

Original Research Report

Statistical Model of Dynamic Markers of the Alzheimer's Pathological Cascade

Steve Balsis, PhD,^{1,*} Lisa Geraci, PhD,¹ Jared Benge, PhD,² Deborah A. Lowe, PhD,¹ Tabina K. Choudhury, MS,¹ Robert Tirso, BS,¹ Rachelle S. Doody, MD, PhD,³ for the Alzheimer's Disease Neuroimaging Initiative

¹Department of Psychological and Brain Sciences, Texas A&M University, College Station. ²Baylor Scott and White Neurosciences Institute, Temple, Texas. ³Department of Neurology, Baylor College of Medicine, Alzheimer's Disease and Memory Disorders Center, Houston, Texas.

*Address correspondence to: Steve Balsis, PhD, Department of Psychological and Brain Sciences, Texas A&M University College Station, TX 77843. E-mail: balsis@tamu.edu

Received: February 2, 2017; Editorial Decision Date: November 2, 2017

Decision Editor: Shevaun Neupert, PhD

Abstract

Objectives: Alzheimer's disease (AD) is a progressive disease reflected in markers across assessment modalities, including neuroimaging, cognitive testing, and evaluation of adaptive function. Identifying a single continuum of decline across assessment modalities in a single sample is statistically challenging because of the multivariate nature of the data. To address this challenge, we implemented advanced statistical analyses designed specifically to model complex data across a single continuum.

Method: We analyzed data from the Alzheimer's Disease Neuroimaging Initiative (ADNI; $N = 1,056$), focusing on indicators from the assessments of magnetic resonance imaging (MRI) volume, fluorodeoxyglucose positron emission tomography (FDG-PET) metabolic activity, cognitive performance, and adaptive function. Item response theory was used to identify the continuum of decline. Then, through a process of statistical scaling, indicators across all modalities were linked to that continuum and analyzed.

Results: Findings revealed that measures of MRI volume, FDG-PET metabolic activity, and adaptive function added measurement precision beyond that provided by cognitive measures, particularly in the relatively mild range of disease severity. More specifically, MRI volume, and FDG-PET metabolic activity become compromised in the very mild range of severity, followed by cognitive performance and finally adaptive function.

Conclusion: Our statistically derived models of the AD pathological cascade are consistent with existing theoretical models.

Keywords: Alzheimer's disease, Brain, Biomarkers, Cognition, Dementia, Item response theory, Magnetic resonance imaging, Model, Psychometrics, Statistics.

Introduction

Alzheimer's disease (AD) is a progressive disease with a continuum of severity ranging from preclinical (asymptomatic) forms to mild cognitive impairment (MCI; [Albert et al., 2011](#); [Petersen et al., 2001](#)) and ultimately dementia ([McKhann et al., 2011](#)). Although the emergence or increased severity

of certain clinical symptoms (e.g., deterioration of the ability to complete daily activities) defines the conversion between these diagnostic stages (i.e., preclinical, MCI, dementia), disease severity itself is a continuum with a lack of sharp borders.

Along this continuum, various imaging, cognitive, and functional changes emerge. For example, AD is associated

with progressive neurodegeneration, which results in volumetric losses in particular brain regions, as measured by structural neuroimaging (e.g., magnetic resonance imaging [MRI] scans). These volumetric changes or cerebral atrophy are thought to begin in the entorhinal cortex of the medial temporal lobe (Devanand et al., 2012), followed by the hippocampus (Desikan et al., 2010; Schuff et al., 2009), caudate nucleus, amygdala, parahippocampus, and posterior cingulate (Johnson, Fox, Sperling, & Klunk, 2012). These sites of neurodegeneration are affected early in the disease, such that by the time patients receive a diagnosis of dementia, entorhinal cortex volumes are already reduced by approximately 20–30% and hippocampal volumes are reduced by approximately 15–25% (Johnson et al., 2012). As the disease progresses, the lateral ventricles expand and atrophy spreads to the temporal neocortex, temporoparietal association areas, and frontal lobe (Scahill, Schott, Stevens, Rossor, & Fox, 2002; Whitwell et al., 2007).

AD is also characterized by changes in fluorodeoxyglucose positron emission tomography (FDG-PET) glucose metabolism that consist of synaptic dysfunction and reduced regionalized cerebral activity. Declines in cerebral activity correspond to a certain extent with the pattern of MRI volumetric losses, generally beginning in the hippocampus, posterior cingulate, and precuneus and spreading to temporoparietal cortices, frontal regions, and even occipital cortex (Chen et al., 2010; Mosconi et al., 2008).

In addition to volumetric and functional changes in the brain, AD is associated with decreases in cognitive performance, as measured by standard neuropsychological assessments (Sperling & Johnson, 2012; Wilson et al., 2012). Typically, early cognitive declines occur in episodic memory, followed by declines in attentional control, executive function, language, and visuospatial ability (Carter, Caine, Burns, Herholz, & Lambon Ralph, 2012). These cognitive changes eventually lead to declines in everyday functioning, which are hallmarks of dementia (Reisberg et al., 2001).

Given that AD exists and is measured across all of these different domains, comprehensive measurement models are needed to quantify how well the different marker domains indicate or reflect disease severity throughout the AD continuum. Quantifying and measuring the somewhat abstract concept of “AD severity” to create these models presents certain methodological and statistical challenges. However, many of these challenges can be met using item response theory (IRT), a statistical framework that is suitable for analyzing continuous constructs such as AD severity.

One of the fundamental challenges to creating a measurement model of AD severity is the need to define a single continuum of disease severity that extends from overt signs of the disease back through “normal” functioning by leveraging indicators of the disease across different assessment modalities (MRI volume, FDG-PET metabolic activity, cognitive performance, and adaptive function). In IRT terms, an indicator can be thought of as a particular variable through which an individual’s standing on a latent

construct can be measured or conveyed (Embretson & Reise, 2000). A latent construct can be defined as a phenomenon that exists but is difficult to observe directly. In lieu of direct observations, indicators of the construct are identified and grouped to represent the presence and level of the construct. Latent constructs can range from psychiatric phenomena such as depression to astronomical constructs such as gravity. Both of these constructs exist, but we know about them indirectly via specified indicators. Alzheimer’s disease is no different in this regard. It exists and we know about it through a variety of indicators. Existing measurement models often use data from only one assessment modality (e.g., MRI volume, FDG-PET metabolic activity, cognitive performance, or adaptive function) to represent the latent construct of AD and then use another as an external referent, the variable chosen to provide a metric for the latent construct (Johnson & Meade, 2007). This approach is used partly to preserve the unidimensionality of the construct of interest. Adding indicators across modalities introduces multidimensionality, which reduces unidimensional model fit. However, AD does not manifest solely across two domains of function and is not solely measured using two discrete assessment modalities. Defining the latent construct of AD severity using only data from one domain (e.g., MRI volume) and describing its association to another (e.g., adaptive function) limits our conceptualization and representation of the complex pathological processes that comprise AD severity. As such, the resulting challenge is to create a model of AD that includes indicators that exist across multiple assessment modalities, while statistically linking each indicator to the same disease continuum. To meet this challenge, we can define the latent construct of interest (AD severity) within one type of assessment modality, such as cognition, and then link indicators from the remaining assessment modalities to that same continuum in a manner that helps us index (rather than define) that continuum.

Using indicators from different types of assessments to index the same AD continuum offers key benefits. First, it enables investigators to measure the continuum thoroughly; for example, indicators from one assessment such as cognitive performance might be most sensitive to detecting AD-related changes in the moderate range of disease severity. Meanwhile, indicators from another type of assessment such as MRI volume might be more sensitive to the disease in a relatively milder form. Using both sets of indicators provides greater coverage of the latent continuum. Second, this approach enables statistically coherent comparisons among the indicators. When cognitive performance indicators and MRI volume indicators are mapped to the same latent continuum, for example, investigators can quantify the extent to which these markers indicate the disease and determine the level of severity at which each marker is sensitive to the continuum. Whereas previous studies have used variables across assessments as external referents to draw general conclusions about the associations among them

(Vemuri, Wiste, & Weigand, 2009), the current approach provides additional utility.

There is much to be gained not just measurement-wise by using the current statistical approach to measuring Alzheimer's disease progression. A comprehensive theoretical model of Alzheimer's disease progression requires that we consider each of these markers and delineate when they begin to change in a way that indicates the same disease continuum. For example, in a recent publication by the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for AD (Sperling & Johnson, 2012), a framework was proposed that depicts the hypothesized progression of disease-related abnormalities across different markers (e.g., brain structure, synaptic dysfunction, cognition). According to this general model, markers of synaptic dysfunction and brain structure change early, followed by changes in cognition and function (Jack et al., 2010, 2013).

Theoretical models are extremely valuable in helping to organize the existing data and provide testable hypotheses regarding the progression of AD. Yet, existing models are based on findings from a variety of studies that often restrict examination of AD markers to a single assessment (e.g., MRI volume alone) rather than combining multiple markers in a single study that are all statistically linked to the same disease continuum. Using a single large sample from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, we statistically determined the measurement capacities of four AD marker assessments (MRI volume, FDG-PET metabolic activity, cognitive performance, and adaptive function) and derive a model of the AD cascade across these assessments.

We know that the AD cascade is detectable using various assessment modalities of disease indicators. Approximations of the manifestations of these indicators have a sigmoidal relationship to the continuum of the disease (Jack et al., 2013). Thus, to model the disease, we need statistical machinery that can identify a core defining dimension of the disease within one type of assessment, and the ability of the other assessments to indicate this dimension with sigmoidal monotonically increasing function curves. IRT provides the needed machinery to model the disease in this way. IRT uses multiple markers to statistically define a single latent dimension (in this case, AD severity) and simultaneously determine the degree to which individual markers are related to that dimension with sigmoidal curves (Jack et al., 2013).

The current study used IRT-based analyses of data from the ADNI to statistically model the progression of Alzheimer's disease. The first aim of the study was to establish a measurement model across assessment tools and quantify the degree to which indicators from each assessment provide AD-related information. The second aim was to use this measurement model to analyze where these markers become dysfunctional across the spectrum of disease severity, from preclinical AD to full dementia.

Methods

Data used in the preparation of this article were obtained from the ADNI database, adni.loni.usc.edu. The ADNI was launched in 2003 as a public-private partnership. The initial goal of ADNI was to recruit 800 participants but ADNI has been followed by two other initiatives, ADNI-GO and ADNI-2. To date these three protocols have recruited over 1,500 adults, ages 55–90, to participate in the research. The sample consists of older adults who are cognitively healthy, those with early or late MCI, and those with early AD. Demographic and clinical data used for this study were downloaded from the ADNI data repository (adni.loni.usc.edu) on May 28, 2014. Data for the current analyses come from individuals who completed baseline assessments and had complete data for key cognitive and brain variables described below ($n = 1,056$).

Participants

The analyses for the present study used baseline data from 1,056 participants (470 female, 45%) enrolled across all three ADNI phases. Participants were an average of 72.87 years old ($SD = 6.99$), highly educated ($M = 16.09$, $SD = 2.77$ years), and the majority identified their race as white ($n = 970$, 92%). Other races represented include black or African American ($n = 52$, 5%), Asian ($n = 17$, 2%), American Indian or Alaskan native ($n = 2$, 0%), and Native Hawaiian or Other Pacific Islander ($n = 2$, 0%); 11 participants (1%) reported that they were more than one race, and 2 participants (0%) indicated unknown race. Thirty-six (3%) participants reported their ethnicity as Hispanic or Latino; 1,012 (96%) reported that they were not Hispanic or Latino, and 8 (1%) were unknown.

Baseline diagnoses represented a range of cognitive impairment: 341 (32%) were cognitively normal, 560 (53%) had MCI, and 155 (15%) had presumed Alzheimer's dementia. We included the cognitively normal (CN) participants so that we could model the disease from the continuum of normal aging to pathological aging. In the ADNI, CN participants served as the controls and showed no signs of MCI or dementia. CN participants had normal cognition (defined as Clinical Dementia Rating [CDR-SOB] = 0, Mini Mental State Exam [MMSE] between 24–30, and Wechsler Memory Scale-Revised [WMS-R; Wechsler, 1987] Logical Memory II subscale score above education-adjusted cut-offs). MCI participants had a subjective memory concern and significant amnesic dysfunction (defined by CDR-SOB = 0.5 plus an abnormal score on the WMS-R Logical Memory II subscale). However, MCI participants had sufficiently preserved functional abilities and global cognition (MMSE score from 24–30), such that they did not meet criteria for AD. Participants were diagnosed with probable AD if they met National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria (McKhann et al., 1984). At the time of the ADNI

diagnosis, AD participants demonstrated deficits in their global cognition, memory functioning (based on scores on the WMS-R Logical Memory II subscale), and showed significant concerns with memory (reported by the participant, study partner, or clinician). Finally, participants were excluded from the analyses if they had a history of significant neurologic disease (including multi-infarct dementia and subdural hematoma).

Measures

We assessed four marker assessment modalities of AD including MRI volume, FDG-PET metabolic activity, cognitive performance, and adaptive function. In the ADNI sample, participants were evaluated using structural MRI scans to assess neuroanatomical volume, fluorodeoxyglucose positron emission tomography (FDG-PET) to assess regional and global glucose metabolism, neuropsychological tests to assess cognitive performance, and the Functional Activities Questionnaire (FAQ; Pfeffer, Kurosaki, Harrah, Chance, & Filos, 1982) to assess activities of daily living (IADL) impairment. The procedures used for each of these domains are briefly described below (full description online at adni.loni.usc.edu).

MRI volume

Structural MRI scans enable volumetric measurements of neuroanatomical regions, which can indicate patterns of volumetric changes and brain atrophy associated with AD. We used MRI volume of four temporal lobe brain regions: entorhinal cortex, hippocampus, fusiform gyrus, and middle temporal gyrus. All values included in the dataset were in cubic millimeters. The range in the dataset for these four variables was determined, and then values were placed into five equal bins (0, 1, 2, 3, and 4) using the unstandardized scores via the interval binning function in SPSS (IBM Corporation, 2014), which creates equal-sized bins, controlling for baseline intracranial volume, gender, and age. These are the values that we parameterized to represent the MRI domain.

FDG-PET metabolic activity

The ADNI includes FDG-PET brain scans, which indicate regional cerebral metabolic activity using positron emission tomography of fluorodeoxyglucose, a glucose analog. We examined hypometabolism averaged across five key regions indicative of pathological change in AD: left angular gyrus, right angular gyrus, bilateral posterior cingular, left inferior temporal gyrus, and right inferior temporal gyrus. The range in the dataset for this variable was determined, and then values were placed into five equal bins (0, 1, 2, 3, and 4) using the unstandardized scores via the interval binning function in SPSS (IBM Corporation, 2014), which creates equal-sized bins. These are the values that we parameterized to represent the FDG-PET metabolic activity domain.

Cognitive performance

Neuropsychological measures of memory, language, visuospatial abilities, and executive function capture the breadth of cognitive decline that occurs in AD and are widely used in clinical research to assess cognitive dysfunction. We used measures that capture each of these cognitive domains. To assess memory ability, we used data from the Alzheimer's disease Assessment Scale – Cognition (ADAS-Cog; Mohs, Rosen, & Davis, 1983; Rosen, Mohs, & Davis, 1984) and Delayed Recall and Word Recognition subtests and Rey Auditory Verbal Learning Test (Rey, 1964), Immediate Recall, Delayed Recall, and Recognition. Each measure was placed into up to five equal bins; those scores were summed across tests. The range in the dataset for this variable was determined, and then values were placed into five equal bins (0, 1, 2, 3 and 4) using the unstandardized scores to represent our Memory domain via the interval binning function in SPSS (IBM Corporation, 2014), which creates equal-sized bins. To assess language ability, the same process was carried out for ADAS-Cog Naming (Rosen et al., 1984), Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1983), and Category Fluency-Animals (adapted from the CERAD Verbal Fluency test; Morris et al., 1989). To assess visuospatial ability, we used the same procedure for ADAS-Cog Constructional Praxis (Rosen et al., 1984) and Clock Drawing Test (Goodglass & Kaplan, 1983) Command and Copy. Finally, to assess executive function, we used ADAS-Cog Number Cancellation (Rosen et al., 1984) and Trail Making Test A and B (Reitan, 1958; Reitan & Wolfson, 1985). These are the values that we parameterized to represent the cognitive performance assessment. Additional measures used to verify the latent continuum and characterize the sample were the Mini Mental State Examination (Folstein, Folstein, & McHugh, 1975) and the CDR-SOB (O'Bryant et al., 2008).

Adaptive function

To measure adaptive function, we used the Functional Activities Questionnaire (FAQ; Pfeffer et al., 1982) which measures an individual's ability to perform instrumental activities of daily living (IADLs) such as writing checks, paying bills, working on a hobby, turning off the stove after use, etc. The range in the dataset was determined, and then values were placed into five equal bins (0, 1, 2, 3 and 4) using the unstandardized scores via the interval binning function in SPSS (IBM Corporation, 2014), which creates equal-sized bins. These are the values that we parameterized to represent the adaptive function assessment.

Though this binning approach may be useful for empirically derived theoretical models, we are not recommending it for other purposes.

Analyses

For our analyses, we defined the latent continuum in terms of four distinct sets of cognitive markers that span four key

dementia domains (memory, language, visuospatial ability, and executive function). Then, using IRT software (IRT-LR-DIF; Thissen, 2001), which is available online <http://www.unc.edu/~dthissen/dl.html>, we estimated the item parameters for each of those domains or “items”. We then used those items as “anchors” to define our latent continuum and determined the extent to which the following markers indicate that latent continuum: Four cognitive markers—Memory, Language, Visuospatial Ability, and Executive Function; Four items for MRI – Entorhinal Cortex, Hippocampus, Fusiform Gyrus, Middle Temporal Gyrus; One item for FDG-PET; and One item for Functional (FAQ score). The non-cognitive indicators were linked to the latent continuum via IRT procedures (Kolen & Brennan, 2004). IRT assumes unidimensionality of the data that define the core latent dimension. While different domains of cognitive functioning exist, because of the eventual global deterioration of cognition in AD it is possible that they statistically emerge as a single factor, which would allow for the unidimensional criteria for IRT to be met. As such, we conducted exploratory and confirmatory factor analyses to test the unidimensionality of the data that defined the latent continuum. Specifically, we tested a four-variable model, in which each of the ordinal cognitive indicator (i.e., Memory, Language, Visuospatial Ability, and Executive Function) served as the variables.

To test if these four variables covaried sufficiently together to meet the assumption of unidimensionality, we first used an exploratory factor analysis in SPSS to determine if the ratio of first to second eigenvalue was greater than the 3:1 ratio suggested by Embretson and Reise (2000). Results from the exploratory factor analysis indicated the data were indeed unidimensional enough for IRT analyses; the first eigenvalue was 2.25 and the second was 0.72, with a ratio of 3.13. Next we conducted a confirmatory factor analysis to test whether the ordinal data were sufficiently unidimensional for IRT analyses. Hu and Bentler (1999) have shown that hypothesized structural models provide a relatively good fit to the observed data when the Tucker-Lewis Index (TLI; Tucker & Lewis, 1973) and comparative fit index (CFI; Bentler, 1990) values are close to 0.95 and the root mean squared error of approximation (RMSEA; Steiger & Lind, 1980) is less than 0.06. Using these recommended cutoffs, we concluded that the data (CFI = 1.00, TLI = 0.99, RMSEA = 0.04) were excellent fits to the specified unidimensional models. A non-significant chi-square test can be used to provide further support for this determination. The Chi-Square test 4.94 (2), $p > .05$ was indeed non-significant, confirming that the data were sufficiently unidimensional for IRT analyses. Taken together, these analyses suggested that the data were robustly unidimensional and suitable for the main analysis.

Next, we conducted IRT analyses in IRT-LR-DIF (Thissen, Steinberg, & Gerrard, 1986) using Samejima's (1969) graded response model to determine as anchor items, the item parameters for the four variables (Memory,

Language, Visuospatial Ability, Executive Function). Then we used the same program to determine how well the following variables indicated that latent continuum, and established item parameters for them: MRI hippocampus, MRI entorhinal cortex, MRI fusiform gyrus, MRI middle temporal gyrus, FDG-PET, and FAQ.

In other words, once we defined the latent continuum, the non-cognitive indicators were linked to the continuum using the IRT-LR-DIF software program (Thissen, 2001). This software application can be used to establish parameters for “anchor” items that define the latent continuum of interest and then test the extent to which individual “candidate” items index that latent continuum by setting the scale for the parameters via the anchor items. Using this program, all parameters for all items can be established, but any IRT capable program can be used for the same purpose. Because the same set of items is used to define the latent continuum, each item has parameters that are directly comparable and quantify the extent to which the items “index” rather than “define” the latent continuum. Figure 1 displays the model used for establishing the parameters.

Results

First, we derived information curves, which indicate the degree of measurement precision, for each of the four assessments: cognitive performance, MRI volume, FDG-PET metabolic activity, and adaptive function (see Figure 2). In IRT, information is a form of reliability that quantifies the extent to which an indicator or a set of indicators can distinguish among individuals across different levels of the latent construct being measured (AD severity). Higher levels of information indicate that an indicator or a set of indicators distinguish relatively well among individuals of at different standings along the latent construct. The units of measurement in Figure 2 is standard deviations of the latent construct at hand. So 0.0 would represent the average among of cognitive dysfunction across these four cognitive indicators. With that explanation as a backdrop, results from Figure 3 indicated that each assessment modality provided varying degrees of information across the continuum of AD-related cognitive dysfunction. Cognitive measures contributed information across the spectrum of

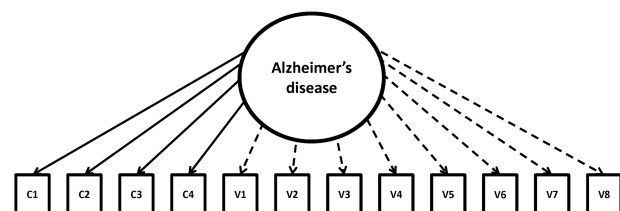


Figure 1. Model of Alzheimer's disease represented across cognitive and non-cognitive variables. Note. c1–c4 represent cognitive variables. v1–v8 represent non-cognitive variables. Solid lines indicate anchor items. Dashed lines indicate secondary items.

AD severity, increasing in the relatively mild to moderate stages of the disease. Our MRI volume and FDG-PET indicators provided information at relatively lower levels, but contributed information across the entire spectrum of AD-related cognitive dysfunction. FAQ provided information in the relatively moderate range of AD-related cognitive dysfunction.

To confirm that adding cross-domain indicators (MRI volume, FDG-PET metabolic activity, and adaptive function) provides significant information across the continuum of AD-related cognitive dysfunction, we graphed two information curves. The top information function represents the total information yielded when indicators across all four assessment modalities are used to measure AD-related cognitive dysfunction. The bottom information function represents information yielded from cognitive measures alone. Figure 3 illustrates the degree to which indicators of MRI, FDG-PET, and FAQ add measurement precision above and beyond that provided by cognitive measures. Inspection of the curves showed that adding indicators from these three

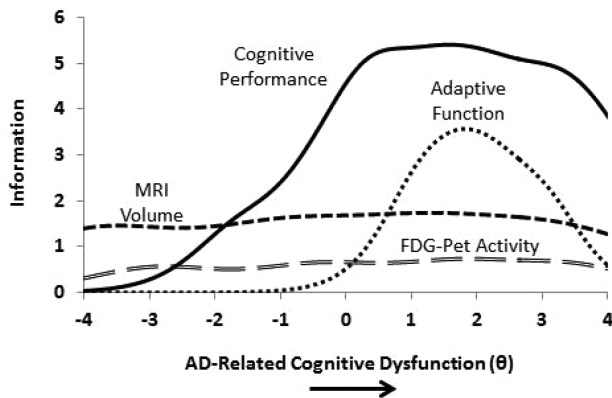


Figure 2. Measurement model of dynamic markers of the Alzheimer's pathological cascade. FDG-PET, fluorodeoxyglucose positron emission tomography; MRI, magnetic resonance imaging; X-axis is the latent variable theta. Reprinted from Balsis, Choudhury, Geraci, Bengel, & Patrick, 2018.

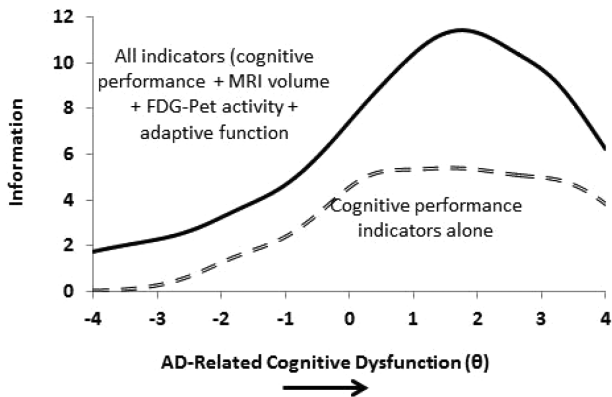


Figure 3. MRI, FDG-PET, and adaptive function indicators contribute measurement precision above and beyond that provided by cognitive performance indicators alone. FDG-PET, fluorodeoxyglucose positron emission tomography; MRI, magnetic resonance imaging; X-axis is the latent variable theta.

domains increased information across the entire spectrum of the latent construct beyond that provided simply by the cognitive performance domain.

Figure 4 illustrates precisely where dysfunction becomes evident in measures from these different domains, and reveals a cascade of decline that informs theoretical models of the disease process. Specifically, the figure shows that the probability of reduced brain volume is higher at relatively mild levels of disease severity, with diminished brain activation indexing similar levels of disease severity. Cognitive performance and adaptive function impairments then become apparent at higher levels of the disease. Of note, decrements in brain volume reveal the initial signs of AD pathology well before (i.e., two standard deviations below) the point at which sophisticated cognitive tests index disease-related dysfunction; the brain pathology precedes the apparent clinical manifestations of the disease.

To validate our latent continuum used in all analyses above, we used Multilog (Thissen, 1991) to estimate scores for individuals along the latent continuum. These scores are known as *maximum a posterioris* scores (MAPs), which are known along the theta continuum (θ ; X-axis of Figures 2–4) based on the pattern of dysfunction for each participant. Table 1 shows that θ scores help to validate the latent continuum as they differ significantly across diagnostic group in the expected way, with cognitively normal individuals scoring lower than those with MCI, and those with MCI scoring lower than those with AD. Finally, Table 1 shows that individuals in these groups have the expected differences in ADAS-Cog, CDR-SOB, and MMSE scores; All of these measures share significant correlations with θ , the latent core continuum of cognitive dysfunction ($r = .74$ ADAS-Cog; $r = .69$ with CDR-SOB; $r = -.64$ with MMSE); all p values $< .001$). Of particular note, there was a significant progression of θ scores across diagnostic category (Normal, MCI, AD), $F(1,055) = 444.64, p < .001$. Cognitively normal individuals had scores of $M = -.58$ ($SD = .56$) and those with MCI had θ scores of $M = .03$ ($SD = .66$), where the cognitive curve begins to indicate

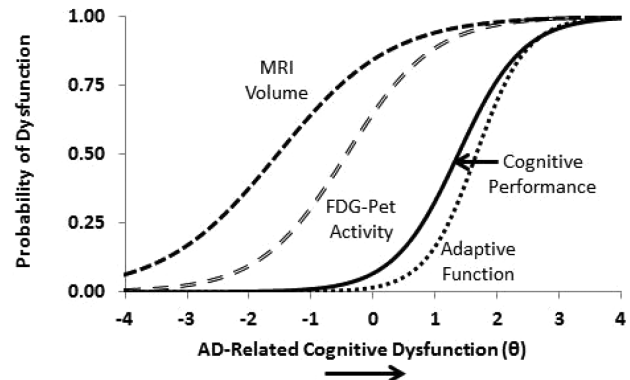


Figure 4. Statistical model of dynamic markers of the Alzheimer's pathological cascade. FDG-PET, fluorodeoxyglucose positron emission tomography; MRI, magnetic resonance imaging; X-axis is the latent variable theta.

Table 1. Alzheimer's Disease-Related Cognitive Dysfunction Across Dementia Categories

	Normal ($n = 341$) M (SD)	MCI ($n = 560$) M (SD)	AD ($n = 155$) M (SD)	F	
ADAS-Cog	5.91 (2.99)	9.58 (4.33)	19.34 (6.37)	513.77	$p < .001$
CDR-SOB	0.04 (0.14)	1.48 (0.86)	4.47 (1.69)	1,278.35	$p < .001$
MMSE	28.96 (1.24)	27.84 (1.75)	23.37 (2.05)	629.14	$p < .001$
Theta (θ)	-0.58 (0.56)	0.03 (0.66)	1.19 (.56)	444.64	$p < .001$

Note. AD = Alzheimer's disease; ADAS-Cog = Alzheimer's Disease Assessment Scale; CDR-SOB = Clinical Dementia Rating Scale – Sum of the Boxes; MMSE = Mini Mental State Examination; MCI = Mild Cognitive Impairment; $N = 1,056$.

the continuum of AD severity. Individuals with CDR-SOB scores greater than 0.5 have scores of $M = 1.15$ ($SD = .56$).

To test whether the findings would replicate, we divided the sample in half (for each half, $n = 528$) and then conducted IRT-based likelihood-ratio differential item functioning (DIF) testing (Thissen, Steinberg, & Gerrard, 1986). This type of DIF testing can be used to determine whether the parameters that define the curves in our models replicate across subsamples. DIF testing of this kind involves statistically comparing IRT models with G^2 difference tests. For each item, a model with item parameters constrained equal across the two subsamples is compared with a model that permits item parameters to vary between the two subsamples. For one item at a time, the graded response model was tested. If the constraints significantly decrease model fit, there is evidence of omnibus DIF (DIF with respect to a , b , or both) for that item. We applied a Bonferroni correction across all 12 items to correct for possible false positives ($p = .05/12 = .0041$). The analyses were conducted as described, treating the same cognitive items as “anchors” to define the latent continuum. Neither the anchor items nor the candidate items showed statistically significant DIF, indicating that the parameters in our models replicate. Results from this split-half analysis suggest that the findings are robust in this sample

Discussion

It has been suggested that MRI volume, FDG-PET metabolic activity, cognitive performance, and adaptive function are differentially affected by the progression of AD. The current study had two objectives. The first objective was to establish a measurement model across these assessment modalities and quantify the degree to which indicators from each provide AD-related measurement information. The second objective was to use this measurement model to analyze where these marker domains become dysfunctional across the spectrum of disease severity, from pre-clinical AD to full dementia. Figures 2 and 3 reveal that AD-related cognitive dysfunction can be indexed across the aforementioned domains, while Figure 4 indicates where along the latent continuum dysfunction begins to occur across domains.

Figure 4 illustrates a cascade of decline that informs theoretical models of the disease process, showing that MRI

volume decreases in the very mild spectrum of severity, followed by FDG-PET metabolic activity. Decrements in these neurophysiological biomarkers are followed by decrements in cognitive performance and finally adaptive function. In other words, the decrements in brain volume reveal the initial signs of AD pathology two standard deviations before standard cognitive tests capture the dysfunction; the brain pathology precedes the apparent clinical manifestations of the disease.

Our measurement model was dependent on the ability of the ADNI neuropsychological battery to provide information about the constructs along the latent continuum. Cognitive tests that can detect subtle changes in the very early stages of the disease continuum would result in increased information yielded by the cognitive indicators. In Figures 2 and 3, this would result in an upward shift of the cognitive performance curve at lower levels of disease severity. Because Figure 4 is a model that is linked to our understanding of these phenomena inasmuch as our current measures inform our understanding of these phenomena, changes in Figures 2 and 3 could result in a possible leftward shift of the cognitive performance curve in Figure 4. In other words, the current statistical model and the placement of the curves in that model are dependent on the ability of the tests to capture these phenomena across the spectrum of the disease. In this way, these findings can guide us in terms of measurement development. As we develop more sensitive neuropsychological tests of cognition, our understanding of the AD cascade, as modeled in Figure 4, might change; these findings may guide investigators who might wish to develop cognitive measures that detect the disease in its relatively mild form.

Limitations and Future Directions

The current study has limitations. As with all research, these findings require replication. Future work should examine whether these results generalize to different samples, particularly samples that differ demographically. Despite the large size, the sample was overwhelmingly white with only a few participants who identified themselves as black or African American ($n = 52$) and very few other minorities. It is also worth noting that despite the comprehensive approach to examining multiple AD markers, we did not examine all existing markers of AD progression. The

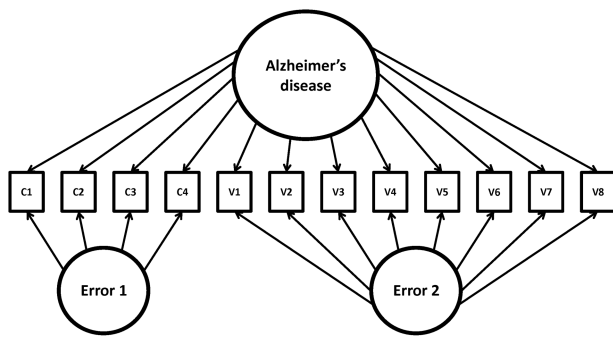


Figure 5. SEM model of Alzheimer's disease represented across cognitive and non-cognitive variables. Note. c1–c4 represent cognitive variables. v1–v8 represent non-cognitive variables. Solid lines indicate anchor items. Dashed lines indicate secondary items.

current study examined a subset of MRI volume measurements and FDG-PET regional metabolic activity readings, selected cognitive performance assessments, and one adaptive function test. Although there are other markers one could include, the MRI and FDG-PET data in this study accounted for considerable variance within both domains and are of clinical relevance. Future studies could also use a similar statistical approach to model other AD markers and how they relate to one another as severity increases, for a more comprehensive overview of the disease. For now, our findings provide a statistical model of the progression of AD showing that biomarkers indicate the disease in its mildest form followed by the more apparent cognitive and functional dysfunction.

One might want to explore the extent to which parameters from the current model (Figure 1) dovetail with parameter estimates for the main dimension of a bifactor model (Figure 5). To explore this, we randomly selected half of the sample, calculated parameters estimates via the main model (Figure 1) according to the procedures outlined in the method section, and compared those parameters to the parameters that define the main dimension in this secondary model (Figure 5). FlexMIRT (Cai, 2013) was used to calculate this second set of parameters. The parameters of the main dimension for Figure 5 correlated significantly with the parameters of those derived from Figure 1, $r = 0.95$, $p < .05$, suggesting that the rank order of the parameters are nearly redundant across the two approaches. Furthermore, we tested the fit of the model; allowing the errors to correlate produced adequate fit according to Hu and Bentler (1999) standards, CFI = 0.98, TLI = 0.97, RMSEA = 0.07. Taken together with the replication, all results point to the same main findings.

Conclusions

Theoretical models have been proposed to describe the effect of AD progression on key markers of AD severity, such as MRI volume, FDG-PET metabolic activity, cognitive performance and functional impairment (Jack et al., 2010, 2013). The current study provides statistical confirmation of this general cascade of Alzheimer's pathology and shows how changes within assessment modalities are associated

with the spectrum of cognitive dysfunction in AD within a single analysis of one sample. In addition to providing a statistical framework to support the existing theoretical models, the current statistical approach more generally provides a framework to develop a finer-grained understanding of the progression of AD. Further, this statistical framework provides key insight about of how these key assessments may change in relation to one another as AD progresses.

These findings carry with them certain implications for the clinical diagnosis of AD as well. Attempting to describe AD using incremental, semi-discrete categories may prove sufficient when examining the disease at a macro level, but each AD patient is unique. Our statistical model illustrates a more granular continuum that could be used to provide more precise insight into a patient's relative level of AD severity, given their unique constellation of disease markers. Furthermore, our measurement model illustrates that we may gain understanding of the disease by considering indicators of the disease across assessment markers domains in a single model of disease progression. This finding supports the clinical utility of including a variety of assessment tools in a comprehensive AD evaluation.

Funding

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Acknowledgements

Data used in preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report.

The first author is responsible for data analyses. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Conflict of Interest

The authors report no conflicts of interest.

References

- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A.,...Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7, 270–279.
- Balsis, S., Choudhury, T. K., Geraci, L., Bengel, J. F., & Patrick, C. J. (2018). Alzheimer's disease assessment: a review and illustrations focusing on item response theory techniques. *Assessment*, 25, 360–373.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246.
- Cai, L. (2013). *FlexMIRT® version 2: flexible multilevel multidimensional item analysis and test scoring [Computer software]*. Chapel Hill, NC: Vector Psychometric Group.
- Carter, S. F., Caine, D., Burns, A., Herholz, K., & Lambon Ralph, M. A. (2012). Staging of the cognitive decline in Alzheimer's disease: insights from a detailed neuropsychological investigation of mild cognitive impairment and mild Alzheimer's disease. *International Journal of Geriatric Psychiatry*, 27, 423–432. doi:10.1002/gps.2738
- Chen, K., Langbaum, J. B., Fleisher, A. S., Ayuthanont, N., Reschke, C., Lee, W., Liu, X.,...Reiman, E. M. (2010). Twelve-month metabolic declines in probable Alzheimer's disease and amnesic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: findings from the Alzheimer's Disease Neuroimaging Initiative. *Neuroimage*, 51, 654–64.
- Desikan, R. S., Sabuncu, M. R., Schmansky, N. J., Reuter, M., Cabral, H. J., Hess, C. P.,...Dale, A. M.; Alzheimer's Disease Neuroimaging Initiative. (2010). Selective disruption of the cerebral neocortex in Alzheimer's disease. *PLoS One*, 5, e12853. doi:10.1371/journal.pone.0012853
- Devanand, D.P., Bansal, R., Liu, J., Hao, X., Pradhaban, G., & Peterson, B. S. (2012). MRI hippocampal and entorhinal cortex mapping in predicting conversion to Alzheimer's disease. *Neuroimage*, 60, 1622–1629. doi:10.1016/j.neuroimage.2012.01.075
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.
- Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders* (2nd ed.). Philadelphia, PA: Lea & Febiger.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1–55.
- IBM Corporation. (2014). *SPSS Statistics Version 23.0 (for Windows) [computer program]*. Armonk, NY: IBM Corporation.
- Jack, C. R. Jr, Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W.,...Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, 9, 119–128. doi:10.1016/S1474-4422(09)70299-6
- Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C.,...Trojanowski, J. Q. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology*, 12, 207–16.
- Johnson, K. A., Fox, N. C., Sperling, R. A., & Klunk, W. E. (2012). Brain imaging in Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, 2: a006213.
- Johnson, E. C., & Meade, A. W. (2007). *The role of referent indicators in tests of measurement invariance*. Paper presented at the 22nd Annual Meeting of the Society for Industrial and Organizational Psychology, New York.
- Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston naming test: Assessment of aphasia and related disorders*. Philadelphia, PA: Lea & Febiger.
- Kolen, M.J. & Brennan, R.L. (2004). *Test equating, scaling, and linking*. New York, NY: Springer.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34, 263–269.
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H.,...Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3), 263–269. doi:10.1016/j.jalz.2011.03.005
- Mohs, R. C., Rosen, W. G., & Davis, K. L. (1983). The Alzheimer's disease assessment scale: an instrument for assessing treatment efficacy. *Psychopharmacology Bulletin*, 19, 448–450.
- Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G.,...Clark, C. (1989). The consortium to establish a registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, 39, 1159–1163.
- Mosconi, L., Tsui, W. H., Herholz, K., Pupi, A., Drzezga, A., Lucignani, G.,...de Leon, M. J. (2008). Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. *Journal of Nuclear Medicine*, 49, 390–398. doi:10.2967/jnumed.107.045385.
- O'Bryant, S. E., Waring, S. C., Cullum, C. M., Hall, J., Lacritz, L., Massman, P. J.,...Doody, R. (2008). Staging dementia using clinical dementia rating scale sum of boxes scores: a Texas Alzheimer's research consortium study. *Archives of Neurology*, 65, 1091–1095. doi:10.1001/archneur.65.8.1091
- Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V.,...Winblad, B. (2001). Current concepts in mild cognitive impairment. *Archives of Neurology*, 58, 1985–1992.
- Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H. Jr, Chance, J. M., & Filos, S. (1982). Measurement of functional activities in older adults in the community. *Journal of Gerontology*, 37, 323–329.

- Reisberg, B., Finkel, S., Overall, J., Schmidt-Gollas, N., Kanowski, S., Lehfeld, H.,...Erzigkeit, H. (2001). The Alzheimer's disease activities of daily living international scale (ADL-IS). *International Psychogeriatrics*, *13*(2), 163–81.
- Reitan, R. M. (1958). Validity of the Trail Making test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*, 271–276.
- Reitan, R. & Wolfson, D. (1985). *The halstead-reitan neuropsychological test battery*. Tucson, AZ: Neuropsychology Press.
- Rey, A. (1964). *L'examen clinique en psychologie [The clinical psychological examination]*. Paris, France: Presses Universitaires de France.
- Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *The American Journal of Psychiatry*, *141*, 1356–1364. doi:10.1176/ajp.141.11.1356
- Samejima, F. (1969). *Estimation of latent ability using a Response Pattern of Graded Scores (Psychometric Monograph No. 17)*. Richmond, VA: Psychometric Society.
- Scahill, R. I., Schott, J. M., Stevens, J. M., Rossor, M. N., & Fox, N. C. (2002). Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI. *Proceedings of the National Academy of Sciences*, *99*(7), 4703–7.
- Schuff, N., Woerner, N., Boreta, L., Kornfield, T., Trojanowski, J. Q.,...Weiner, M. W.; Alzheimer's Disease Neuroimaging Initiative. (2009). MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. *Brain*, *132*, 1067–1077. doi:10.1093/brain/awp007
- Sperling, R. A., & Johnson, K. A. (2012). Dementia: new criteria but no new treatments. *The Lancet Neurology*, *11*, 4–5. doi:10.1016/S1474-4422(11)70272-1
- Steiger, J.H., & Lind, J.C. (1980). *Statistically based tests for the number of common factors*. Paper presented at: *Annual Meeting of the Psychometric Society*; Iowa City, IA.
- Thissen, D. (1991). *MULTILOG user's guide: Multiple, categorical item analysis and test scoring using item response theory*. Skokie, IL: Scientific Software International.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10.
- Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group-mean differences: the concept of item bias. *Psychological Bulletin*, *99*, 118–128.
- Vemuri, P., Wiste, H. J., Weigand, S. D. Jr; Alzheimer's Disease Neuroimaging Initiative. (2009). MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. *Neurology*, *73*, 294–301. doi:10.1212/WNL.0b013e3181af79fb
- Wechsler, D. (1987). *Wechsler Memory Scale – Revised*. San Antonio, TX: The Psychological Corporation.
- Whitwell, J. L., Przybelski, S. A., Weigand, S. D., Knopman, D. S., Boeve, B. F., Petersen, RC,...Jack, C. R. (2007). 3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease. *Brain*, *130*, 1777–1786. doi:10.1093/brain/awm112
- Wilson, R. S., Segawa, E., Boyle, P. A., Anagnos, S. E., Hize, L. P., & Bennett, D. A. (2012). The natural history of cognitive decline in Alzheimer's disease. *Psychology and Aging*, *27*, 1008–1017. doi:10.1037/a0029857