB. Computational modeling of neuropsychological test performance to disentangle impaired cognitive processes in cancer patients. *J Natl Cancer Inst*. 2021;113(1):99-102.
8. Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. *Percept Mot*. 1958;8(3):271-276.
9. Sweet JJ, Heilbronner RL, Morgan JE, et al. American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *Clin Neuropsychol*. 2021;35(6):1053-1106.

10. Underwood E, Rochon P, Moineddin R, et al. Cognitive sequelae of endocrine therapy in women treated for breast cancer: a meta-analysis. *Breast Cancer Res Treat*. 2018;168(2):299-310.
11. Presciuttini R, Danesi R, Bocci G. May endocrine therapy be associated with cognitive impairment in breast cancer patients? *Clin Cancer Drugs*. 2018;5(2):75-86.
12. Burstein HJ, Lacchetti C, Anderson H, et al. Adjuvant endocrine therapy for women with hormone receptor-positive breast cancer: ASCO clinical practice guideline focused update. *J Clin Oncol*. 2019;37(5):423-438.
13. Burstein HJ, Somerfield MR, Barton DL, et al. Endocrine treatment and targeted therapy for hormone receptor–positive, human epidermal growth factor receptor 2–negative metastatic breast cancer: ASCO guideline update. *J Clin Oncol*. 2021;39(35):3959-3977.
14. Rocca WA, Lohse CM, Smith CY, Fields JA, Machulda MM, Mielke MM. Association of premenopausal bilateral oophorectomy with cognitive performance and risk of mild cognitive impairment. *JAMA Netw Open*. 2021;4(11):e2131448.
15. Berkowitz MJ, Thompson CK, Zibecchi LT, et al. How patients experience endocrine therapy for breast cancer: an online survey of side effects, adherence, and medical team support. *J Cancer Surviv*. 2021;15(1):29-39.
16. Yussof I, Tahir NAM, Hatah E, Shah NM. Factors influencing five-year adherence to adjuvant endocrine therapy in breast cancer patients: a systematic review. *Breast*. 2022;62:22-35.
17. Ryan C, Wefel JS, Morgans AK. A review of prostate cancer treatment impact on the CNS and cognitive function. *Prostate Cancer Prostatic Dis*. 2020;23(2):207-219.
18. Schagen SB, Tsvetkov AS, Compter A, Wefel JS. Cognitive adverse effects of chemotherapy and immunotherapy: are interventions within reach? *Nat Rev Neurol*. 2022;18(3):173-185.
19. Bernal JDK, Recchia F, Yu DJ, et al. Physical activity and exercise for cancer-related cognitive impairment among individuals affected by childhood cancer: a systematic review and meta-analysis. *Lancet Child Adolesc Health*. 2023;7(1):47-58.

20. Campbell K, Zadravec K, Bland KA, Chesley E, Wolf F, Janelsins MC. The effect of exercise on cancer-related cognitive impairment and applications for physical therapy: systematic review of randomized controlled trials. *Phys Ther*. 2020;100(3):523-542.

List of publications

Chapter 2 published as:
Lee Meeuw Kjoe PR, Agelink van Rentergem JA, Vermeulen IE, Schagen SB. How to Correct for Computer Experience in Online Cognitive Testing? *Assessment*. 2021;28(5):1247-1255.

All authors contributed to the study conception and design. Analyses were performed and the first draft of the manuscript was written by PL and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Chapter 3 published as:
Lee Meeuw Kjoe PR, Vermeulen IE, Agelink van Rentergem JA, van der Wall E, Schagen SB. Standardized item selection for alternate computerized versions of Rey Auditory Verbal Learning Test (-based) word lists. *Journal of Clinical and Experimental Neuropsychology*. 2022;44(9):681-701.

All authors contributed to the study conception and design. Analyses were performed and the first draft of the manuscript was written by PL and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Chapter 4 published as:
Agelink van Rentergem JA*, Lee Meeuw Kjoe PR*, Vermeulen IE, Schagen SB. Subgroups of cognitively affected and unaffected breast cancer survivors after chemotherapy: a data-driven approach [published online ahead of print January 14, 2023]. *Journal of Cancer Survivorship*. 2023.

All authors contributed to the study conception and design. Analysis were performed by JAvR and PL. The first draft of the manuscript was written by JAvR and PL and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Chapter 5 published as:
Lee Meeuw Kjoe PR, van der Wall E, Schagen SB. Endocrine therapy with or without CDK4/6 inhibitors in women with hormone-receptor positive breast cancer: what do we know about the effects on cognition? *Clinical Breast Cancer*. 2022;22(3):191-199.

All authors contributed to the study conception and design. The first draft of the manuscript was written by PL and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Lee Meeuw Kjoe PR, van der Wall E, Schagen SB. De mogelijke invloed van endocriene therapie en CDK4/6-remmers op cognitie bij het mammacarcinoom. *Nederlands Tijdschrift Voor Oncologie*. 2021;18(2):42-50.

All authors contributed to the study conception and design. The first draft of the manuscript was written by PL and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Chapter 6 published as:
Lee Meeuw Kjoe PR*, Kieffer JM*, Small BJ, Boogerd W, Schilder CM, van der Wall E, Meershoek-Klein Kranenbarg E, van de Velde CJH, Schagen SB. Effects of tamoxifen and exemestane on cognitive function in postmenopausal patients with breast cancer. *JNCI Cancer Spectrum*. In press.

CvdV, WB, EM, CS and SS were involved in the conception and design of the study. PL and JK assessed and verified the data. PL, JK, EvdW and SS drafted the manuscript. BS, WB, CS, EM and CvdV were involved in the final version of the manuscript. All authors have read and approved the manuscript.

Chapter 7 published as:

Terra L*, Lee Meeuw Kjoe PR*, Agelink Van Rentergem JA, Beekman MJ, Heemskerk-Gerritsen BAM, van Beurden M, Roeters van Lennep JE, van Doorn HC, De Hullu JA, Mourits MJE, van Dorst EBL, Mom CH, Slangen BFM, Gaarenstroom KN, van der Kolk LE, Collée JM, Wevers MR, Ausems MGEM, van Engelen K, van de Beek I, Berger LPV, van Asperen CJ, Gomez Garcia EB, Maas AHEM, Hooning MJ, van der Wall E, van Leeuwen FE**, Schagen SB**. Long-term effects of premenopausal risk-reducing salpingo-oophorectomy on cognition in women with high familial risk of ovarian cancer: A cross-sectional study [published online ahead of print January 30, 2023]. *BJOG: An International Journal of Obstetrics and Gynaecology*. 2023.

FvL, SS, AM, and MJH were involved in the conception and design of the study. LT, PL and JAvR assessed and verified the data. LT, PL, JAvR, SS and FvL drafted the manuscript. MB, MvB JRvL, HvD, JdH, EvD, CM, BS, KG, MM, LvdK, MC, MW, MA, KvE, IvdB, LB, CvA, EG, AM, MH, BHK and EvdW were involved in the final version of the manuscript. All authors have read and approved the manuscript.

Other publication

Agelink van Rentergem JA, Vermeulen IE, Lee Meeuw Kjoe PR, Schagen SB. Computational modeling of neuropsychological test performance to disentangle impaired cognitive processes in cancer patients. *Journal of the National Cancer Institute*. 2021;113(1), 99-102.

*First authors contributed equally.

**Last authors contributed equally.

Nederlandse samenvatting

Steeds meer onderzoek laat zien dat sommige mensen met kanker tijdens of na de behandeling (zoals bijvoorbeeld chemotherapie en endocriene therapie) te maken krijgen met cognitieve problemen. Cognitieve problemen zijn problemen met het denken, zoals verminderde concentratie, geheugenproblemen en moeite met plannen. Cognitieve problemen kunnen leiden tot moeilijkheden bij de terugkeer naar werk en in het dagelijks functioneren. Hoewel er tegenwoordig steeds meer aandacht is voor cognitieve problemen tijdens of na chemotherapie, is nog veel onduidelijk over mogelijke cognitieve bijwerkingen van andere therapievormen, zoals bijvoorbeeld endocriene therapie.

Om het inzicht in en de aanpak van cognitieve problemen bij kanker en kankerbehandeling te verbeteren, is enkele jaren geleden de Amsterdam Cognition Scan (ACS) ontwikkeld. De ACS bestaat uit zeven online cognitieve tests die patiënten zelfstandig vanuit huis op hun computer of laptop kunnen maken. De ACS meet verschillende cognitieve functies, zoals aandacht, geheugen en executieve functies. Alle tests zijn gebaseerd op veelgebruikte pen-en-papier tests. Door de online testafname is de ACS een gebruiksvriendelijke tool die grootschalig onderzoek binnen de oncologie mogelijk maakt.

Voordat de ACS grootschalig en internationaal ingezet kan worden, dient de ACS verder ontwikkeld te worden. Het eerste doel van dit proefschrift was doorontwikkeling van de ACS. Het tweede doel was implementatie van cognitieve tests in studies naar cognitieve effecten van endocriene therapie.

In dit hoofdstuk worden de resultaten van de in dit proefschrift beschreven onderzoeken kort samengevat. Dit wordt separaat gedaan voor beide doelstellingen.

Deel I: Doorontwikkeling van de Amsterdam Cognition Scan

Samenvatting van de bevindingen

Een belangrijke doorontwikkeling van de ACS is het verkrijgen van inzicht in de wijze waarop rekening kan worden gehouden met computerervaring bij de interpretatie van scores op gecomputeriseerde tests. Scores op gecomputeriseerde cognitieve tests worden mogelijk beïnvloed door de mate van computerervaring. Het is echter nog onduidelijk hoe een correctie voor computerervaring toegepast dient te worden. In **hoofdstuk 2** wordt een studie beschreven waarin is onderzocht hoe bij het analyseren van online cognitieve testscores gecorrigeerd kan worden voor computerervaring. Computerervaring werd gemeten door het testen van computervaardigheden (typesnelheid en klikken en slepen met de computermuis) en zelf-gerapporteerde frequentie van computergebruik (aantal uren computergebruik per week). De invloed van beide maten werd onderzocht op zowel de ACS als de equivalente traditionele pen-en-papier tests. De resultaten lieten zien dat betere geteste computervaardigheden samenhingen met betere scores op zowel de ACS als op de traditionele, voornamelijk op snelheid-gebaseerde tests. Meer zelf-gerapporteerde computergebruik hing samen met betere scores op de op snelheid-gebaseerde ACS-tests, maar niet met scores op de meeste traditionele tests. Geconcludeerd werd dat voor het rekening houden met computerervaring bij het maken van online cognitieve tests, het beste gebruik kan worden gemaakt van een zelf-rapportage maat. Omdat computergebruik na diagnose mogelijk geen betrouwbare weergave geeft van eerdere computerervaring (immers, het computergebruik kan sterk veranderen rondom diagnose en behandeling), wordt aanbevolen te corrigeren voor computergebruik vóór de diagnose in plaats van voor actueel computergebruik.

Een andere belangrijke doorontwikkeling van de ACS is een standaardprocedure voor de ontwikkeling van de verbale geheugentest, de enige talige subtest van de ACS. Middels deze standaardprocedure kan de ACS vertaald worden naar andere talen. De verbale geheugentest van de ACS maakt gebruik van een woordenlijst die deelnemers gevraagd wordt te onthouden. Woordenlijsten dienen onderling vergelijkbaar te zijn voor een betrouwbare meting van het geheugen. In **hoofdstuk 3** wordt een studie beschreven naar woordselectiecriteria voor nieuwe woordenlijsten voor de verbale geheugentest in de ACS. Dertien woordselectiecriteria werden geïdentificeerd, twee nieuwe Amerikaans-Engelse woordenlijsten en een Nederlandse woordenlijst werden ontwikkeld en de criteria werden gevalideerd door online scores op de nieuwe woordenlijsten te vergelijken met scores op veelgebruikte traditionele woordenlijsten. Het bleek dat toepassing van de woordselectiecriteria leidde tot twee zeer vergelijkbare nieuwe Amerikaans-Engelse woordenlijsten. De nieuwe Nederlandse woordenlijst was op alle uitkomsten zeer vergelijkbaar met de traditionele Nederlandse versie. Geconcludeerd werd dat de woordselectiecriteria de ontwikkeling van nieuwe woordenlijsten in de ACS faciliteert, ook in verschillende talen.

Een laatste belangrijke doorontwikkeling van de ACS is een verbetering van de interpretatie van testscores. Er wordt in het algemeen aangenomen dat slechts een subgroep van patiënten met kanker te maken krijgt met cognitieve problemen. Het is echter nog onduidelijk of dit daadwerkelijk een subgroep van patiënten betreft die verschilt van andere subgroepen op één of meer kenmerken (zoals bijvoorbeeld leeftijd) of dat deze 'subgroep' ontstaat door de typering van testscores als normaal of afwijkend. In **hoofdstuk 4** wordt een studie beschreven bij borstkankerpatiënten die behandeld zijn met chemotherapie waarin het bestaan van cognitieve subgroepen werd onderzocht. De resultaten van subgroepanalyses werden vergeleken met traditionele normatieve vergelijkingen (vergelijkingen van scores met die van een grote groep mensen met vergelijkbare leeftijd en opleidingsniveau zonder ziekte). De resultaten toonden consequent twee

subgroepen middels toepassing van verschillende clustermethoden: De ene had normale cognitieve scores, terwijl de andere lagere scores had. De groep met lagere scores bestond uit 45% van de patiënten. De resultaten van het subgroepclustermodel kwamen overeen met de traditionele normatieve vergelijkingsmethode die twee kleine afwijkende scores op tests van verschillende cognitieve domeinen vereist. Geconcludeerd werd dat het percentage patiënten dat cognitief is aangedaan mogelijk hoger is dan verwacht en dat mogelijk een minder strenge grens gebruikt dient te worden in deze populatie voor de identificatie van afwijkende testscores.

Deel II: Implementatie van cognitieve tests in studies naar cognitieve effecten van endocriene therapie

Om de bijwerkingen van verschillende kankerbehandelingen op het cognitief functioneren te onderzoeken, moet de afname van cognitieve tests geïmplementeerd worden in studies naar verschillende vormen van kankerbehandeling. Eén van die vormen waarbij onderzoek naar cognitieve effecten nodig is, is endocriene therapie. Endocriene therapie is een belangrijke behandelvorm van hormoongevoelige borstkanker. Vijfentachtig procent van alle vrouwen met invasieve borstkanker in Nederland heeft een hormoongevoelige borstkanker. Dit type borstkanker wordt hormoongevoelig genoemd omdat bij deze vrouwen de borstkanker groeit door aanwezigheid van oestrogenen. Endocriene therapie gaat de werking van oestrogenen tegen, zoals tamoxifen (welke oestrogeenreceptoren blokkeert) en aromataseremmers (welke de synthese van oestrogenen tegengaat). Oestrogenen spelen echter een belangrijke rol bij het normale functioneren van de hersenen. Endocriene therapie kan dus mogelijk gepaard gaan met cognitieve bijwerkingen.

Samenvatting van de bevindingen

Endocriene therapie (ET) wordt in toenemende mate gecombineerd met 'cycline-afhankelijke kinasen 4 en 6' (CDK4/6)-remmers bij vrouwen met uitgezaaide borstkanker. Om deze reden wordt in **hoofdstuk 5** een overzicht gegeven van potentieel nadelige effecten op het cognitief functioneren van ET en CDK4/6-remmers bij patiënten met hormoongevoelige borstkanker. Eerst worden mechanismen beschreven die ten grondslag kunnen liggen aan cognitieve effecten van bovengenoemde behandelingen. Zowel ET als CDK4/6-remmers kunnen de cognitie beïnvloeden door directe (bv. verhoogde cytokinesecretie) en/of indirecte effecten (bv. vermoeidheid). Voorts wordt steeds meer bewijs gevonden voor nadelige cognitieve effecten van ET, vooral op het verbale geheugen. Er zijn nog geen studies uitgevoerd naar cognitieve effecten van CDK4/6-remmers met objectieve cognitiemetingen. De conclusie was dat 1) de literatuur wijst op nadelige cognitieve effecten van ET bij pre-, peri- en postmenopauzale vrouwen met borstkanker, 2) cognitieve effecten kunnen verschillen tussen ET-middelen, met mogelijk minder gunstige uitkomsten bij tamoxifengebruik in vergelijking met aromataseremmersgebruik (zoals exemestaan), en 3) longitudinale studies naar cognitieve effecten van de gecombineerde ET-CDK4/6-remmers nodig zijn om vast te stellen of er cognitieve bijwerkingen zijn om vervolgens patiënten te kunnen informeren over potentiële cognitieve bijwerkingen op korte en lange termijn.

In **hoofdstuk 6** wordt een studie beschreven waarin de korte- en langetermijn cognitieve effecten van tamoxifen en exemestaan worden onderzocht bij vrouwen met hormoongevoelige borstkanker aan de hand van gegevens van de Tamoxifen and Exemestane Adjuvant Multinational (TEAM) trial: een prospectieve fase-III gerandomiseerde klinische studie. Deelnemers van de TEAM trial werden gerandomiseerd tussen enerzijds 2.5 tot 3 jaar behandeling met tamoxifen gevolgd door 2 tot 2.5 jaar behandeling met exemestaan of anderzijds 5 jaar behandeling met exemestaan.

Cognitief functioneren werd driemaal over de tijd gemeten middels traditionele pen-en-papier cognitieve tests. Het bleek dat vrouwen die een sequentiële behandeling ondergingen met tamoxifen en exemestaan een lichte achteruitgang op korte en lange termijn vertoonden op verschillende gemeten cognitieve functies, terwijl vrouwen die exemestaan monotherapie ondergingen geen achteruitgang lieten zien. De nadelige cognitieve effecten van tamoxifen en tamoxifen gevolgd door exemestaan werden gevonden in afwezigheid van veranderingen in zelf-gerapporteerde cognitieve klachten. Geconcludeerd werd dat tamoxifen alleen en na switch naar exemestaan geassocieerd was met cognitieve bijwerkingen, wat duidt op een carry-over effect op cognitie. Meer studies zijn nodig om de klinische relevantie van de cognitieve effecten te onderzoeken.

Een andere vorm van ET is een Risico Reducerende Salpingo-Ovariëctomie (RRSO). Een RRSO wordt aanbevolen op relatief jonge leeftijd bij draagsters van een BRCA1/2 mutatie wegens een verhoogd risico op eierstokkanker. Omdat de RRSO wordt aanbevolen vóór de menopauze komen deze vrouwen vervroegd in de overgang waardoor de aanmaak van oestrogenen fors afneemt. Enkele eerdere studies suggereerden een verhoogde kans op cognitieve problemen op latere leeftijd bij vrouwen die vóór de menopauze een RRSO hebben ondergaan. In **hoofdstuk 7** wordt een studie beschreven naar het langetermijneffect van een RRSO op het cognitief functioneren. Deze studie werd uitgevoerd bij vrouwen met een familiair verhoogd risico op eierstokkanker die vóór of na de menopauze een RRSO hadden ondergaan. De ACS en een vragenlijst werden afgenomen om objectief en subjectief cognitief functioneren te meten. Er werden geen verschillen gevonden in objectief cognitief functioneren tussen vrouwen met een premenopauzale RRSO ≤ 45 jaar of een postmenopauzale RRSO ≥ 54 jaar, noch tussen vrouwen met een RRSO tussen 41-45 jaar of vóór 40 jaar. Hoewel meer cognitieve klachten werden gerapporteerd door vrouwen met een premenopauzale RRSO in vergelijking met vrouwen met een postmenopauzale RRSO, verdween dit verschil in analyses bij vrouwen

van vergelijkbare leeftijd (60-70 jaar). De conclusie was dat ongeveer 18 jaar na de RRSO geen verband werd gevonden tussen een premenopauzale RRSO en cognitie, hetgeen geruststellend is voor vrouwen met een familiair verhoogd risico op eierstokkanker.

Conclusie

Er is meer onderzoek nodig naar de cognitieve bijwerkingen van kanker en kankerbehandelingen. Online cognitieve tests, zoals de ACS, kunnen bij deze studies een belangrijke rol spelen. In dit proefschrift wordt de ACS doorontwikkeld en worden cognitieve tests geïmplementeerd bij studies naar cognitieve effecten van endocriene therapie. Steeds meer mensen overleven kanker of leven langer met de ziekte door vroege opsporing en effectievere behandelingen. Dit maakt de optimalisatie van de kwaliteit van leven voor deze groeiende groep in onze samenleving van zeer groot belang.

Dankwoord

Mijn dankwoord gaat ten eerste uit naar alle mensen die deel hebben genomen aan het in dit proefschrift beschreven onderzoek. Zonder hun onbaatzuchtige deelname is wetenschappelijk onderzoek niet mogelijk. Voorts wil ik iedereen bedanken die op welke wijze dan ook heeft bijgedragen aan de totstandkoming van dit proefschrift. Een aantal mensen wil ik hierbij in het bijzonder benoemen:

Sanne, dank voor je vertrouwen en begeleiding de afgelopen jaren. Hoe druk je ook bent, ik heb altijd het gevoel gehad dat ik je kon benaderen voor vragen en je had altijd goede oplossingen paraat voor elke situatie. We hebben wat afgereisd: Sydney, Stockholm, Lille, New York, Denver en San Antonio, hoe leuk! Het was een eer om bij jou en Elsken te promoveren.

Elsken, dank voor jouw aanstekelijke enthousiasme, deskundigheid en zorgzaamheid. Ik ben onder de indruk van jouw kennis. Na elke meeting met jou had ik extra motivatie om nog harder te werken, en meetings waren altijd gezellig – zoals de cappuccino meetings op het NKI. Grotendeels door jou staat cognitie meer op de kaart in de oncologie. Hopelijk heb je wat gehad aan mijn “Zweedse modetips”.

Ivar, dank voor jouw enthousiasme en creativiteit. Jij legt verbanden tussen zaken die anderen niet direct zien. Mede daardoor en door jouw positiviteit waren overleggen met jou altijd interessant.

Dank aan de leden van mijn promotiecommissie voor het plaatsnemen in de commissie en het beoordelen van het manuscript.

Dank aan alle co-auteurs voor de samenwerkingen. Dank aan Joost, Lara en Jacobien voor onze artikels samen. Leuk om te zien hoeveel ik kon leren van jullie door samen aan een manuscript te werken (inclusief nieuwe statistische programma’s en advanced Word-skills).

Joost, kamergenoot en paranimf, dank voor jouw begeleiding, gezelligheid en vele interessante gesprekken. Samenwerken met jou was ware verrijking van de hersenen- zowel cognitief als fysiek (minstens drie keer per dag 7 verdiepingen op en af). Mooi hoe jij de recentste onderzoeksbevindingen in jouw eigen dagelijks leven onderzoekt à la N=1.

Maryse, a.k.a. Excel-guru, dank voor zowel jouw onderzoeksassistentschap als jouw overname van het ACS project als promovenda. Door jouw can-do attitude weet ik zeker dat met jou aan het roer de ACS optimaal zal worden doorontwikkeld.

Nick, dank voor jouw expertise en samenwerking. Door jouw werk als programmeur is de technologie achter de ACS sterk en solide.

Dank aan Kelly, Marina en Esmée voor jullie interessante masterscripties over de ACS – jullie data leveren een belangrijke bijdrage aan de doorontwikkeling van de ACS.

Voorts, dank aan de filmcrew Kick, Niels en Matthijs, visagiste Mirjam, actrices en (stem-)acteurs Anna, Paula, Melline, Tamara, Jennie, Claire, Anders, Thomas, Chris en Iben voor jullie bijdrage aan de mooie instructie- en pauzevideo's van de ACS.

Tack så mycket Yvonne, Maria, Renske en Anna for the Swedish ACS collaboration (and the hospitality in Karolinska), thanks so much Jeff and Kyle for the American one, mange tak Cecilie for the Danish collaboration en dank aan Florien (ha, gewoon in het Nederlands) voor de Britse samenwerking.

Dank aan de cognitiegroep (Sanne, Michiel, Marianne, Jacobien, Joost, Emmie, Kimberly, Kete, Annette en anderen) voor leuke meetings, borrels en retreat. Dank aan alle PSOE collega's voor interessante meetings en een fijne sfeer op H8. Dank aan Joost, Kimberly en Johanna voor de gezelligheid

als kamergenoten– soms kwamen we haast niet aan werken toe omdat we zoveel leuke gesprekken hadden. Dank aan Ellen, Katinka en Emine voor een door covid alternatief maar toch gezellig PSOE Party Committee jaar.

Dank aan Liselore, Josephine en Noor voor het gezamenlijk organiseren van een geslaagde, eerste BOOG Young Investigator meeting. Dank aan Agnes en Elise voor deze kans – hopelijk is dit de start van een mooie reeks nieuwe events. Dank ook, mede aan Gabe, Inge en Jeany, voor de SONIA-EfFECT samenwerking, en mede aan Judith, voor de DIRECT II samenwerking.

Dank aan alle collega's bij iPractice voor de mogelijkheid om mijn proefschrift af te ronden naast de klinische werkzaamheden. Een speciale dank aan Silvija, voor jouw waardevolle support. En ik kijk uit naar onze onderzoeksresultaten bij iPractice naar de langetermijneffecten van de behandelingen.

De bakermat van mijn wetenschappelijke ambities lag in de VU. Dr. Deijen, mentor tijdens mijn master, dank voor de mogelijkheid om mijn masteronderzoek in Cambridge uit te voeren en de gezellige begeleiding tijdens en na. Dank aan Sietske en Roos voor een mooie en leerzame introductie bij het VU Alzheimercentrum in het reilen en zeilen van een onderzoek.

Dank aan alle vrienden, in het bijzonder Walter, Wouter, Dominiek en Christiaan voor jullie jarenlange vriendschappen vanaf de kleuterschool en middelbare school. Liza, jij woont na Cambridge inmiddels al wat jaartjes in Australië, maar het is altijd vanouds gezellig als je weer in het land bent. Claartje, dank voor jouw goede tips voor de klinische werksetting en het is altijd leuk om weer met je bij te kletsen. Dank aan alle ooms, tantes, neven en nichten (Joëlle en Maud, inmiddels eveneens praktisch mijn buren, en Hanna, ik kijk altijd met heel veel trots naar jouw films en voorstellingen) voor jullie interesse en steun tijdens het traject. Een speciale dank aan mijn grote academische voorbeelden: Ellen, mijn tante, en Eric, mijn oom, voor

jullie aansporing en waardevolle adviezen in zowel wetenschappelijke context alsook daarbuiten.

Dank aan Edward, mijn broer (wat een eer dat de adviseur van de plaatsvervangend secretaris-generaal van Volksgezondheid mijn paranimf is!), en Earl en Annelies, mijn wijze en lieve ouders, die mij altijd gestimuleerd en gesteund hebben. Wij lachen wat af met elkaar en staan altijd klaar voor elkaar – dat warmt mijn hart. En natuurlijk aan mijn kleine Nino en Indy, trouwste viervoeters in de hemel en op aarde, die mij vaak lieten zien hoe betrekkelijk alles ook maar is.

And last but not least, thank you to my sweet Fiona, not only for the cover (what a beautiful painting it has become) but also for all the support while working on both our dissertations. I look forward to read yours. 我愛你.

About the author

Philippe Romano Lee Meeuw Kjoe was born on the 25th of June, 1991 in Amsterdam, the Netherlands. He attended secondary school at the Coornhert Lyceum in Haarlem and graduated Gymnasium in 2009. In the same year, Philippe started his bachelor in Psychology at the VU University Amsterdam. In 2012, he graduated with honours and started a research master in Cognitive Neuropsychology at the VU University. Philippe conducted his master research at the University of Cambridge in the United Kingdom. During his master, he started a second master in Clinical Neuropsychology at the VU. In 2015, Philippe obtained his master degree from both masters. He started working as a research assistant and as a neuropsychologist in the Alzheimer Center of the VUmc hospital. In 2018, he started his PhD-project at the Netherlands Cancer Institute – Antoni van Leeuwenhoek, resulting in the current thesis. From 2022, Philippe works as a psychologist at iPractice in the basic mental health care setting. His aim is to combine clinical work with research to strive to his mission to offer patients the best possible care.