UCSF

Title

Data-driven regions of interest for longitudinal change in frontotemporal lobar degeneration

Permalink

https://escholarship.org/uc/item/4g86v189

Authors

Pankov, Aleksandr
Binney, Richard J
Staffaroni, Adam M
et al.

Publication Date

DOI

Peer reviewed

CrossMark

Data-driven regions of interest for longitudinal change in frontotemporal lobar degeneration

Aleksandr Pankov BS[a,d,*], Richard J. Binney PhD[b], Adam M. Staffaroni MS[b], John Kornak PhD[a], Suneth Attygalle MS[b], Norbert Schuff PhD[c], Michael W. Weiner MD[c], Joel H. Kramer PsyD[b], Bradford C. Dickerson MD[e], Bruce L. Miller MD[b], Howard J. Rosen MD[b,*]

[a] Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA, USA
[b] Department of Neurology, Memory and Aging Center, University of California, San Francisco, San Francisco, CA, USA
[c] Department of Radiology, University of California, San Francisco, San Francisco, CA, USA
[d] Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA, USA
[e] Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

ARTICLE INFO

ABSTRACT

Current research is investigating the potential utility of longitudinal measurement of brain structure as a marker of drug effect in clinical trials for neurodegenerative disease. Recent studies in Alzheimer's disease (AD) have shown that measurement of change in empirically derived regions of interest (ROIs) allows more reliable measurement of change over time compared with regions chosen a-priori based on known effects of AD on brain anatomy. Frontotemporal lobar degeneration (FTLD) is a devastating neurodegenerative disorder for which there are no approved treatments. The goal of this study was to identify an empirical ROI that maximizes the effect size for the annual rate of brain atrophy in FTLD compared with healthy age matched controls, and to estimate the effect size and associated power estimates for a theoretical study that would use change within this ROI as an outcome measure. Eighty six patients with FTLD were studied, including 43 who were imaged twice at 1.5 T and 43 at 3 T, along with 105 controls (37 imaged at 1.5 T and 67 at 3 T). Empirically-derived maps of change were generated separately for each field strength and included the bilateral insula, dorsolateral, medial and orbital frontal, basal ganglia and lateral and inferior temporal regions. The extent of regions included in the 3 T map was larger than that in the 1.5 T map. At both field strengths, the effect sizes for imaging were larger than for any clinical measures. At 3 T, the effect size for longitudinal change measured within the empirically derived ROI was larger than the effect sizes derived from frontal lobe, temporal lobe or whole brain ROIs. The effect size derived from the data-driven 1.5 T map was smaller than at 3 T, and was not larger than the effect size derived from a-priori ROIs. It was estimated that measurement of longitudinal change using 1.5 T MR systems requires approximately a 3-fold increase in sample size to obtain effect sizes equivalent to those seen at 3 T. While the results should be confirmed in additional datasets, these results indicate that empirically derived ROIs can reduce the number of subjects needed for a longitudinal study of drug effects in FTLD compared with a-priori ROIs. Field strength may have a significant impact on the utility of imaging for measuring longitudinal change.

1. Introduction

FTLD is a neurodegenerative disorder that typically presents as one of three clinical variants: 1) the behavioral variant of frontotemporal dementia (bvFTD), characterized by progressive impairment in socioemotional function, 2) the semantic variant of primary progressive aphasia (svPPA), characterized by progressive loss of knowledge about words and objects, and 3) the nonfluent variant of PPA (nfvPPA), characterized by progressive impairment of articulation and speech (Rascovsky et al., 2011). FTLD is at least as common as Alzheimer's disease (AD) in people under the age of 65 (Borroni et al., 2009; Knopman et al., 2004; Papageorgiou et al., 2009; Ratnavalli et al., 2002; Rosso et al., 2003). It has a profound effect on the lives of patients and their families; one that can be considered more detrimental than the effects of more typical degenerative disease such as AD because it is associated with an earlier age of onset (Papageorgiou et al., 2009) and more rapid rate of decline (Roberson et al., 2005). As there are no approved treatments for FTLD,

* Corresponding author at: University of California, San Francisco, Memory and Aging Center, Suite 190, Box 1207, 675 Nelson Rising Lane, San Francisco, CA 94158, USA. Tel.: +1 415 476 5567.

these features reinforce the importance of developing therapies for this devastating disorder.

Brain imaging is a powerful tool in neurodegenerative disease. MRI and PET, the most commonly used tools, can be used to support diagnosis, and measures derived from brain images correlate with the type and severity of symptoms in each patient (Tartaglia et al., 2011). These observations have led to studies examining the utility of brain imaging as a tool for clinical trials, which have demonstrated that longitudinal imaging can be used to track change in neurodegenerative disorders more reliably than clinical measures such as cognitive testing (Weiner et al., 2013).

One limitation of brain imaging is that each image produces hundreds or thousands of data-points corresponding to spatial locations in the brain, posing a significant hurdle for defining imaging-based biomarkers (Friston et al., 1996). One of the most common approaches to reduce the large-scale data in imaging studies is to limit measures of change to aggregated estimates over regions-of-interest (ROIs), which tend to be chosen based on prior knowledge about the regions that are most severely affected in each disease. Thus, in AD, ROIs chosen often include the hippocampus, entorhinal cortex, and temporoparietal regions (Dickerson et al., 2011). In FTLD, the frontal and/or temporal lobes have been used (Gordon et al., 2010; Krueger et al., 2010). The regions most severely affected in each disease, however, tend to be those affected earliest (Jack et al., 1997; Seeley et al., 2008). When a disorder moves beyond the earliest stages, it is possible that regions affected early begin to slow their rate of change while other regions, previously only mildly affected, begin to accelerate their decline (Brambati et al., 2009; Rohrer et al., 2012; Schuff et al., 2012). Thus, ROIs chosen based on regions that are most strongly associated with the disease may not be ideally placed to include the regions changing most rapidly. In AD, it has recently been demonstrated using both MRI and PET that ROIs chosen based on empirical data about the regions of most robust change improve statistical power to detect change compared with ROIs chosen based on their general association with the disease (Chen et al., 2010; Hua et al., 2009).

The goal of this study was to empirically identify the most reliable regions of change in FTLD using longitudinal imaging data and to quantify the power estimates for longitudinal imaging studies that would use an ROI derived from these regions. We also compared these power estimates to those that would be calculated based on anatomically based ROIs. We included longitudinal image pairs from a patient cohort accrued between 1999 and 2013, including images obtained at both 1.5 Tesla (T) and 3 T (same field strength within patients). Our original intent was to assess reproducibility by generating the ROI in the 1.5 T dataset and testing the reproducibility of the estimate of change in the 3 T dataset. However, we found important differences in the effect size of change estimates across field strengths that could have implications for future studies. Thus, we have estimated data-driven ROIs for each field strength separately and included the analyses at both field strengths in order to report on those differences.

2. Methods

2.1. Subjects

Our overall objective was to include as many patients as possible. We chose to include patients with the behavioral variant of frontotemporal dementia (bvFTD, n = 37 [14 in 1.5 T and 23 in 3 T]) and the semantic variant of primary progressive aphasia (svPPA, n = 49 [29 in 1.5 T and 20 in 3 T]), along with healthy age-matched comparison subjects (n = 105 [38 in 1.5 T and 67 in 3 T]). These two variants of FTLD were chosen because they are often caused by similar neuropathological changes (Seelaar et al., 2011) and prior studies have shown significant overlap between these two disorders in both behavioral symptomatology and anatomical regions of damage (Liu et al., 2004; Rosen et al., 2002). Because of this overlap, we reasoned that

there are likely to be shared regions of longitudinal change between these variants, justifying our combining them into a single group. Recent clinical trials have also combined the two groups (Boxer et al., 2013). Patients were recruited between 1998 and 2013 through ongoing studies (AG019724, AG032306, AG023501) at the University of California San Francisco (UCSF) Memory and Aging Center (MAC). Diagnosis for these studies is based on a multidisciplinary evaluation incorporating neurological, neuropsychological, and nursing assessment (Rosen et al., 2002). Structural neuroimaging is not used to make the syndromic diagnosis, but only to exclude other causes of brain damage, such as strokes or tumors.

Two groups of cognitive healthy comparison (HC) subjects were used for analysis. One group of 105 HCs was included who had longitudinal imaging at 1.5 T (n = 38) and 3 T (n = 67), and these controls were used to create maps of disease specific change in FTLD (see below). A second group of 272 HCs was used to generate z-scores in order to create composite metrics of change in cognitive scores over time (described below). All HC data were obtained from a cohort of subjects recruited at the MAC via advertisements and community events. Controls undergo the same evaluation as patients and are required to have no clinically significant cognitive or behavioral complaints, performance within one standard deviation of normal on all cognitive tasks, and to bring a knowledgeable informant to verify the absence of clinically significant cognitive or behavioral problems. Controls were excluded if they had a history of significant mood disorders, clinically-significant alcohol or drug use, significant vascular disease, visual problems that would impair test performance, other neurological conditions, and self-reported deficits in cognition.

To be included in the present study, a subject was required to have had two T1-weighted MRI scans acquired with the same scanner and pulse sequence. We aimed for an inter-scan interval of 1 year. However, in the interest of sample size, we included 2 subjects that had an interval greater than 2 years (one at 2.12 years and one at 2.27 years) as they otherwise met criteria. All estimated change maps were divided by the inter-scan interval in years (see below) to become maps of estimated annual rate of relative volume change, so that differences due to inter-scan interval effects would be somewhat mitigated. Images were inspected for quality, including ensuring whole brain coverage and looking for excessive motion artifact. In the years prior to 2008, research subjects underwent MRI at 1.5 T. Subsequently, MRI was acquired at 3 T. Subjects were divided into two sets based on magnet strength. No patient was included in both the 1.5 T and 3 T sets. The majority of the patients had assessment of CNS amyloid burden, usually with PET amyloid imaging using Pittsburgh B compound or, less commonly AV45, but sometimes using CSF. Also, some patients have gone on to autopsy. Patients were excluded if they had evidence of amyloid at a level consistent with AD based on any of these assessments. All subjects provided informed consent, and the clinical and imaging protocols were approved by the UCSF Committee on Human Research.

2.2. Clinical assessment

Patients received an extensive battery of cognitive and behavioral assessments. In order to compare the utility of MRI-based measures with the value of clinical variables for tracking disease progression, a subset of these measures was used. Prior studies have indicated that quantification of daily function and cognitive composite scores is among the most useful clinical variables for this purpose (Knopman et al., 2008). Functional state was quantified using the Clinical Dementia Rating (CDR) (Morris, 1997), which was used here to generate a continuous variable based on the sum of the individual ratings for functional domains, typically referred to as the sum-of-the-boxes (CDR-SB). General intellectual functioning was quantified using the Mini-mental state examination (MMSE; Folstein et al.,1975) The cognitive tasks

assessed a variety of abilities typically affected in bvFTD and svPPA, which can be divided into four broad domains:

1 Episodic memory: California Verbal Learning Test, Second Edition, Short Form (CVLT; D.C. Delis et al., 2000). The total immediate recall over four learning trials, delayed memory, and discriminability scores were used for analysis. Discriminability summarizes hits and false-positives on recognition performance. Modified Rey–Osterrieth Figure memory task (Benson figure; Kramer et al., 2003).
2 Language/semantic knowledge: Boston Naming Test (BNT; Kaplan et al., 1983), Semantic fluency (D. Delis et al., 2001), Peabody Picture Vocabulary Test — Revised (PPVT-R; Dunn & Dunn, 1997), Pyramids and Palm Trees (PPT; Howard & Patterson, 1992). WAIS —III, Information subtest (Wechsler, 1997).
3 Visuospatial: Modified Rey–Osterrieth Figure Copy (Benson Figure) (Kramer et al., 2003), Number Location from the Visual Object and Space Perception (VOSP) battery (Warrington & James, 1991), Beery-Buktenica Developmental Test of Visual–Motor Integration (Beery, 1997) — participants copy a series of basic geometric figures of increasing complexity. Scoring is based on objective criteria. WAIS-III Block Design (WAIS-BlockD) (Wechsler, 1997).
4 Executive functions: Modified trailmaking set-shifting task (Trails; Kramer et al., 2003). A logarithmic transformation was performed to normalize the data. Design Fluency (Filled Dots Condition) from the Delis–Kaplan Executive Function System (D. Delis et al., 2001), Lexical Fluency (Birn et al., 2010), Stroop color–word interference (Stroop, 1935), and a backward digit span task (Digits BW).

2.3. Image acquisition

1.5 T MRI was acquired on a Siemens Magnetom VISION system (Siemens, Iselin, NJ) equipped with a standard quadrature head coil. A volumetric magnetization prepared rapid gradient-echo (MPRAGE) sequence was used to acquire T1-weighted images of the entire brain (coronal slice orientation; slice thickness = 1.5 mm; in-plane resolution = 1.0×1.0 mm; matrix = 256×256; time to repetition (TR) = 10 ms; echo time (TE) = 4 ms; inversion time (TI) = 300 ms; flip angle = 15°). MRI was acquired on a 3.0 T Siemens Tim Trio system (Siemens, Iselin, NJ) equipped with a 12-channel receiver head coil. A volumetric MPRAGE sequence was used to acquire T1-weighted images of the entire brain (coronal slice orientation; slice thickness = 1.0 mm; in-plane resolution = 1.0×1.0 mm; matrix = 240×256; TR = 2300 ms; TE = 3 ms; TI = 900 ms; flip angle = 9°).

2.4. Creation of cognitive composites and effect sizes for clinical variables

In order to reduce the number of variables used to assess longitudinal change, a single composite measure was created for each of the four domains listed above based on mean z-scores compared with a normative sample of age-matched heal. In the first step, means and standard deviations were created for each cognitive measure, based on a group of 272 healthy control participants (mean age: 70.56 (7.0); mean education: 17.5 (2.0)) who had received these measures. This control group was separate from the control group used for the image analysis. Next, z-scores were created for svPPA participants at each time point by subtracting the HC mean from their performance and dividing by the standard deviation of HC participants. Composite scores were created based on the available data for each participant even if some data were missing. An episodic memory composite was created if a participant had at least 3 of 4 measures at a given time point. A semantic memory composite was created if participants had at least 2 of 5 measures at a given time point. A spatial composite was created if participants had at least 2 of 4 measures. Lastly, an executive composite was created for a participant if they had at least 3 of 5 measures at a given time point.

For each cognitive composite measure, we calculated the Cohen's d effect size measure to calculate the change between time points for

the FTLD and HC groups assuming unequal variances. The HC group was included in order to estimate the FTLD-specific change, which is analogous the approach we used to estimate rates of atrophy in FTLD (see below). The HC group used for the image analysis was used for this estimate of clinical change in HCs. We then used this effect size measure ($\alpha = 0.05$, $\beta = 0.8$) to estimate the necessary sample size to detect a 20% reduction in the change in clinical measures in the FTLD group.

2.5. Image processing

Longitudinal changes in regional brain volume were estimated using the Pairwise Longitudinal Registration Toolbox implemented in SPM12b (Ashburner & Ridgway, 2012), which addresses concerns regarding asymmetric bias in pair-wise longitudinal registration (Thomas, 2010; Yushkevich et al., 2010). The process begins with intrasubject registration using iterative and interleaved rigid-body alignment, diffeomorphic warping and correction for differential intensity inhomogeneity to generate a within-subject template representing an average of the subject's two scans with respect to position, shape and intensity non-uniformity. Two Jacobian determinant maps are then computed; one that encodes the relative difference in volume between the first scan and the average template, and another that describes the relative volume between the second scan and the template. These maps are generated in the same image space as the within-subject average. Computing the difference between these two Jacobian determinants provides a map of relative change in volume between scan one and scan two at each spatial location. These change maps were divided by the inter-scan interval (in units of years) to become maps of annual rate of relative volume change. Each subject's average image was bias corrected and the brain was partitioned into gray matter, white matter, and cerebrospinal fluid (CSF), using SPM12b's unified segmentation procedure. The contraction/expansion maps were then multiplied with the gray matter probabilistic tissue segmented maps, in within-subject average space, to restrict analyses to cortical and subcortical gray matter.

To allow statistical analysis across subjects, all image data were transformed to a standardized space. Mappings from the gray matter and white matter segments of the within-subject averages (all patients and control subjects) to an iteratively evolving study-specific population mean of these tissues were estimated using the DARTEL (diffeomorphic anatomical registration through an exponentiated lie algebra) toolbox (Ashburner, 2007). DARTEL minimizes the geodesic distance from each patient to the population mean. Thus, between-population asymmetries in registration, which could also lead to erroneous population effects, were addressed. An affine mapping between the population mean and MNI space (defined by SPM12b's Prior Tissue Probability Map) was also estimated and combined with each subject-to-population mean mapping for warping average images and volume expansion/contraction

Table 1
Basic demographics in subject groups.

		N	Sex (M/F)	Mean age at time 1 (S.D.)	Mean interscan interval in years (S.D.)
1.5 T	All subjects	81	42/39	64.7 (7.21)	1.08 (0.25)
	Healthy	38	15/23	66.42 (7.89)	1.06 (0.17)
	All FTLD	43	27/16	63.19 (6.28)	1.11 (0.28)
	bvFTD	14	10/4	61.36 (7.03)	1.14 (0.37)
	svPPA	29	17/12	64.07 (5.7)	1.14 (0.37)
3 T	All subjects	110	58/52	66.5 (7.78)	1.03 (0.24)
	Healthy	67	33/34	69.28 (6.53)	1.05 (0.21)
	All FTLD	43	25/18	62.16 (7.65)	0.99 (0.28)
	bvFTD	23	13/10	60.65 (7.8)	1.00 (0.29)
	svPPA	20	12/8	63.9 (7.26)	0.99 (0.27)

rate maps to MNI space. The rate change maps were then warped to population-in-MNI space using the abovementioned mapping composition, and resampled to 1.5 mm³ without 'volume-preserving' modulation. No spatial smoothing was applied. Subsequent analysis was done using only the gray-matter maps of each patient.

2.6. Generation and evaluation of a data-driven ROI

Our data-driven ROI generation procedure follows (in spirit) from the approaches of Chen et al. and Hua et al. (Chen et al., 2010; Hua et al., 2009), adopting some minor modifications, to identify a single

Table 2
Baseline and 1-year clinical data in svPPA and bvFTD groups at each field strength.

Measure[a]	svPPA				bvFTD			
	1.5 T		3 T		1.5 T		3 T	
	Baseline (n, S.D.)	Followup (n, S.D.)	Baseline (n, S.D.)	Followup (n, S.D.)	Baseline (n, S.D.)	Followup (n, S.D.)	Baseline (n, S.D.)	Followup (n, S.D.)
General functioning								
MMSE	22.42	20.16	25.06	23.11	22.57	21.08	24.52	20.22
	(24; 7.98)	(25; 7.88)	(16; 5.12)	(18; 5.99)	(14; 9.1)	(12; 8.71)	(21; 4.07)	(18; 7.54)
CDR-SB	3.73	5.13	3.68	4.53	6.5	10.41	7.22	9.5
	(24; 2.29)	(23; 2.90)	(19; 2.19)	(20; 2.53)	(14; 3.36)	(11; 4.18)	(23; 3.04)	(22; 3.15)
Memory scores								
CVLT-immediate recall (Max = 36)	16.0	13.17	18.87	16.2	20.5	20.2	22.0	18.62
	(18; 6.65)	(18; 6.91)	(15; 5.51)	(15; 5.38)	(10; 6.82)	(10; 9.92)	(13; 6.1)	(13; 6.49)
CVLT-long delay (Max = 9)	2.83	1.67	3.4	1.87	4.6	4.4	5.38	4.08
	(18; 2.28)	(18; 2.47)	(15; 2.17)	(15; 2.2)	(10; 2.41)	(10; 3.66)	(13; 1.5)	(13; 2.4)
CVLT−Recognition Discrim. (d'')	1.58	1.57	1.84	1.41	2.43	2.41	2.05	1.56
	(18; 1.15)	(17; 0.83)	(15; 1.06)	(14; 1.03)	(10; 0.73)	(10; 1.0)	(13; 1.18)	(13; 1.15)
Benson recall (Max = 17)	7.33	7.37	6.4 (15; 4.73)	5.2	9.7	7.2	7.15	6.78
	(18; 4.12)	(16; 5.3)		(10; 4.49)	(10; 4.47)	(10; 6.27)	(13; 4.54)	(9; 6.46)
Memory composite (z-score)	−2.98	−3.46	−2.61	−3.52	−1.67	−2.01	−1.94	−2.86
	(18; 1.42)	(18; 1.56)	(15; 1.43)	(15; 1.51)	(10; 1.37)	(10; 2.18)	(13; 1.47)	(13; 1.68)
Language scores								
BNT (Max = 15)	4.58	3.2	6.0	4.35	12.9	12.0	13.15	12.54
	(19; 4.13)	(20; 3.33)	(16; 3.76)	(17; 3.69)	(10; 2.81)	(10; 4.22)	(13; 1.46)	(13; 2.9)
Semantic fluency (Max in 1 min)	7.17	4.72	7.81	6.56	12.5	10.7	11.31	7.69
	(18; 4.3)	(18; 3.27)	(16; 4.23)	(16; 5.62)	(10; 6.54)	(10; 7.18)	(13; 3.07)	(13; 4.94)
PPVT-R (Max = 16)	9.73	6.4	9.94	7.92	14.11	14.0	14.54	13.0
	(11; 4.5)	(15; 4.37)	(16; 4.02)	(12; 4.03)	(9; 2.32)	(8; 2.0)	(13; 1.71)	(9; 2.74)
PPT (Max = 52)	40.08	40.0	42.41	39.44	47.56	46.67	46.75	45.5
	(13; 7.96)	(13; 6.3)	(17; 7.13)	(9; 9.04)	(9; 6.91)	(6; 8.21)	(12; 4.25)	(6; 5.54)
WAIS-Information (Max = 28)	8.71	7.29	10.63	9.6	16.56	16.71	17.08	16.8
	(14; 5.66)	(14; 5.85)	(16; 5.37)	(10; 6.9)	(9; 7.54)	(7; 8.06)	(12; 5.71)	(5; 3.42)
Semantic−Language composite (z-score)	−5.8	−7.48	−5.17	−6.38	−1.69	−2.12	−1.68	−2.61
	(21; 2.59)	(21; 3.08)	(18; 2.72)	(18; 2.69)	(10; 2.22)	(10; 2.41)	(13; 1.27)	(13; 1.83)
Visuospatial scores								
WAIS-Block D (Max = 68)	35.08	33.67	36.67	37.0	30.44	25.78	29.11	22.14
	(12; 9.24)	(15; 11.76)	(12; 15.47)	(9; 11.64)	(9; 12.24)	(8; 13.97)	(9; 13.82)	(7; 14.89)
Benson Copy (Max = 17)	15.94	15.83	15.45	15.67	15.1	14.8	15.7	14.9
	(16; 1.06)	(18; 1.15)	(11; 0.69)	(12; 0.89)	(10; 2.42)	(9; 1.75)	(10; 0.82)	(10; 1.45)
VOSP (Max = 10)	9.47	9.69	9.18	9.73	8.80	7.78	7.8	6.89
	(15; 1.06)	(16; 0.48)	(11; 1.08)	(11; 0.91)	(10; 1.03)	(8; 3.49)	(10; 1.99)	(9; 2.37)
Beery (Max = 16)	15.07	13.44	13.75	12.73	13.33	11.63	14.33	11.57
	(15; 1.16)	(16; 2.53)	(12; 2.01)	(11; 4.29)	(9; 2.24)	(7; 3.38)	(9; 1.58)	(7; 3.6)
Visuospatial composite (z-score)	0.2	−0.06	−0.17	−0.3	−0.5	−1.49	−0.53	−1.45
	(18; 0.69)	(18; 0.69)	(13; 0.75)	(13; 1.33)	(10; 1.12)	(9; 1.4)	(10; 0.87)	(10; 1.54)
Executive scores								
Design fluency (Max in 1 min)	8.0	6.13	7.92	5.89	6.0	5.0	6.7	3.56
	(13; 3.42)	(15; 3.72)	(13; 2.72)	(9; 3.76)	(9; 2.4)	(7; 3.3)	(10; 3.06)	(9; 2.24)
Lexical fluency (Max in 1 min)	8.13	5.13	7.0	7.0	9.91	8.91	9.3	5.4
	(15; 4.0)	(15; 3.7)	(13; 2.24)	(13; 2.92)	(11; 5.59)	(10; 5.89)	(10; 3.5)	(10; 4.45)
Trails set-shifting (Time to complete)	64.33	69.87	49.75	55.15	73.3	84.18	63.4	86.7
	(15; 37.55)	(15; 38.42)	(12; 29.62)	(13; 31.87)	(10; 42.69)	(10; 44.91)	(10; 41.36)	(10; 43.58)
Stroop (Trials in 1 min)	35.07	30.25	39.31	37.33	25.0	24.6	27.1	15.89
	(14; 10.1)	(12; 15.52)	(13; 18.54)	(6; 17.94)	(11; 13.78)	(9; 16.15)	(10; 10.95)	(9; 14.69)
Digits BW (Max Span)	4.73	4.71	5.15	4.85	4.18	3.55	4.2	3.9
	(15; 0.96)	(14; 1.33)	(13; 0.8)	(13; 1.21)	(11; 1.66)	(10; 1.81)	(10; 1.23)	(10; 1.29)
Executive composite (z-score)	−1.33	−1.79	−1.02	−1.48	−1.65	−2.41	−1.57	−2.47
	(15, 0.81)	(15, 1.1)	(13, 0.75)	(13, 0.95)	(11, 1.0)	(10, 1.53)	(10, 0.91)	(10, 1.21)

a See Methods (Section 2.2) for test names corresponding to abbreviations.

data-driven region that can most reliably quantify longitudinal change in FTLD. We first generated Student's t-statistic (allowing heterogeneous group variances) at each voxel in standardized space to quantify the difference in expansion/contraction between FTLD patients and controls. A map was then generated from the t-statistics. A cross-validated procedure was used to determine the threshold for generating a data-driven ROI that led to maximal effect size. The algorithm proceeded as follows (performed separately in the 1.5 and 3 T datasets):

1 The data for each of controls and patients were randomly divided into training and test sets; here, 16% of the data were used in the test set (n = 13 for 1.5 T and n = 18 for the 3 T dataset). In addition, a restriction was applied such that the proportion of FTLD to normal samples was required to be more than 1/3, but less than 2/3.

2 A series of ROIs was then generated in the training set by thresholding the t-maps over a set of levels ranging from 3.5 to the maximum observed t-statistic in increments of 0.01 units.

3 The effect size for the mean difference in rate of change between the FTLD and controls was then calculated for each ROI of the training set. The curve for the plot of effect sizes for each map represents the relationship between the t-statistic cutoff and the corresponding effect size for each resulting ROI.

4 The ROI associated with the t-statistic cutoff corresponding to the maximum effect size is selected.

5 The effect size of the ROI from step 4 is then calculated in the test set to obtain an unbiased effect size estimate.

It should be noted that the t-statistic cutoff corresponding to the maximal empirical effect size estimate becomes highly variable over neighboring thresholds because of high variability in estimated effect sizes at high t-thresholds; because only a small number of voxels form an ROI at high thresholds. To mitigate this effect and generate a stable estimate of maximum effect size we smoothed the effect size curve plotted against threshold. However, even lowess regression did not sufficiently down-weight the influence of high thresholds. We therefore implemented a heuristic method to identify the maximum effect size. Specifically, a lowess regression was performed after iteratively excluding a top set of t-statistic thresholds (from 0 to 10% of the highest voxels in increments corresponding to those associated with the t-thresholds). At each iteration, the lowess-smoothed maximum was calculated, and the overall maximum was taken as the median of all the smoothed maximums. This approach was able to identify the location of the maximum in reasonable agreement with the choice that one would make visually as being the maximum of the relatively smooth part of the curve.

Steps 1–5 were then repeated 1024 times, reassigning patients into the training and test sets each time. At the end of the process, we have a set of "optimal" ROIs (i.e., across training/test set partitions). To estimate the consensus ROI from the ensemble of cross-validated measurements, we weighted the contribution of each voxel to the data-driven ROI as the proportion of cross-validation partitions (weighted by the effect size for that cross-validation sample) in which the voxel contributes to the "optimal" ROI. Thus, the resulting map has a stronger representation from voxels consistently contributing to the overall effect size across cross-validation samples and weaker representation from voxels whose contribution was more variable.

In order to assess the potential impact of using a data-driven ROI of change for future clinical trials, we compared the effect size observed using the data-driven ROIs to the effect sizes obtained by measuring change within a-priori ROIs based on cerebral anatomy. For this purpose, we used frontal, temporal, combined frontal and temporal, and whole brain masks as regions of interest relevant to FTLD. These ROIs were obtained from the AAL brain atlas supplied with the WFU-PickAtlas software package (Maldjian et al., 2003).

3. Results

3.1. Group demographics and clinical assessments

Demographic characteristics of the 1.5 and 3 T patient groups are presented in Table 1. The two patient groups had similar ratios of males to females. There was no statistically significant difference between the groups in age at the time of the first scan acquisition (p = 0.5). The difference in inter-scan interval between groups was small and approached statistical significance (0.05 year, or about 19 days, p = 0.065). The control subjects at both 1.5 and 3 T were not statistically significantly different from their respective patient groups with respect to inter-scan interval (p = 0.32 and p = 0.22, respectively). Control subjects were, however, statistically significantly older than the patients in both the 1.5 and 3 T comparisons p = 0.043 and p < 0.001, respectively. The control subjects scanned at 1.5 T did not statistically significantly differ from those scanned at 3 T in inter-scan interval (p = 0.96). The 3 T control subjects were significantly older than those scanned at 1.5 T (p = 0.048).

In terms of cognitive and behavioral data, scores were generally what would be expected. SvPPA patients tended to score more poorly on measures of episodic memory and on the language/semantic tasks compared with bvFTD (Table 2). BvFTD patients performed slightly worse on measures of visuospatial and executive function compared with svPPA. All groups declined in all domains of testing. When comparing participants across field-strengths, there were no obvious differences in baseline performance within diagnostic groups (Table 2).

3.2. Change maps and effect sizes

As expected, there was a clear relationship between the t-statistic threshold and the effect size for the associated ROI (Fig. 1). As can be
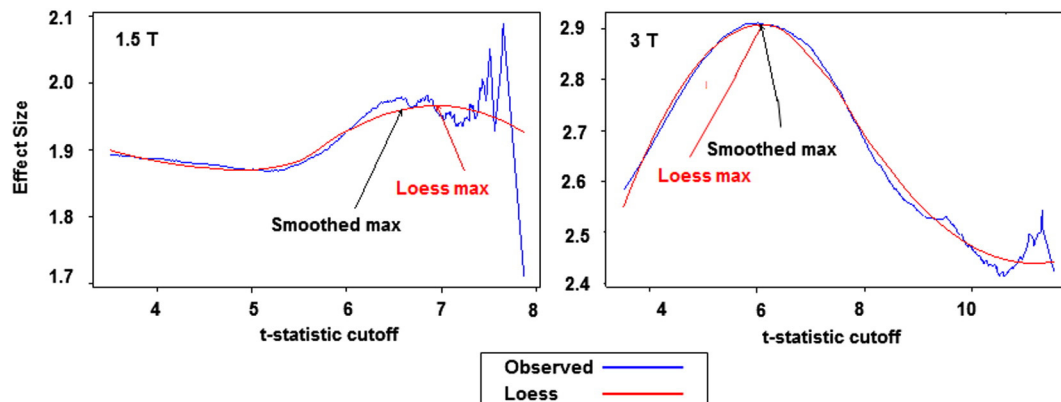

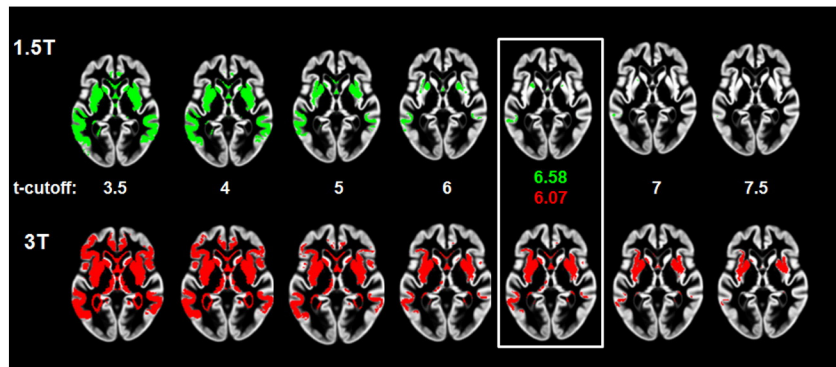**Fig. 1.** Curves for effect size associated with increasing t-statistic cutoff for 1.5 and 3 T.

Fig. 2. Non-cross-validated ROIs associated with a range of t-cutoffs at two field strengths. ROIs chosen as optimal based on effect size curves in Fig. 1, and associated t-cutoff, are enclose in the white box.

seen in Fig. 2, which depicts non-crossvalidated voxel maps at the same axial section associated with various t-cutoff thresholds for each field strength from one iteration of step 2, longitudinal change is detectable in the cortical and subcortical regions of the frontal lobes as well as in the temporal lobes in both analyses. In both datasets, the typical peak effect sizes were associated with t-score cutoffs between 6 and 7. In both the 1.5 T and 3 T data, estimates of effect size at the highest t-thresholds varied greatly, stemming from the small number of voxels at those thresholds. For instance, in the 1.5 T graph in Fig. 1, it is clear that t-cutoff thresholds above approximately 7 are associated with increased variance in effect size. This was the major motivation for using a smoothed estimate for the maximum effect size. While the maps of change are spatially similar at the two field strengths, the extent of the map at the optimal threshold is larger at 3 T compared with 1.5 T.

Fig. 3 depicts the optimal ROIs created at 1.5 and 3 T after crossvalidation. The ROI maps are displayed on a scale from 0 to 1, representing the proportion of times during the cross-validation procedure that the voxel contributed to the optimal ROI. At 3 T, a large number of voxels were identified in the optimal ROI in more than 50% of the iterations, including voxels in the posterior portions of the orbitofrontal cortex, medial frontal region, bilateral insula and underlying striatum, and in the inferior and inferolateral portions of the temporal lobe, excluding the temporal pole and medial temporal regions. The number of voxels at 1.5 T that were included in a high proportion of the optimal

ROIs is much smaller, although these voxels overlapped in location with those seen at 3 T.

Fig. 4 illustrates the distribution of the effect sizes in the test sets estimated during the crossvalidation iterations and Table 2 compares the effect sizes obtained using the statistical ROIs (taken as the mean of the test effect sizes from the cross-validation procedure) with those obtained using anatomically-based ROIs. Also included in the table are sample size estimates for a hypothetical study designed to detect a 20% or 40% reduction in rate of decline between two time points 1 year apart. In both datasets the effect sizes from the data driven ROIs are among the largest although at 1.5 T some of the a-priori ROIs produced larger effect sizes. At 3 T the data-driven ROI outperformed all other ROIs. The most notable difference was in the effect sizes obtained at 1.5 vs. 3 T. For instance the effect size for the data-driven ROI was 1.6 at 1.5 T vs. 2.7 at 3 T. We also estimated the effect size when applying the crossvalidated change map created at 3 T to the 1.5 T dataset, and this resulted in the best effect size at 1.5 T, which was 1.75.

3.3. Effect sizes for clinical measures

Table 3 reports the effect sizes and sample size estimates required to achieve a 20% reduction in rate of decline for MMSE, CDR-SB and the memory, language, spatial, and executive composite scores. The effect sizes are substantially larger than those required for imaging.
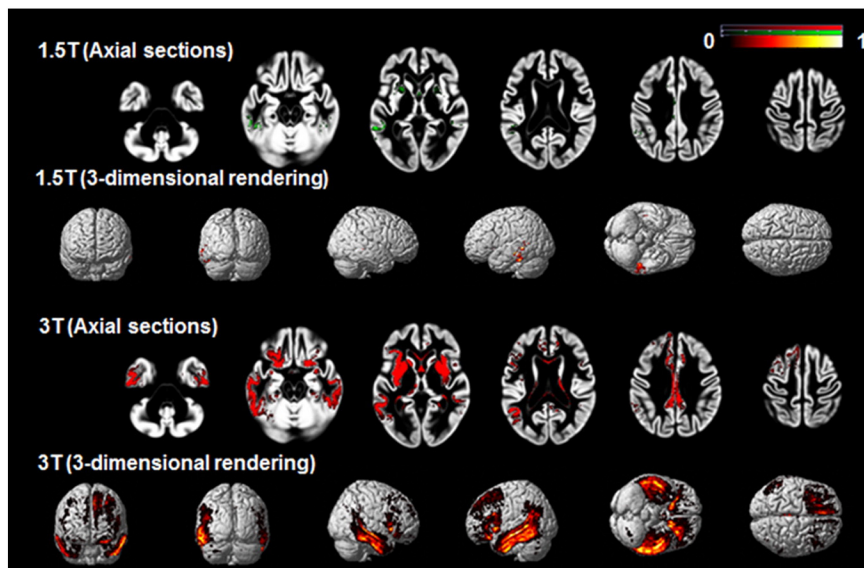

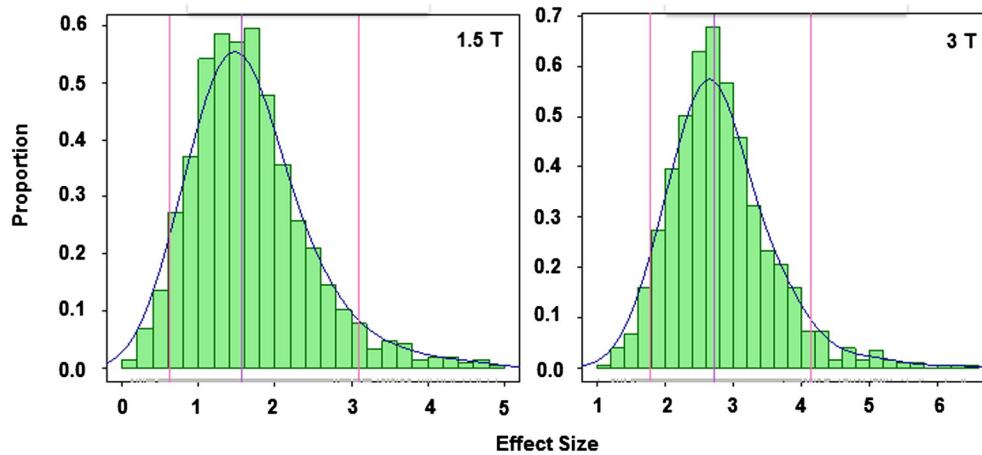**Fig. 3.** Crossvalidated ROIs created at 1.5 and 3 T.

Fig. 4. Histogram of the test set effect sizes for the crossvalidated data. Median and 5th and 95th quantiles are marked by vertical lines.

4. Discussion

The aim of this project was to create an ROI with maximal effect size for measuring change in cortical volume in FTLD, and to compare the effect-size for measuring change within this region to the effect size in a-priori defined anatomical ROIs. The hope is that these data-driven ROI will prove powerful in detecting treatment effects in clinical trials, resulting in reduced sample sizes relative to a-priori defined anatomical ROIs. The approach we implemented generated an ROI that included medial and inferolateral portions of the frontal lobes, insula, striatum, and lateral and inferior portions of the temporal lobes, but no voxels in the medial temporal regions and temporal poles. For images acquired at 3 T, the estimated effect size for longitudinal studies within these data-driven ROI was higher than that generated by all other anatomically-based ROIs examined. We identified a large difference in the reliability of change measures at 1.5 T resulting in smaller ROIs with lower effect sizes compared with 3 T. The findings identify new ROIs that may be able to improve the efficiency of clinical trials in FTLD through MRI volumetric change measurement.

The specific regions identified in the change maps were what would be expected from prior cross-sectional and longitudinal studies of FTLD. Prior cross-sectional analyses have demonstrated that svPPA and bvFTD share regions of cortical volume loss in the medial and orbital frontal regions and insula when compared with controls (Liu et al., 2004; Rosen et al., 2002). Few, if any, studies have examined regions of overlapping longitudinal volume loss in svPPA and bvFTD, but two studies of longitudinal change in svPPA identified regions of change in the medial and orbital frontal regions as well as the temporal lobes (Brambati et al., 2009). The absence of detectable change in the most anterior and medial portions of the temporal lobes is also consistent with prior studies of svPPA, which have interpreted these findings as evidence of a floor effect in volume loss in this disorder (Brambati et al., 2009; Rohrer et al., 2008).

The differences in the estimated change maps between 1.5 and 3 T are potentially important. Large differences in effect size were present

using both the statistically-derived ROI and a-priori ROIs. It should be noted, however, that images from two specific MRI scanners were compared, thus we cannot confirm that the effects are due to field strength per se — they could be machine specific due to other scanner differences. The particular approach to quantification of longitudinal decline using SPM may also be particularly sensitive to the effects of field strength. That said, the approach taken in this paper for quantification of change would be a reasonable approach for a clinical trial, and if the particular methodology were to have such a major impact on volumetric calculations, it would be important to consider that in the design of any future clinical trial. In addition, the sample size calculations derived from our analysis of the 1.5 T data were similar to prior estimates derived from 1-year longitudinal studies of FTLD using different methodologies, which estimated a sample size of 50 (Gordon et al., 2010) to 55 (Knopman et al., 2009) per group to detect a 40% effect on longitudinal decline, compared with our estimate of 40 per group.

It is also notable that the effect sizes for change in most clinical measures were also substantially larger in the 3 T group compared with the 1.5 T group. Upon looking at the characteristics of the groups, there is no obvious explanation for this. The groups were similar in age and sex distribution, and overall similar in baseline cognitive characteristics, except for a slightly higher MMSE at baseline in the 3 T group. In addition, while the effect sizes were larger for cognitive measures in the 3 T group, they were larger in the 1.5 T group for the CDR-SB. Thus, it is difficult to conclude that there was a meaningful difference in the two groups that would predictably affect rates of decline, and we would attribute these differences to chance. Nevertheless, the differences in rates of cognitive decline could be an indicator that some factor other than field strength could have accounted for the difference in rates of brain atrophy between the two groups. Because of all these caveats, it will be important to confirm the effect of field strength on estimates of brain atrophy in other datasets.

One potential benefit from the use of imaging as a marker of longitudinal decline is that increased precision could result in improved effect sizes when compared with clinical measures of change (Weiner et al.,

Table 3
Effect sizes (ES) and sample size calculations for rate of atrophy in a-priori and data-driven ROIs (sample sizes are those required for each arm).

Region	1.5 T			3 T		
	Effect size	Sample size 20% reduction	Sample size 40% reduction	Effect size	Sample size 20% reduction	Sample size 40% reduction
Frontal lobe	1.11	319	81	1.53	170	44
Temporal lobe	1.57	161	41	2.30	76	20
Frontal and temporal	1.66	145	37	2.20	83	22
Lateral ventricles	1.09	332	84	1.50	176	45
Whole brain	1.59	157	40	2.31	75	20
All regions combined	1.60	154	40	2.33	74	20
Data-driven (mean of test ES)	~1.67	155	40	~2.81	55	15

Table 4
Effect sizes and sample size calculations for rate of change in clinical measures.

Clinical measure	1.5 T			3 T			Entire group		
	Sample size (FTLD, HC)	Effect size	Sample size 20% reduction	Sample size (FTLD, HC)	Effect size	Sample size 20% reduction	Sample size (FTLD, HC)	Effect size	Sample size 20% reduction
CDR-SB	30, 18	1.15	298	41, 24	0.84	558	71, 42	0.98	410
MMSE	34, 24	1.05	357	31, 23	0.86	532	65, 47	0.95	436
Memory score	28, 4	0.54	1347	28, 3	1.23	261	56, 7	0.85	545
Language score	31, 23	0.76	681	31, 21	1.39	205	62, 44	0.95	436
Spatial score	28, 23	0.52	1453	23, 7	0.66	902	51, 30	0.54	1347
Executive score	26, 23	1.03	371	23, 22	1.56	163	49, 45	1.29	237

2013). This was generally confirmed in our analysis. For instance, we found that measures of clinical change in a placebo-controlled trial would require 410 subjects per arm using the CDR-SB and 237 subjects per arm using an executive composite score to detect a 20% reduction in rate of change in FTLD (Table 4, entire group). These estimates are roughly consistent with prior studies that estimated sample sizes of 251 (Knopman et al., 2008) and 298 (Gordon et al., 2010) for CDR-SB and 174 (Knopman et al., 2008) for cognitive composite scores to detect a slightly larger effect size of 25%. In contrast, our analysis indicates that a study measuring rates of atrophy using a statistically derived ROI in T1-weighted images acquired at 3 T would require 55 people to detect the same effect.

Our results confirm that this approach for generating data-driven ROIs is reproducible (based on the general shape of the non-crossvalidated ROIs at the two field strengths) and generates expected patterns of longitudinal atrophy based on the known patterns of disease in FTLD and the limited prior data on longitudinal change. Our most dramatic finding of significant differences in the reliability of change maps at different field strengths should be followed up using images acquired on other scanners to more thoroughly establish the effect of field strength on the reliability of longitudinal change estimates.

Acknowledgments

References

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38 (1), 95–113. http://dx.doi.org/10.1016/j.neuroimage.2007.07.00717614438.

Ashburner, J., Ridgway, G.R., 2012. Symmetric diffeomorphic modeling of longitudinal structural MRI. Front. Neurosci. 6, 197. http://dx.doi.org/10.3389/fnins.2012.0019723386806.

Beery, K.E., 1997. The Beery–Buktenica Developmental Test of Visual–Motor Integration. Administration, Scoring and Teaching Manual. fourth edition Modern Curriculum Press, Parsippany, NJ.

Birn, R.M., Kenworthy, L., Case, L., Caravella, R., Jones, T.B., Bandettini, P.A., Martin, A., 2010. Neural systems supporting lexical search guided by letter and semantic category cues: a self-paced overt response fMRI study of verbal fluency. Neuroimage 49 (1), 1099–1107. http://dx.doi.org/10.1016/j.neuroimage.2009.07.03619632335.

Borroni, B., Grassi, M., Archetti, S., Costanzi, C., Bianchi, M., Caimi, L., Caltagirone, C., Di Luca, M., Padovani, A., 2009. BDNF genetic variations increase the risk of Alzheimer's disease-related depression. J. Alzheimers Dis. 18 (4). http://dx.doi.org/10.3233/JAD-2009-119119661618.

Boxer, A.L., Knopman, D.S., Kaufer, D.I., Grossman, M., Onyike, C., Graf-Radford, N., Miller, B.L., 2013. Memantine in patients with frontotemporal lobar degeneration: a multicentre, randomised, double-blind, placebo-controlled trial. Lancet Neurol. 12 (2), 149–156. http://dx.doi.org/10.1016/S1474-4422(12)70320-423290598.

Brambati, S.M., Rankin, K.P., Narvid, J., Seeley, W.W., Dean, D., Rosen, H.J., Gorno-Tempini, M.L., 2009. Atrophy progression in semantic dementia with asymmetric temporal involvement: a tensor-based morphometry study. Neurobiol. Aging 30 (1), 103–111. http://dx.doi.org/10.1016/j.neurobiolaging.2007.05.01417604879.

Chen, K., Langbaum, J.B., Fleisher, A.S., Ayutyanont, N., Reschke, C., Lee, W., Alzheimer's Disease Neuroimaging Initiative, 2010. Twelve-month metabolic declines in probable Alzheimer's disease and amnestic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: findings from the Alzheimer's

disease neuroimaging initiative. Neuroimage 51 (2), 654–664. http://dx.doi.org/10.1016/j.neuroimage.2010.02.06420202480.

Delis, D., Kaplan, E.B., Kramer, J., 2001. The Delis–Kaplan Executive Function System. the psychological corporation, San Antonio, TX.

Delis, D.C., Kramer, J.H., Kaplan, E., Ober, B.A., 2000. California Verbal Learning Test. second edition the psychological corporation, San Antonio, TX.

Dickerson, B.C., Stoub, T.R., Shah, R.C., Sperling, R.A., Killiany, R.J., Albert, M.S., Detoledo-Morrell, L., 2011. Alzheimer-signature MRI biomarker predicts AD dementia in cognitively normal adults. Neurology 76 (16), 1395–1402. http://dx.doi.org/10.1212/WNL.0b013e3182166e9621490323.

Dunn, L.M., Dunn, L.M., 1997. Peabody Picture Vocabulary Test — Third Edition. American Guidance Service, Circle Pines, MN.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state". A practical method for grading the mental state of patients for the clinician. J. Psychiat. Res. 12 (3), 189–198. http://dx.doi.org/10.1016/0022-3956(75)90026-61202204.

Friston, K.J., Holmes, A., Poline, J.B., Price, C.J., Frith, C.D., 1996. Detecting activations in PET and fMRI: levels of inference and power. Neuroimage 4 (3 Pt 1), 223–235. http://dx.doi.org/10.1006/nimg.1996.00749345513.

Gordon, E., Rohrer, J.D., Kim, L.G., Omar, R., Rossor, M.N., Fox, N.C., Warren, J.D., 2010. Measuring disease progression in frontotemporal lobar degeneration: a clinical and MRI study. Neurology 74 (8), 666–673. http://dx.doi.org/10.1212/WNL.0b013e3181d1a87920177120.

Howard, D., Patterson, K., 1992. Pyramids and Palm Trees: A Test of Semantic Access from Pictures and Words. Thames Valley Publishing Company, Bury St Edmunds, Suffolk.

Hua, X., Lee, S., Yanovsky, I., Leow, A.D., Chou, Y.Y., Ho, A.J., Alzheimer's Disease Neuroimaging Initiative, 2009. Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects. Neuroimage 48 (4), 668–681. http://dx.doi.org/10.1016/j.neuroimage.2009.07.01119615450.

Jack Jr., C.R., Petersen, R.C., Xu, Y.C., Waring, S.C., O'Brien, P.C., Tangalos, E.G., Kokmen, E., 1997. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. Neurol. 49 (3), 786–794. http://dx.doi.org/10.1212/WNL.49.3.7869305341.

Kaplan, E., Goodglass, H., Wintraub, S., 1983. The Boston Naming Test. Lea & Febiger, Philadelphia.

Knopman, D.S., Jack Jr., C.R., Kramer, J.H., Boeve, B.F., Caselli, R.J., Graff-Radford, N.R., Mercaldo, N.D., 2009. Brain and ventricular volumetric changes in frontotemporal lobar degeneration over 1 year. Neurology 72 (21), 1843–1849. http://dx.doi.org/10.1212/WNL.0b013e3181a7123619470967.

Knopman, D.S., Kramer, J.H., Boeve, B.F., Caselli, R.J., Graff-Radford, N.R., Mendez, M.F., Mercaldo, N., 2008. Development of methodology for conducting clinical trials in frontotemporal lobar degeneration. Brain 131 (11), 2957–2968. http://dx.doi.org/10.1093/brain/awn23418829698.

Knopman, D.S., Petersen, R.C., Edland, S.D., Cha, R.H., Rocca, W.A., 2004. The incidence of frontotemporal lobar degeneration in Rochester, Minnesota, 1990 through 1994. Neurology 62 (3), 506–508. http://dx.doi.org/10.1212/01.WNL.0000106827.39764.7E14872045.

Kramer, J.H., Jurik, J., Sha, S.J., Rankin, K.P., Rosen, H.J., Johnson, J.K., Miller, B.L., 2003. Distinctive neuropsychological patterns in frontotemporal dementia, semantic dementia, and Alzheimer disease. Cogn. Behav. Neurol. 16 (4), 211–218. http://dx.doi.org/10.1097/00146965-200312000-0000214665820.

Krueger, C.E., Dean, D.L., Rosen, H.J., Halabi, C., Weiner, M., Miller, B.L., Kramer, J.H., 2010. Longitudinal rates of lobar atrophy in frontotemporal dementia, semantic dementia, and Alzheimer's disease. Alzheimer Dis. Assoc. Disord. 24 (1), 43–48. http://dx.doi.org/10.1097/WAD.0b013e3181a6f10119571735.

Liu, W., Miller, B.L., Kramer, J.H., Rankin, K., Wyss-Coray, C., Gearhart, R., Rosen, H.J., 2004. Behavioral disorders in the frontal and temporal variants of frontotemporal dementia. Neurology 62 (5), 742–748. http://dx.doi.org/10.1212/01.WNL.0000113729.77161.C915007124.

Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19 (3), 1233–1239. http://dx.doi.org/10.1016/S1053-8119(03)00169-112880848.

Morris, J.C., 1997. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. Int. Psychogeriatr. 9 (Suppl. 1), 173–176. http://dx.doi.org/10.1017/S104161029700487094447441.

Papageorgiou, S.G., Kontaxis, T., Bonakis, A., Kalfakis, N., Vassilopoulos, D., 2009. Frequency and causes of early-onset dementia in a tertiary referral center in Athens. Alzheimer Dis. Assoc. Disord. 23 (4), 347–351. http://dx.doi.org/10.1097/WAD.0b013e31819e6b2819568157.

Rascovsky, K., Hodges, J.R., Knopman, D., Mendez, M.F., Kramer, J.H., Neuhaus, J., 2011. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 134 (9), 2456–2477. http://dx.doi.org/10.1093/brain/awr17921810890.

Ratnavalli, E., Brayne, C., Dawson, K., Hodges, J.R., 2002. The prevalence of frontotemporal dementia. Neurology 58 (11), 1615–1621. http://dx.doi.org/10.1212/WNL.58.11.161512058088.

Roberson, E.D., Hesse, J.H., Rose, K.D., Slama, H., Johnson, J.K., Yaffe, K., Miller, B.L., 2005. Frontotemporal dementia progresses to death faster than Alzheimer disease. Neurology 65 (5), 719–725. http://dx.doi.org/10.1212/01.wnl.0000173837.82820.9f16157905.

Rohrer, J.D., Clarkson, M.J., Kittus, R., Rossor, M.N., Ourselin, S., Warren, J.D., Fox, N.C., 2012. Rates of hemispheric and lobar atrophy in the language variants of frontotemporal lobar degeneration. J. Alzheimers Dis. 30 (2), 407–411. http://dx.doi.org/10.3233/JAD-2012-11155622406442.

Rohrer, J.D., McNaught, E., Foster, J., Clegg, S.L., Barnes, J., Omar, R., Fox, N.C., 2008. Tracking progression in frontotemporal lobar degeneration: serial MRI in semantic dementia. Neurology 71 (18), 1445–1451. http://dx.doi.org/10.1212/01.wnl.0000327889.13734.cd18955688.

Rosen, H.J., Gorno-Tempini, M.L., Goldman, W.P., Perry, R.J., Schuff, N., Weiner, M., Miller, B.L., 2002. Patterns of brain atrophy in frontotemporal dementia and semantic dementia. Neurology 58 (2), 198–208. http://dx.doi.org/10.1212/WNL.58.2.19811805245.

Rosso, S.M., Donker Kaat, L., Baks, T., Joosse, M., de Koning, I., Pijnenburg, Y., van Swieten, J.C., 2003. Frontotemporal dementia in The Netherlands: patient characteristics and prevalence estimates from a population-based study. Brain 126 (9), 2016–2022. http://dx.doi.org/10.1093/brain/awg20412876142.

Schuff, N., Tosun, D., Insel, P.S., Chiang, G.C., Truran, D., Aisen, P.S., Alzheimer's Disease Neuroimaging Initiative, 2012. Nonlinear time course of brain volume loss in cognitively normal and impaired elders. Neurobiol. Aging 33 (5), 845–855. http://dx.doi.org/10.1016/j.neurobiolaging.2010.07.01220855131.

Seelaar, H., Rohrer, J.D., Pijnenburg, Y.A., Fox, N.C., van Swieten, J.C., 2011. Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review. J. Neurol. Neurosurg. Psychiatry 82 (5), 476–486. http://dx.doi.org/10.1136/jnnp.2010.21222520971753.

Seeley, W.W., Crawford, R., Rascovsky, K., Kramer, J.H., Weiner, M., Miller, B.L., Gorno-Tempini, M.L., 2008. Frontal paralimbic network atrophy in very mild behavioral variant frontotemporal dementia. Arch. Neurol. 65 (2), 249–255. http://dx.doi.org/10.1001/archneurol.2007.3818268196.

Stroop, J.R., 1935. Studies of interferences in serial verbal reactions. J. Exp. Psychol. 18 (6), 643–662. http://dx.doi.org/10.1037/h0054651.

Tartaglia, M.C., Rosen, H.J., Miller, B.L., 2011. Neuroimaging in dementia. Neurotherapeutics 8 (1), 82–92. http://dx.doi.org/10.1007/s13311-010-0012-221274688.

Thomas, A., 2010. The conditional independences between variables derived from two independent identically distributed Markov random fields when pairwise order is ignored. Math. Med. Biol. 27 (3), 283–288. http://dx.doi.org/10.1093/imammb/dqp02219942608.

Warrington, E.K., James, M., 1991. The Visual Object and Space Perception Battery. Thames Valley Test Company, Bury St Edmunds.

Wechsler, D., 1997. Wechsler Adult Intelligence Scale—Third Edition (WAIS-III). the psychological corporation, San Antonio, TX.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Alzheimer's Disease Neuroimaging Initiative, 2013. The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. Alzheimers Dement. 9 (5), e111–e194. http://dx.doi.org/10.1016/j.jalz.2013.05.176923932184.

Yushkevich, P.A., Avants, B.B., Das, S.R., Pluta, J., Altinay, M., Craige, C., Alzheimer's Disease Neuroimaging Initiative, 2010. Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: an illustration in ADNI 3 T MRI data. Neuroimage 50 (2), 434–445. http://dx.doi.org/10.1016/j.neuroimage.2009.12.00720005963.