Longitudinal Neuroimaging Measures of Volumetric Change across the Frontotemporal Dementia Spectrum

A thesis submitted to University College London for the degree of Doctor of Philosophy

Elizabeth Gordon

Dementia Research Centre

Institute of Neurology

Queen Square

UCL

2020

I, Elizabeth Gordon confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

……..

Abstract

Frontotemporal dementia (FTD) is a common cause of young onset dementia, encompassing several clinical, genetic and pathological subgroups. Currently there are no treatments, but there are promising candidates in development. However, proven biomarkers of disease progression in FTD are lacking and urgently needed to facilitate these trials.

Investigating large sporadic and genetic FTD cohorts, this thesis provides a comprehensive comparison of longitudinal neuroimaging measures of structural change within the clinical, genetic and pathological FTD subgroups. Effect size and sample size estimates are computed to explore the feasibility of these brain measures as surrogate markers of disease progression in order to detect disease-modifying treatment effects.

The first project compares 17 automated techniques for extracting whole-brain atrophy measures. Many of the techniques showed great promise, producing sample sizes of substantially less than 100 patients required to detect a disease-modifying effect. Significant differences in performance were found between both techniques and patient subgroups, highlighting the importance of informed biomarker choice in matching the optimal marker to the patient group to be enrolled in a trial.

In the following chapters, I explored lobar and subcortical change across the disease spectrum. The different patient subgroups presented with unique profiles of change but, interestingly, automated measures of temporal lobe, caudate and thalamic atrophy proved to be particularly sensitive markers of change, producing low sample size estimates across the FTD subgroups. Importantly, I found significantly increased rates of amygdala, hippocampus, caudate and thalamic atrophy in differing patterns across presymptomatic mutation carriers, providing the first comprehensive assessment of the utility of such markers for early therapeutic intervention at this ideal stage before symptoms develop.

In summary, this work expands current knowledge and builds on the limited longitudinal investigations currently available in FTD, as well as providing valuable information about the potential of non-invasive biomarkers for sporadic and genetic FTD trials.

Impact statement

The work in this thesis provides novel findings that contribute to the FTD imaging research field, both locally and internationally, and also provides data to inform non-invasive biomarker choice for commercial trials testing potential disease-modifying therapies across this patient population.

Chapter 3 presents the first direct head-to-head comparison of automated whole-brain segmentation and atrophy measures in a large cohort encompassing all key FTD subgroups. The results indicate important differences in performance between these widely used techniques that have significant implications for future imaging research, particularly longitudinal research involving patient images. Furthermore, it confirms the utility of many of these fully automated techniques to detect putative disease-modifying effects with only small sample sizes being required over a single year interval. This type of data contributes valuable insights for designing time- and cost-effective trials in terms of biomarker choice to detect a meaningful disease-modifying effect.

Chapter 4 employs a newly developed method for measuring longitudinal lobar changes, not previously validated on a large neurodegenerative patient population. The results confirm the utility of this new technique to reliably measure regional atrophy within each lobe, revealing important differences in the lobar measures that performed best in each FTD patient subgroup. This provided a valuable validation dataset for the developers of this technique. In addition, it demonstrates its utility in automatically delineating and detecting atrophy in these important lobar regions for future neurodegenerative imaging research and provides further data to inform preferential biomarker choice depending on patient enrolment criteria in interventions.

Investigating more focal regions, Chapter 5 presents the first comprehensive FTD dataset investigating automated subcortical measures of amygdala,

hippocampus, caudate and thalamus and accompanying sample size calculations. These structures are important in many of the symptoms and difficulties experienced by FTD patients but have not been previously investigated longitudinally across the different FTD presentations. The dataset provides valuable insights into differential disease progression in these patients.

The work in Chapters 3, 4 and 5 provides the first neuroanatomical investigation into a newly recognised distinct clinical subtype of PPA. This investigation delivers valuable insights into the presentation and progression of this disorder, thus providing the clinical benefit of additional neuroanatomical information to inform diagnosis and prognosis for patients and their families.

Finally, with the advent of clinical trials aimed at preventing or halting disease progression before symptoms occur in genetic FTD, the work presented in Chapter 6 confirms the ability to detect significant longitudinal volume loss in presymptomatic *MAPT*, *C9orf72* and *GRN* mutation carriers many years prior to symptom onset. This is fundamental to the development of sensitive and specific markers that could detect a disease-modifying effect in these neuroanatomical changes before individuals present with the devastating symptoms associated with FTD.

Taken together, this work contributes substantially to our understanding of the differential neuroanatomical evolution and rate of disease progression across FTD and confirms the potential of fully automated measures of volumetric change across global, lobar and subcortical structures. These data have important implications for the design of clinical trials in sporadic and genetic FTD.

Acknowledgements

Firstly, I am thankful to all the patients, families and carers who have tirelessly and selflessly given their time and energy to contribute to the research in our department over so many years. The commitment, generosity and resolve they have shown in the face of such difficult circumstances is an inspiration, and I am indebted to each and every one of them.

I am grateful to my supervisors Dr Jonathan Rohrer, Dr Martina Bocchetta and Dr David Cash whose vision and guidance made this thesis possible. I am particularly thankful to Martina, who has always gone above and beyond to help. Thank you for not only showing unwavering faith in my potential but for being a constant source of encouragement, strength and wisdom. I will be eternally grateful for your insights both in work and in life, and most of all for the privilege of your friendship. My deepest thanks also go to Prof. Nick Fox, who allowed me to continue to manage and lead the image analysis team at the DRC. Your constant support and flexibility in ensuring I continued my professional development alongside my academic pursuits has been of immeasurable benefit. I would also like to thank The Alzheimer's Society who provided funding for this work. Their support and enthusiasm for this work was a great source of motivation.

I have been fortunate to be surrounded by many exceptional colleagues and female scientists during my journey at the DRC and throughout the PhD process. In particular, I'd like to thank Katrina Moore, Martina Bocchetta, Laila Ahsan, Emily Manning, Mica Clarke, Lucy Russell, Ione Woollacott, Caroline Greaves, Rachelle Schafei, Rhian Convery and Heidi Murray-Smith for making it a joy to come into the office every day. Special thanks to Jenny Nicholas for so many hours of statistical help and to Elizabeth Halton for being a beacon of advice and encouragement. I am very fortunate to have so many of my colleagues as good friends.

Very special thanks go to all my friends and family for their support and understanding. In particular, I am immensely grateful to have experienced this PhD journey alongside Katrina Moore and for these years of support and friendship. None of this would be possible without my parents, Carmel and Ian. Thank you for so kindly taking the time to read through this thesis and for always being such an inspiration to this journey over so many years and in so many ways.

Finally, my endless thanks go to my incredible husband, Rob, and to our beautiful little miracle twins, Joshua and Caleb. Rob, your patience, love, wisdom and understanding have meant the world to me and have kept me on the right side of sane during these last few years. Thank you to our gorgeous boys for all the light and love you enrich our lives with. Even during a global pandemic, your beaming faces bring such joy and remind me what it is to know true gratitude. I thank you all.

Table of Contents

Table of Figures

Table of Tables

Abbreviations	
AAO	Age at Symptom Onset
ABSI	Amygdala Boundary Shift Integral
AD	Alzheimer's Disease
ARTFL	Advancing Research and Treatment for FTLD
BCa	Bias-corrected and cccelerated
BET	Brain Extraction Tool
BMAPS	Brain Multi-Atlas Propagation and Segmentation
BSI	Boundary Shift Integral
BV	Brain Volume
bvFTD	behavioural variant Frontotemporal Dementia
C9orf72	chromosome 9 open reading frame 72
CBS	Corticobasal Syndrome
CBSI	Caudate Boundary Shift Integral
CDR	Clinical Dementia Rating scale
CHMP2B	Charged Multivesicular Body Protein 2B gene
CI	Confidence Interval
CMIC	Centre for Medical Image Computing
Coef.	Coefficient
CSF	Cererbospinal Fluid
DIAN	Dominantly Inherited Alzheimer Network
DLPFC	Dorsolateral Prefrontal Vortex
DOF	Degrees of Freedom
DRC	Dementia Research Centre
EYO	Estimated Years from symptom Onset
FAST	FMRIB Automated Segmentation Tool
FLIRT	FMRIB Linear Image Registration Tool
FMRIB	(Oxford Centre for) Functional Magnetic Resonance Imaging of the Brain
FSL	FMRIB Software Library
FTD	Frontotemporal Dementia
FTD-MND	Frontotemporal Dementia - Motor Neuron Disease
FTLD	Frontotemporal Lobar Degeneration
FUS	fused in sarcoma
GBSI	Generalised Boundary Shift Integral
GENFI	Genetic Frontotemporal dementia Initiative
GIF	Geodisic Information Flows
GM	Grey Matter
GRN	Granulin
HBSI	Hippocampus Boundary Shift Integral
HD	Huntingon's Disease

IFG	Inferior Frontal Gyrus
IQR	Interquartile Range
KBSI	K-means Boundary Shift Integral
LEFFTDS	Longitudinal Evaluation of Familial Frontotemporal Dementia Subjects
LNCC	locally normalised cross-correlation
lvPPA	logopenic variant Primary Progressive Aphasia
MALPEM	Multi-Atlas Label Propagation with Expectation-Maximisation-based refinement
MAPER	Multi-atlas propagation with enhanced registration
MAPT	Microtubule-Associated Protein Tau
MCI	Mild Cognitive Impairment
MIDAS	Medical Information Display and Analysis System
MNI	Montreal Neurological Institute
MRI	Magnetic Resonance Imaging
NFL	Neurofilament Light
nfvPPA	non-fluent variant Primary Progressive Aphasia
PACS	Picture Archiving and Communication System
PBVC	Percentage Brain Volume Change
PET	Positon Emission Tomography
PieBSI	Pie Boundary Shift Integral
PPA	Primary Progressive Aphasia
PPA-M	Primary Progressive Aphasia - Mixed Disease
PPA-NOS	Primary Progressive Aphasia - Not Otherwise Specified
PSP	Progressive Supranuclear Palsy
QC	Quality Control
ROI	Region of Interest
SIENA	Structural Image Evaluation, using Normalisation, of Atrophy
SIENAX	Structural Image Evaluation, using Normalisation, of Atrophy – cross-sectional
SPM	Statistical Parametric Mapping
SQSTM1	Sequestosome 1 gene
STAPLE	Simultaneous truth and performance level estimation
STEPS	Similarity and Truth Estimation for Propagated Segmentations
svPPA	semantic variant Primary Progressive Aphasia
TARDBP	transactive response DNA binding protein gene
TBK1	TANK Binding Kinase 1 gene
TBSI	Thalamus Boundary Shift Integral
TDP-43	TAR DNA-binding protein 43
TE	Echo Time
TI	Inversion Time
TIV	Total Intercranial Volume
TR	Repetition Time

UCL	University College London
UK	United Kingdom
VBM	Voxel Based Morphometry
VCP	Valosin-Containing Protein
WM	White Matter
XOR	Exclusive OR

1 Introduction

1.1 The FTD spectrum

Frontotemporal dementia (FTD) refers to a heterogeneous spectrum of clinically, genetically and pathologically diverse but related familial and sporadic neurodegenerative disorders. Together, they constitute a common and devastating cause of young-onset dementia (Onyike & Diehl-Schmid, 2013). FTD occurs with approximately equal frequency as Alzheimer's disease (AD) in individuals under 65 years of age and is associated with significant personal and financial loss for patients, their families and wider society.

Clinically, FTD patients predominantly present with either changes in behaviour, social conduct and personality (behavioural variant (bvFTD)), or language impairment (primary progressive aphasia, PPA). PPA can be further divided into three main subtypes, semantic variant (svPPA), non-fluent variant (nfvPPA) and logopenic variant (lvPPA) (Gorno-Tempini et al., 2011). In addition, there is a subgroup of patients with a presentation that does not fulfil the diagnostic criteria for the three canonical PPA subtypes, referred to as PPA-NOS (Not Otherwise Specified) (Marshall et al., 2018; Rohrer, Rossor, & Warren, 2010). Patients may also develop motor deficits, presenting with features of motor neuron disease (FTD-MND), corticobasal syndrome (CBS) or progressive supranuclear palsy (PSP) (Convery, Mead, & Rohrer, 2019; Woollacott, 2016). As the disease progresses, many patients also develop symptoms that overlap with clinical features characteristic of other FTD subtypes and even within the canonical diagnostic features of bvFTD, further clinical heterogeneity exists. Some patients present with more disinhibition whilst others are more apathetic; some show prominent appetite change whilst other patients do not.

Alongside a clinically variable picture, there are also several genetic and pathological subgroups identified that underlie these clinical manifestations (Figure 1-1). FTD is a highly heritable set of disorders, with approximately a third of patients exhibiting an autosomal dominant form (Greaves & Rohrer,

2019; Rohrer & Warren, 2011). Mutations in the microtubule-associated protein tau (*MAPT*) and progranulin (*GRN*) genes or a hexanucleotide expansion in the chromosome 9 open reading frame 72 (*C9orf72*) gene are the major contributors (Rohrer, Guerreiro et al., 2009; Warren et al., 2013). Several rare pathogenic mutations have also been identified in a small number of families including VCP, TARDBP, TIA1, TBK1, CCNF and CHMP2B (Greaves & Rohrer, 2019). Pathologically, neuronal inclusions usually contain abnormal forms of one of three proteins: tau, TAR DNA-binding protein 43 (TDP-43) or fused in sarcoma (FUS), with the latter presenting less frequently (Lashley, Rohrer, Mead, & Revesz, 2015; Mackenzie et al., 2011, 2010; Seelaar et al., 2010) (Figure 1-1).

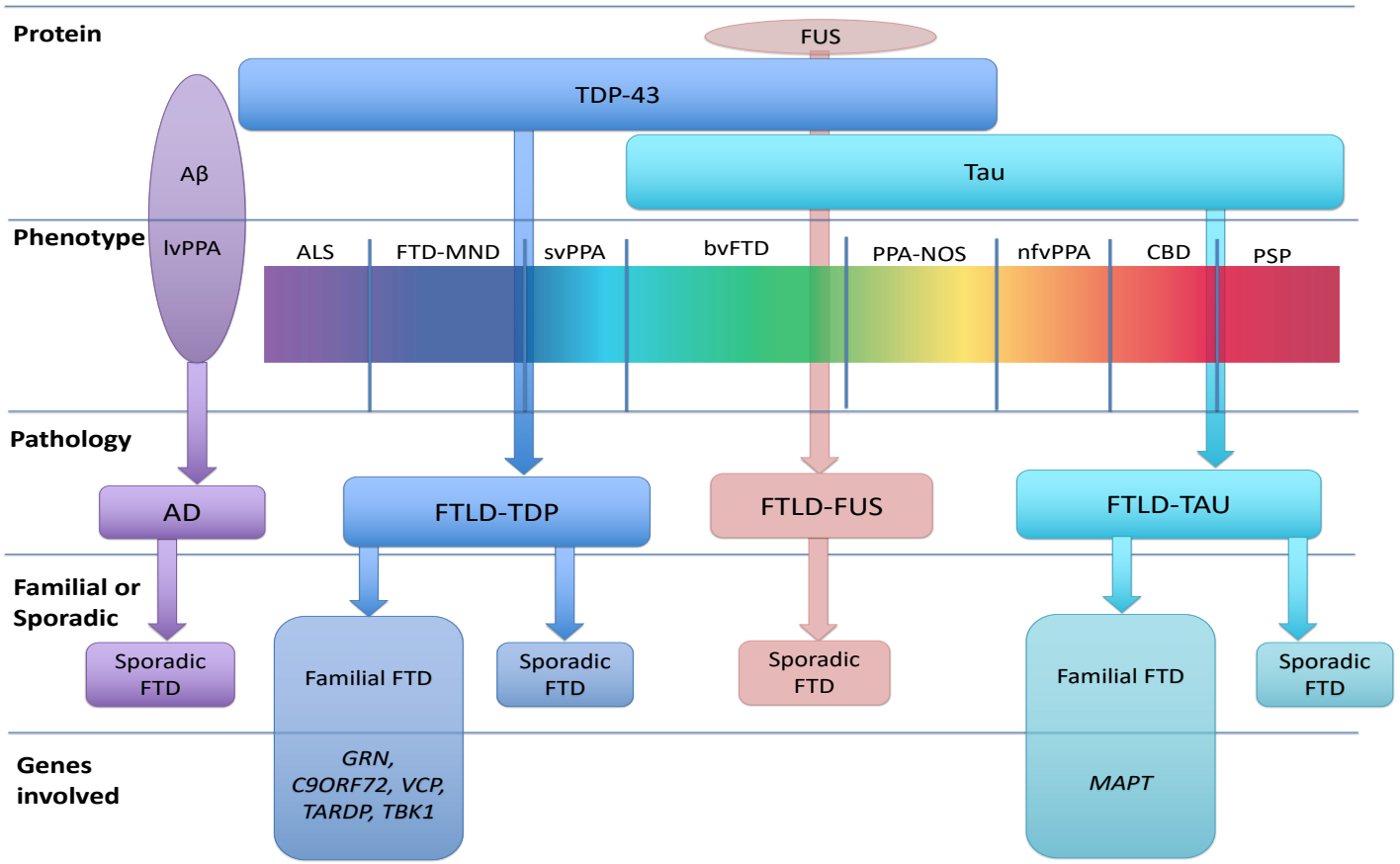

Figure 1-1 Clinical, genetic and pathological spectrum of FTD. Genetic (familial) forms of FTD have predictable pathology: *GRN* mutations and *C9orf72* repeat expansions result in TDP-43 pathology, whereas *MAPT* mutations result in tau pathology. By contrast, variable underlying pathologies and genetic forms are found across the clinical spectrum of FTD. bvFTD, behavioural variant FTD; CBD, corticobasal degeneration; FUS, RNA-binding protein FUS; nfvPPA, non-fluent variant primary progressive aphasia; PSP, progressive supranuclear palsy; svPPA, semantic variant primary progressive aphasia; TDP-43, transactive response DNA-binding protein. Adapted from Meeter, Kaat, Rohrer, & Van Swieten (2017).

This inherent heterogeneity and lack of direct correspondence between the clinical, genetic and pathological presentations not only delays diagnosis and hinders clarity of prognosis for patients and families but it also makes it difficult to target recruitment and recommend suitable endpoints in putative FTD treatment trials. Whilst there are currently no proven disease-modifying treatments available, there are several promising candidates in early stages of clinical trials in humans. With these trials now under way and an increasing number of candidate therapies in development, the need for developing and validating longitudinal non-invasive surrogate markers capable of measuring potential treatment effects is increasingly pressing.

Longitudinal structural changes measured from magnetic resonance imaging (MRI) have been extensively used as surrogate markers of potential treatment effects in multiple clinical trials in AD (Cash et al., 2014) and more recently in Huntington's Disease (HD) (Tabrizi et al., 2019; Zeun, Scahill, Tabrizi, & Wild, 2019). An increasing body of literature suggests that structural MRI measures will prove useful in FTD as well. A recent article investigating longitudinal multimodal imaging and clinical end-points for FTD trials demonstrated that measures of longitudinal structural imaging generally provided the lowest estimated sample sizes required to detect a disease-modifying treatment effect across the clinical spectrum, compared with other imaging and clinical measures (Staffaroni et al., 2019). However, in-depth investigation of longitudinal structural imaging measures across the full FTD spectrum is currently lacking and urgently needed as we move towards treatments directly targeting FTD-associated pathologies.

Whole-brain atrophy rates have been shown to produce feasible sample size estimates in FTD that would reveal clinically meaningful effects (Gordon, Rohrer, & Fox, 2016; Gordon et al., 2010; Knopman et al., 2009). More focal measures of atrophy, such as lobar volumes, have been shown to produce smaller sample sizes in the language variants of FTD (Rohrer et al., 2008); however, comprehensive investigations in bvFTD and genetic- or pathology-

confirmed subgroups is required. Characterised as a focal disease, targeted regional measurements in FTD could provide additional value to the global measures that have been shown to perform well across the full spectrum of patients. Neuroimaging findings of regions particularly vulnerable to volume loss suggest several candidate structures within the frontal and temporal lobe, insula, anterior cingulate and subcortical regions such as the hippocampus, amygdala, caudate and thalamus (Agosta, Canu, Sarro, Comi, & Filippi, 2012; Gordon et al., 2016).

Given the high heritability of FTD, investigations involving presymptomatic individuals who are carriers of one of the known mutations provides a unique opportunity to study the earliest stages of the disease. This, in turn, would allow for assessing the feasibility of early treatment interventions that could slow or halt the progression of the disease before symptom onset and the extensive neuronal damage underlying that onset. Validating measures of change in these presymptomatic FTD populations would prove invaluable for moving towards this goal. Several cross-sectional studies have demonstrated the sensitivity of structural MRI measurements to detect volumetric differences between mutation carriers and non-carrier controls, some more than 10 years before expected symptom onset (Cash et al., 2018; Panman et al., 2019; Popuri et al., 2018; Rohrer et al., 2015); however, longitudinal investigations are currently lacking.

1.2 Structure of the Introduction

The focus of this PhD is to investigate automated neuroimaging methods for quantifying longitudinal structural changes across the different clinical, genetic and pathological subgroups within the FTD spectrum and evaluate their potential as surrogate markers of treatment efficacy for current and upcoming therapeutic interventions. Because FTD incorporates several different clinical presentations, I will first provide an overview of the core FTD symptom profiles. This is followed by an overview of neuroimaging research that has provided insights into the different cross-sectional neuroanatomical profiles of volume loss in the different clinical, genetic and pathological FTD subgroups. Building

on the cross-sectional characterisation, the following sections lead on to explaining the core focus of the current PhD work. This will include an overview of previous investigations of longitudinal neuroanatomical changes in FTD, how these structures are generally measured and why capturing these changes using structural neuroimaging provides promising markers for therapeutic interventions and thus warrants more comprehensive investigation. Finally, the key aims and hypotheses of the thesis will be clarified and how they relate to each PhD projects.

1.3 Behavioural, language and functional change within FTD

Despite its earliest description dating back to 1892, it is only relatively recently that wider efforts have been made to better understand and characterise the extensive range of behavioural, language and functional changes that occur within FTD. The nosology of FTD has undergone multiple revisions to incorporate emerging findings (Mackenzie et al., 2010; Woollacott & Rohrer, 2016) and the following presents an overview of the most current classifications with reference to core diagnostic consensus criteria.

1.3.1 Behavioural variant of Frontotemporal dementia

BvFTD is the most common and most heterogeneous clinical presentation within the spectrum, accounting for approximately half of patients diagnosed (Rascovsky et al., 2011; Rohrer, Guerreiro, et al., 2009). It involves a progressive decline in social conduct, changes in behaviour and personality and difficulty with executive functioning or planning. The initial symptoms may be subtle, and patients often present with relative sparing of other cognitive domains such as visuospatial abilities and episodic memory. This profile can result in misdiagnosis as a psychiatric rather than neurodegenerative condition, especially in those with a younger age of symptom onset. The core behavioural changes fall into five main categories: apathy, disinhibition, loss of empathy, development of compulsive or ritualistic behaviours and changes in food preferences (Table 1-1 summarises the current criteria). To qualify for a diagnosis of possible bvFTD, an individual must present with at least three of

these changes. Where supporting imaging is available and demonstrates characteristic patterns of volume loss, the diagnosis will advance to probable bvFTD (Rascovsky et al., 2011). It is important to note that additional behavioural symptoms including loss of insight (Mendez & Shapira, 2011), changes in sensitivity and awareness to temperature and pain (Fletcher et al., 2015) and several neuropsychiatric symptoms (Schroeter et al., 2014; Seeley, 2019) are common, particularly in *C9orf72* patients, but are not included in the core diagnostic criteria.

Table 1-1 Diagnostic criteria for bvFTD

Behavioural/cognitive symptoms – diagnosis of possible bvFTD requires at least 3 of the following symptoms to be fulfilled: *	Examples of specific symptoms
Early behavioural disinhibition ≥ 1 of:	
Socially inappropriate behaviour	Staring, inappropriate physical contact with strangers, verbal or physical aggression
Loss of manners or decorum	Lack of social etiquette, insensitive or rude comments, preference for crass jokes and slapstick humour, inappropriate choices of clothing or gifts
Impulsive, rash or careless actions	New gambling behaviour, driving or investing recklessly, overspending, gullibility to phishing/Internet scams
Early apathy or inertia ≥ 1 of:	
Apathy	Reduced drive, stops previous hobbies, stops going out, reduced bathing or personal care
Inertia	Lack of persistence or completion of an activity, does not initiate activities or conversations
Early loss of sympathy or empathy ≥ 1 of:	
Diminished response to other people's needs and feelings	Selfish or hurtful comments or actions, inability to perceive when someone is upset, embarrassed, or in pain, reduced appreciation of sarcasm or sophisticated humour
Diminished social interest, interrelatedness, or personal warmth	Emotionally cold or detached, lack of rapport in conversation, loss of interest or affection in relationships with friends or family members, reduced interest in sex
Early perseverative, stereotyped or compulsive or ritualistic behaviour ≥ 1 of:	
Simple repetitive movements	Repetitive rocking, tapping, clapping or rubbing
Complex compulsive or ritualistic behaviours	Hoarding, strict grooming or walking routines, timekeeping and counting, checking or sorting items, cleaning or tidying, new obsessions or interests (usually spiritual, religious, artistic or musical)
Stereotypy of speech	Habitual repetition of particular words, sentences or topics
Hyperorality and dietary changes ≥ 1 of:	
Altered food preferences	Sweet tooth (sweets, biscuits, ice cream), carbohydrates or obsessive food fads
Binge eating, increased consumption of alcohol or cigarettes	Cramming food into mouth, overeating or messy eating, new addictions to alcohol or smoking
Oral exploration or consumption of inedible objects	Pica
Neuropsychological profile – all 3 of:	
Deficits in executive tasks	Vary as per neuropsychological assessment used
Relative sparing of episodic memory	
Relative sparing of visuospatial skills	

*Criteria of possible bvFTD requires that symptoms be persistent or recurrent, rather than single events. As a guideline, 'early' refers to within 3 years of initial symptom onset as per Rascovsky et al. (2011). bvFTD, behavioural variant frontotemporal dementia. Content adapted from Rascovsky et al. (2011), Warren et al. (2013) and Woollacott and Rohrer (2016).

1.3.2 The Primary Progressive Aphasia Syndromes

The PPA syndromes make up the remainder of the clinical presentations, where language dysfunction is the predominant symptom (Table 1-2). Originally only the semantic and non-fluent variants were described (Mesulam, 2001) with the criteria expanding to later include the logopenic variant (Gorno-Tempini et al., 2011).

SvPPA accounts for around 20% of cases of FTD (Gorno-Tempini et al., 2004, 2011; Rohrer, Rossor, & Warren, 2010) and is an almost entirely sporadic disease. The syndrome is characterised by semantic memory deficits, initially presenting as word-finding difficulties. This insidious loss of semantic knowledge for words, objects and concepts affects both spoken and written language. Initially, this loss of semantic knowledge is for highly specialised and low-frequency words, but progressively evolves to affect more common words and concepts often in a step-wise categorical manner (e.g. degrading from petunia → flower → plant → living thing → thing → I don't know). Despite semantic deficits, patients present with fluent spontaneous speech and initially have intact word and sentence repetition.

By contrast, nfvPPA patients present with profound speech production difficulties alongside relatively preserved semantic knowledge. Their spontaneous speech is agrammatic, effortful, slow and likely to contain phonetic errors. Because initial semantic memory is intact, word comprehension and object knowledge is unaffected and allows for the differentiation from svPPA (Gorno-Tempini et al., 2011; Wilson, Galantucci, Tartaglia, & Gorno-Tempini, 2012).

Unlike svPPA and nfvPPA, which are caused most commonly by TDP-43 and Tau pathology respectively, lvPPA is believed to be an atypical variant of AD as demonstrated by post-mortem (Gorno-Tempini et al., 2011), amyloid PET imaging (Mesulam et al., 2008, 2014) and cerebrospinal fluid (CSF) biomarker (Spinelli et al., 2017) investigations. Clinically, lvPPA patients demonstrate frequent pauses, phonological errors, anomia and impaired sentence but

preserved single-word repetition (Norise et al., 2019; Perini et al., 2018). Single-word repetition and comprehension and object knowledge remain intact due to preserved semantic memory, but they have impaired comprehension of longer sentences due to a deficit in phonological memory (Henry & Gorno-Tempini, 2010). LvPPA patients at a more advanced stage may be difficult to differentiate from nfvPPA due to dysfunctional production of speech; however, the lack of agrammatism, preserved prosody and impaired sentence repetition are key differentiating features (Gorno-Tempini et al., 2011).

As with bvFTD, it has been strongly suggested that where possible, clinical diagnosis of the PPA variants be supplemented with neuroimaging findings to lead to the firmer category of 'imaging-supported diagnosis'.

Table 1-2 Clinical features for the three canonical variants of primary progressive aphasia. Contents adapted from Gorno-Tempini et al. (2004).

Clinical features	svPPA	nfvPPA	lvPPA
Spontaneous speech (fluency; errors; grammar; prosody)	Fluent, garrulous and circumlocutory; semantic errors; intact grammar and prosody	Slow and hesitant, effortful +/- apraxic; phonetic errors; may be agrammatic; aprosodic	Hesitant; not effortful or apraxic; frequent word-finding pauses and loss of train of sentence; intact grammar; intact prosody
Naming	Severe anomia with semantic paraphasias and errors	Moderate anomia with phonetic errors	Mild to moderate anomia with visual and verbal errors
Single-word comprehension	Poor	Intact early on, but affected later on	Intact early on
Sentence comprehension	Initially preserved, later on becomes impaired as word comprehension is impaired	Impaired if grammatically complex	Impaired, especially if long
Single-word repetition	Intact if word comprehended	Mild to moderately impaired if polysyllabic, otherwise intact	Relatively intact
Sentence repetition	Intact if words comprehended	Can be effortful; impaired if grammatically complex	Impaired
Reading	Surface dyslexia	Effortful phonological dyslexia +/- phonetic errors	Can have phonological dyslexia
Writing	Surface dysgraphia	Phonological dysgraphia	Phonological/mixed dysgraphia

Finally, not all patients with progressive primary language difficulties clearly fit into the three canonical PPA diagnoses. These patients present with features from across the spectrum of language dysfunction but do not definitively fulfil the diagnostic criteria listed above. This emerging subgroup was initially termed PPA-mixed disease (PPA-M) and more recently 'PPA-not otherwise specified' (PPA-NOS) (Mesulam et al., 2009).

It has been suggested that over time the evolving syndrome may become clearer and conform to one of the three canonical variants, but this is not always the case (Marshall et al., 2018). Some of these patients have a *GRN* mutation (Harris et al., 2013); however, comprehensive clinical, genetic, pathological or neuroanatomical investigations are lacking in this population due both to the relatively new emergence of this categorisation and the rarity of its presentation or recognition in the clinic.

1.4 Capturing these disease-related symptomatic changes

Diagnosis occurs through formal clinical examination by an experienced clinician and often accompanied by neuropsychological testing to elucidate the nuanced behavioural, language and functional difficulties. However, these assessments are only able to provide a snapshot of the patient at the time of testing and can be confounded by behavioural or language difficulties during testing. To obtain a more comprehensive picture of the first and full set of symptomatic presentations that occur within the context of daily living, informant questionnaires are often used. This is particularly important for behavioural and daily care symptoms of which many patients lack insight, but that family and friends are often acutely aware of. Not only do these questionnaires assist in diagnosing a fuller picture of the syndrome, they have proven valuable in elucidating the cross-sectional neural correlates of many of these symptomatic presentations when paired with structural imaging (Rohrer, Crutch, Warrington, & Warren, 2010).

However, there have been few longitudinal studies of non-imaging symptomatic markers of disease progression in FTD (Long et al., 2019;

Sheelakumari et al., 2019). These small studies have shown that scales currently used to measure clinical and behavioural symptoms such as the Neuropsychiatric Inventory (Cosseddu et al., 2020; Gordon et al., 2010; Knopman et al., 2008) or Frontal Assessment Battery (Cummings, 2007) do not appear to be reliable measures of clinical disease progression.

Whilst questionnaires and neuropsychological tests have proven useful for diagnosis, prognosis and researching neuroanatomical profiles or correlates of symptomatic presentations in FTD, they are yet to prove useful as longitudinal biomarkers and are generally insensitive to detecting early preclinical deficits (Mioshi, Hsieh, Savage, Hornberger, & Hodges, 2010). Because of the familial nature of FTD, there is the unique opportunity to focus on markers for presymptomatic trials. A large body of data demonstrates that structural imaging would be a promising avenue to further explore this as demonstrated by the red curve in the 'Theoretical Biomarker Curve' shown below (Figure 1-2).

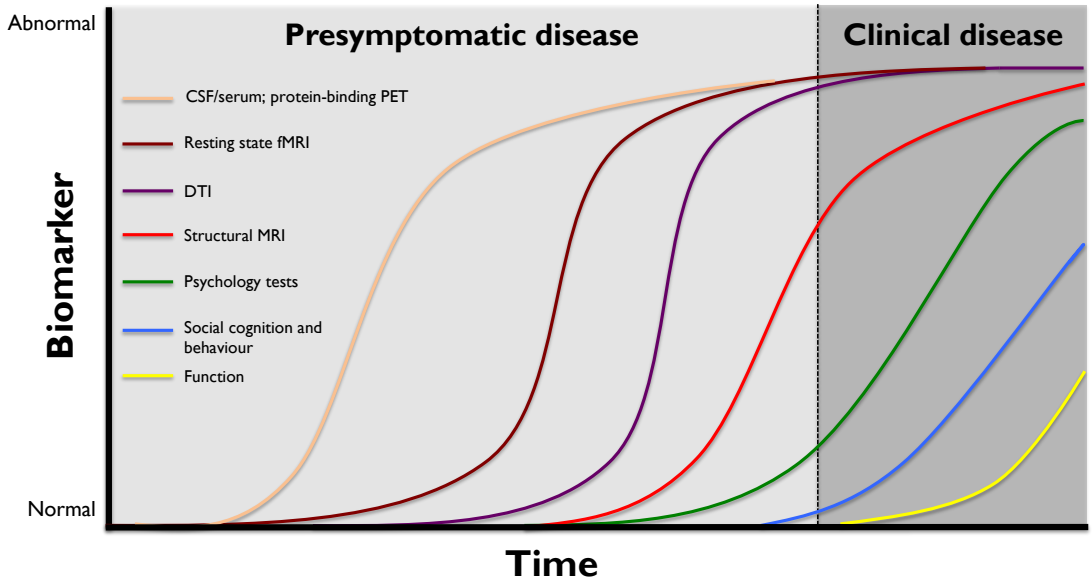

Figure 1-2 Theoretical curve detecting different biomarker changes in FTD.

1.5 Differing distributions: Neuroanatomical signatures in FTD

The recent inclusion of confirmatory neuroimaging into the diagnostic criteria described above demonstrates the added value this modality provides over

and above a symptomatic assessment. Early neuroimaging studies quantifying the pattern and distribution of cerebral loss (often with a focus on grey matter (GM)) from volumetric MRI have shown broadly characteristic patterns of abnormalities across and between FTD subtypes. The majority of studies have aimed at improving differential diagnosis, both distinguishing FTD from other neurodegenerative conditions such as AD, as well as differentiating between the clinical, genetic and pathological subtypes within the FTD spectrum.

1.5.1 Neuroimaging across the clinical syndromes

Volume loss in bvFTD occurs primarily in the frontal and temporal lobes. Several meta-analyses have highlighted the particular involvement of the prefrontal cortex and anterior temporal lobes, along with volume loss in the insula, anterior cingulate, striatum and thalamus (Pan et al., 2012; Schroeter, Raczka, Neumann, & Yves von Cramon, 2007); a pattern that has high sensitivity and specificity in differentiating bvFTD from AD (Jack et al., 2004). An important issue to note here is that this frontotemporal-insular-cingulate involvement was initially demonstrated in a cohort of patients with a range of disease severities. In a condition that is by its very nature neurodegenerative, it is important to distinguish distinct stages of disease-specific volume loss to better understand the initial sites of vulnerability and the evolution of the disease. To better capture this progression or evolution, one group (Seeley et al., 2008) divided their bvFTD cohort into three separate severity stages using the Clinical Dementia Rating scale (CDR) (Morris et al., 1993). The mild group (CDR = 0.5) demonstrated atrophy in the frontal paralimbic cortices. These included reduction in the dorsolateral, rostromedial and orbitofrontal regions, as well as the anterior cingulate, anterior insular, ventral striatum, dorsomedial thalamus and right hippocampus. With increased CDR score and progressing disease severity, regions of volume loss became more extensive in the same regions as the mild group, but also expanded to include more posterior regions in the insular, temporal and parietal lobes (Figure 1-3).

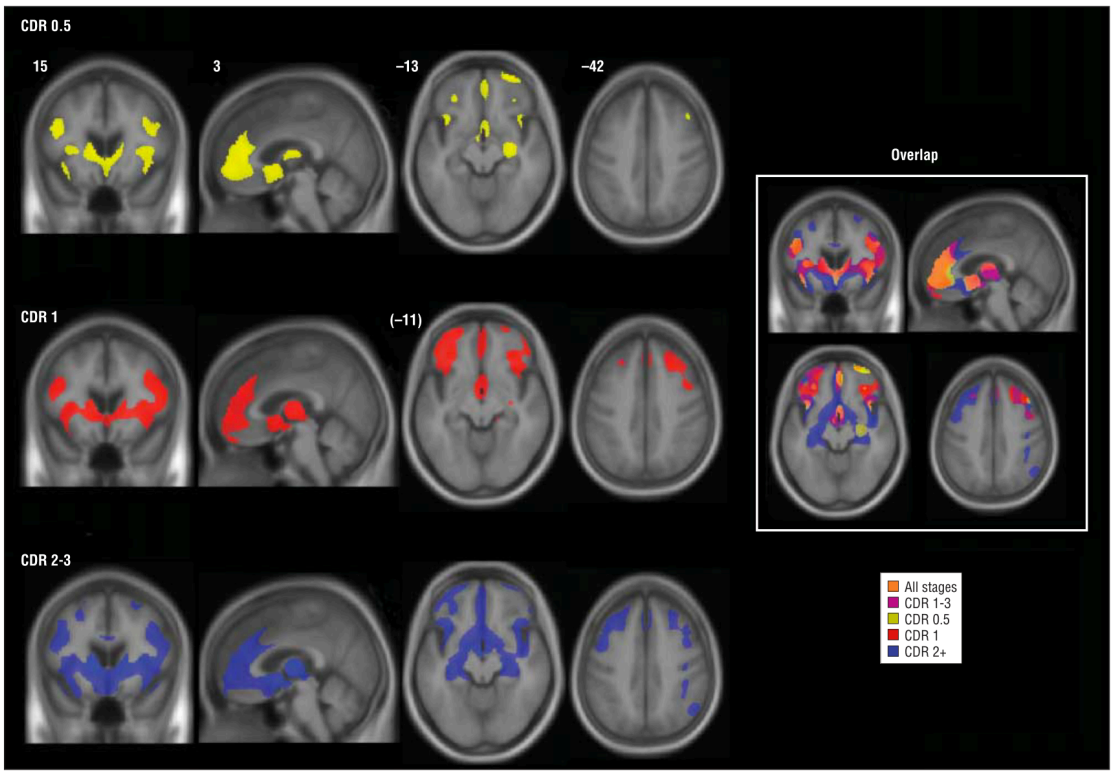

Figure 1-3 Stage-specific grey matter (GM) loss in bvFTD. On the right, the same colour scheme is overlaid on a single image to show the overlap of stage-related atrophy patterns. The right side of the axial and coronal images corresponds to the right side of the brain. Numbers indicate the MNI (Montreal Neurological Institute) co-ordinates of the template brain magnetic resonance image shown. CDR indicates Clinical Dementia Rating; CDR 1–3, mild to severe; CDR 0.5, very mild; CDR 1, mild; and CDR 2+, moderate to severe (reproduced from Seeley et al., 2008).

As the previous meta-analyses suggest, this pattern of frontotemporal-insula-anterior-cingulate involvement is often detected in bvFTD patients and has been proposed to represent a structural and functional network that appears to be particularly vulnerable to neuronal degeneration in these patients.

Despite a consensus about the regional pattern of volume loss at the group level, one important issue is the significant heterogeneity of individual findings in this group, with patients having a variable degree of hemispheric asymmetry, differing predominance of frontal versus temporal lobe atrophy and a varying extent of posterior cortical involvement (Seeley et al., 2008). To address this, two cluster analyses have been performed. These suggest that at least four neuroanatomical subtypes of bvFTD exist: two with predominantly frontal atrophy (a focal frontal-dominant and a more distributed frontotemporal subtype) and two with predominantly temporal involvement (including a focal temporal-dominant and a more widespread frontotemporal-parietal subtype),

which potentially map onto distinct pathogenetic causes of FTD (Figure 1-4). However, an analysis using a subset of these patients was only able to show a clear correlation between the temporal-dominant imaging signature and mutations in the *MAPT* gene (Whitwell et al., 2009, 2012). No other clear relationships could be demonstrated between the pathological profile and neuroanatomical signature, which warrants further investigation in a larger cohort.

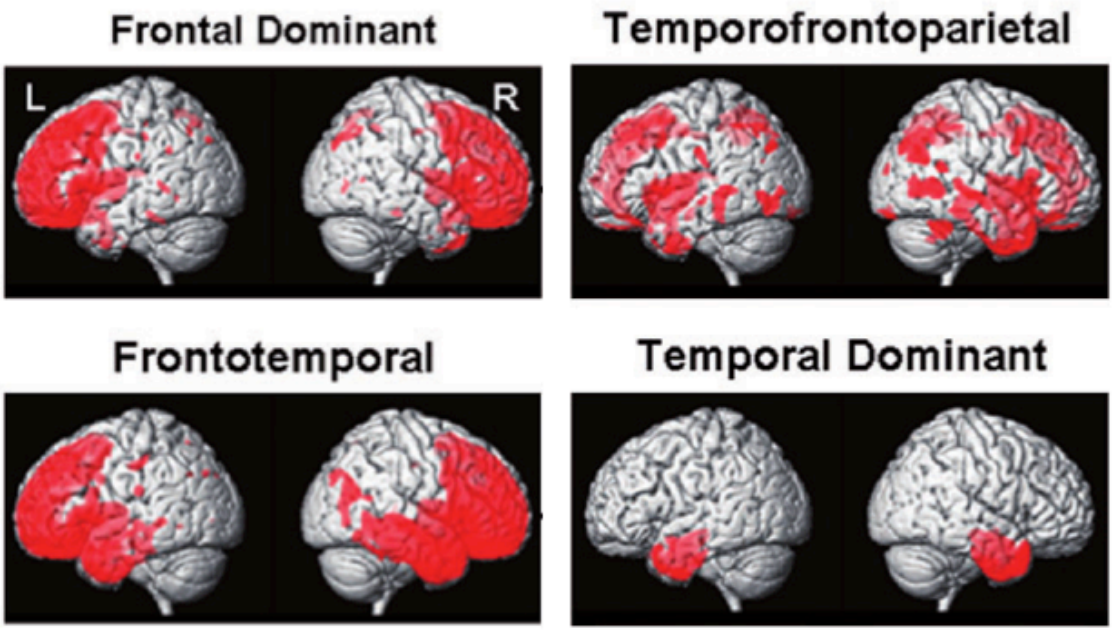

Figure 1-4 Cluster analysis results demonstrating four neuroanatomical variants of bvFTD. Reproduced from Whitwell et al. (2009).

In contrast, svPPA patients have a much more characteristic pattern of asymmetric temporal lobe atrophy, primarily in the anterior and inferior regions (Gorno-Tempini et al., 2004; Rohrer et al., 2009). The most common presentation is prominent left-sided atrophy, although a right-dominant variant has been described (Chan et al., 2009; Josephs et al., 2009). Earliest changes include GM loss in the inferior temporal and fusiform gyri, the temporal pole and the parahippocampal and entorhinal cortex (Brambati et al., 2009; Rogalski et al., 2011; Rohrer et al., 2009). These changes can be striking at presentation with MRI revealing advanced temporal pole atrophy despite the patient's general maintenance of activities of daily living. Involvement of the amygdala and hippocampus has also been reported (Schroeter et al., 2007).

With increasing severity, this pattern extends anteriorly to include orbitofrontal, inferior frontal, insular and anterior cingulate cortices, as well as posteriorly to include temporoparietal regions and corresponding regions in the contralateral hemisphere (Schroeter et al., 2007).

Atrophy profiles in nfvPPA also include predominantly left-hemisphere cortical regions of loss; however, the distribution distinguishes it from svPPA (Schroeter et al., 2007). These include changes in the inferior frontal gyrus (IFG), (particularly the pars opercularis), dorsolateral prefrontal cortex (DLPFC), superior temporal gyrus and insula. Over time, this pattern of cortical loss includes prefrontal and temporal lobe structures in the right hemisphere as well as continued propagation ipsilaterally to encompass anterior frontal, lateral temporal and anterior parietal lobes (Gorno-Tempini et al., 2004; Rogalski et al., 2011), with the involvement of caudate and putamen bilaterally (Gorno-Tempini et al., 2004).

LvPPA patients also present with an asymmetric pattern of atrophy; however, patients generally develop a more posterior profile than svPPA and nfvPPA patients (see Figure 1-5). Volume loss of the temporoparietal, posterior cingulate and precuneus regions is often observed (Gorno-Tempini et al., 2004).

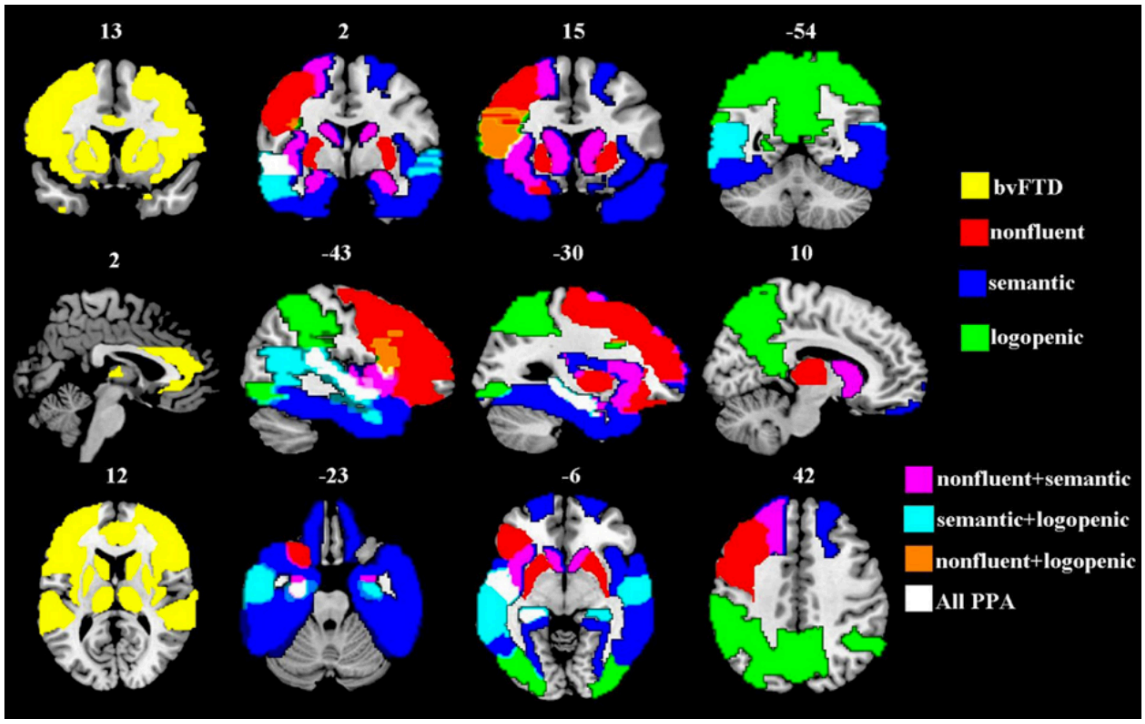

Figure 1-5 Characteristic patterns of grey matter atrophy in the four major clinical FTD subtypes, overlaid on a rendered 3D surface of the Montreal Neurological Institute Template brain. Representative coronal, sagittal and axial slices are displayed in the first, second and third rows, respectively. Images adapted from Agosta et al. (2012).

As discussed above, PPA-NOS is an emerging clinical category with limited neuroanatomical investigations. To date, only a single neuroimaging paper has been published specifically identifying these patients as a subgroup in their analysis investigating thalamic atrophy in FTD (Bocchetta, Gordon et al., 2018). The current project aims to provide some exploratory neuroimaging insights into the pattern of volumetric change with the inclusion of these patients in the FTD cohort results presented in this thesis.

In summary, neuroimaging investigations based on clinical syndromes have proven useful in elucidating broadly distinct and dissociable patterns of atrophy that correspond well with the behavioural and language disturbances that are associated with the syndromes. Despite this, there is both considerable heterogeneity within a syndrome at an individual level as well as some common areas affected across all syndromes, likely reflecting the various distinct but overlapping genetic contributions and pathological processes underpinning the clinical spectrum as a whole. When genetic and pathological information is available, investigations including these levels of classification

have proved highly valuable in extending our understanding of the neuroanatomy affected in FTD.

1.5.2 Neuroimaging in genetic FTD

Patients with *MAPT* mutations present primarily with focal symmetrical anterior temporal and orbitofrontal lobe atrophy (Cash et al., 2018), with involvement of the insula and anterior cingulate also reported (Gorno-Tempini et al., 2004) (see Figure 1-6). Subcortical involvement of the caudate, amygdala and thalamus have also been highlighted (Bocchetta, Gordon, et al., 2018; Bocchetta, Iglesias, et al., 2018; Cash et al., 2018; Gorno-tempini et al., 2004; Mahoney, Beck, et al., 2012; Whitwell et al., 2012). In addition, preliminary findings suggest a more lateral temporal lobe signature for *MAPT* mutations in the coding region, whilst those affecting the splicing of exon 10 target the medial temporal lobes (Ghetti et al., 2015; Spina et al., 2008). In these latter cases, the marked symmetrical loss of hippocampal volume can result in misdiagnosis as early-onset AD (Ghetti et al., 2015; Spina et al., 2008).

C9orf72 mutations show a more distributed symmetric pattern of atrophy, predominantly involving dorsolateral, medial frontal and orbitofrontal lobes, with additional loss in anterior temporal, parietal and occipital lobes, as well as in the thalamus and cerebellum (Cash et al., 2018; Lee et al., 2014; Mahoney, Downey, et al., 2012; Sha et al., 2012). Subcortically, bilateral thalamic atrophy has been particularly emphasised as a feature in *C9orf72* patients (Bocchetta, Gordon, et al., 2018; Cash et al., 2018; Lee et al., 2014).

Unlike *MAPT* or *C9orf72*, *GRN* mutations commonly present with markedly asymmetric atrophy of the temporal, inferior frontal and inferior parietal lobes (see Figure 1-9). These atrophy profiles not only differentiate *MAPT, C9orf72* and *GRN* patients at the group level but have shown some success in classifying the different mutation carriers at the single-subject level (Whitwell et al., 2012).

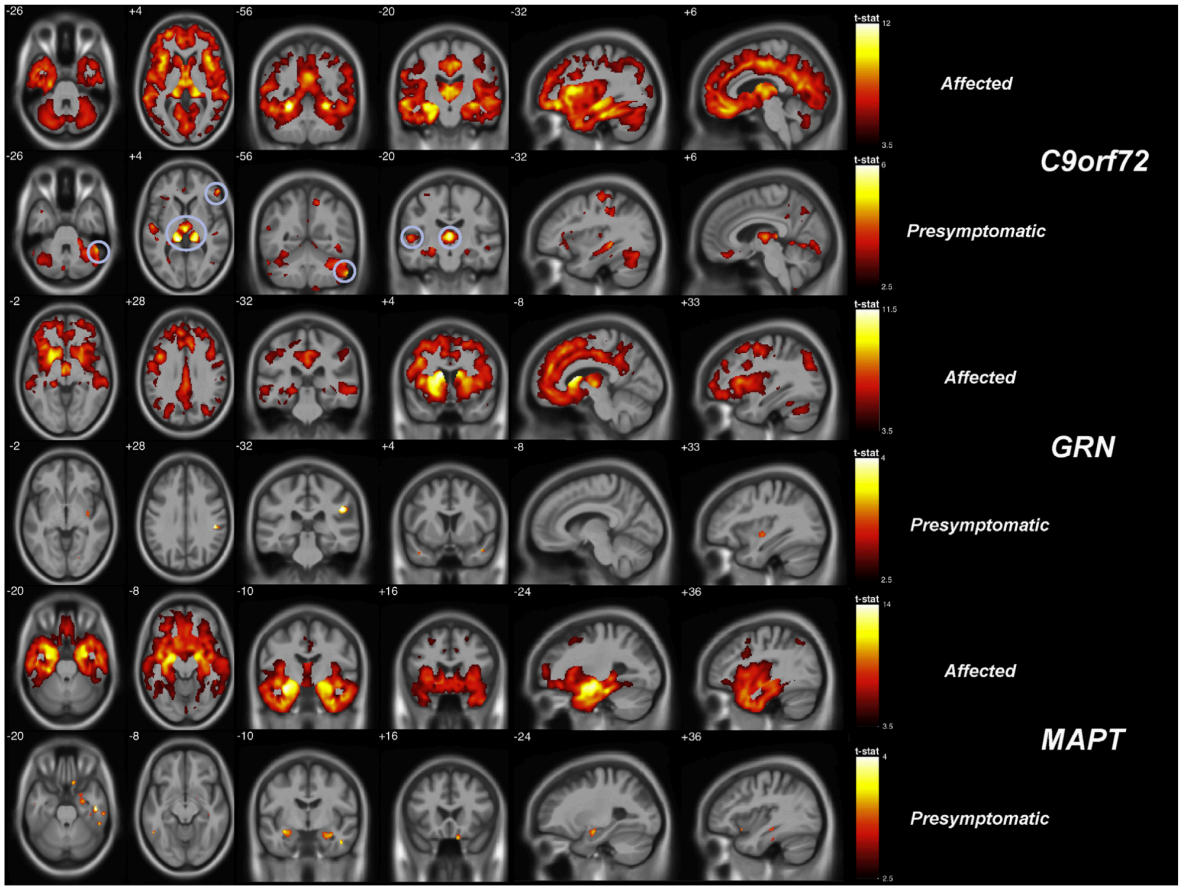

Figure 1-6 GM differences by mutation and clinical status in symptomatic (odd rows, $p < 0.05$ FWE-corrected) and presymptomatic (even rows, $p < 0.001$ uncorrected) carriers compared to noncarriers. Comparisons to the C9orf72 carriers are in the top 2 rows (with findings at $p < 0.05$, FWE-corrected circled in the presymptomatic group), the *GRN* carriers in the middle 2 rows, and *MAPT* carriers in the bottom 2 rows. Reproduced from Cash et al., (2018).

1.5.2.1 Imaging markers of presymptomatic genetic FTD

Because FTD is highly heritable, there is the unique opportunity to investigate the preclinical time window of the disease and to target presymptomatic individuals when underlying pathology is not too advanced. By intervening at this early stage, it may be possible to slow or prevent the downstream cascade of resulting pathology and neuronal damage before it begins to affect individuals in a clinically relevant manner. Sensitive and reliable biomarkers of presymptomatic disease progression to detect potential treatment effects are lacking and urgently needed as we move into the era of early-phase trials in genetic FTD. Neuroimaging markers may prove a viable option over other markers. For example, cognitive or functional markers are less likely to be sensitive enough to identify a change in presymptomatic individuals without

clear deficits and therefore even less likely to detect any modification caused by the drug. In contrast, there is growing evidence from other neurodegenerative diseases such as AD and HD, that changes in imaging biomarkers are measurable several years before symptom onset (Bateman et al., 2012; Tabrizi et al., 2009; Zeun et al., 2019), suggesting their potential for measuring disease-modifying effects for interventions targeting presymptomatic individuals.

Comprehensive analyses of neuroimaging biomarkers have been notably absent in familial FTD, particularly in the presymptomatic phase. However, in recent years there has been increased focus on large-scale international observational studies monitoring genetic FTD families to investigate patterns of presymptomatic change to support the discovery and validation of such biomarkers. These include the Genetic Frontotemporal dementia Initiative (GENFI) (http://genfi.org.uk/) and the Longitudinal Evaluation of Familial Frontotemporal Dementia Subjects, which has recently been combined with the Advancing Research and Treatment for FTLD studies (LEFFTDS/ARTFL). (https://www.rarediseasesnetwork.org/cms/artfl). The GENFI study has demonstrated detection of neuroanatomical changes using structural imaging measures as early as 25 years before expected symptom onset (Rohrer et al., 2015).

Rohrer and colleagues (2015) report an ordered series of neuroanatomical changes across all genetic groups before expected symptom onset (based on mean expected onset within each family). Volumetric analyses revealed the presence of insular atrophy in mutation carriers 10 years before the expected onset of symptoms. This was followed closely by volume loss in the temporal lobe (10 years before expected onset) as well as frontal-lobe and whole-brain atrophy five years before the expected onset of first clinical symptoms (Figure 1-7). Subcortical involvement was also reported, with the striatum, thalamus, hippocampus and amygdala all demonstrating significant volume reduction in mutation carriers compared with gene-negative family members (Figure 1-8).

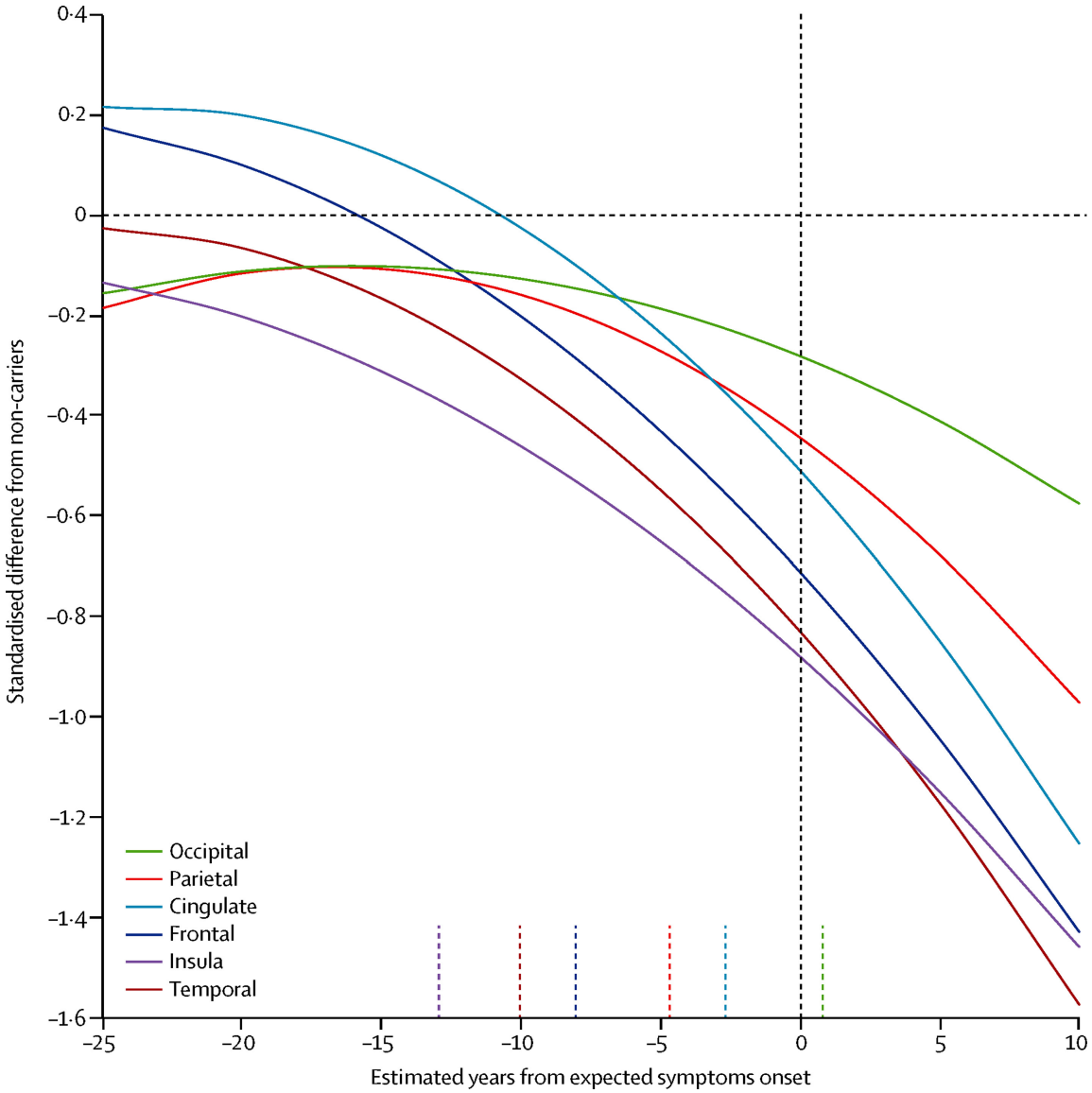

Figure 1-7 Standardized difference between all mutation carriers and non-carriers in cortical grey matter volumetric imaging measures, plotted against estimated years from expected onset of symptoms. Individual data points not plotted to prevent disclosure of genetic status. The time at which the upper 95% confidence interval for each curve crosses zero on the y-axis (i.e., the point at which a significant difference exists between mutation carriers and non-carriers) is shown on the x-axis. Reproduced with permission from Rohrer et al. (2015).

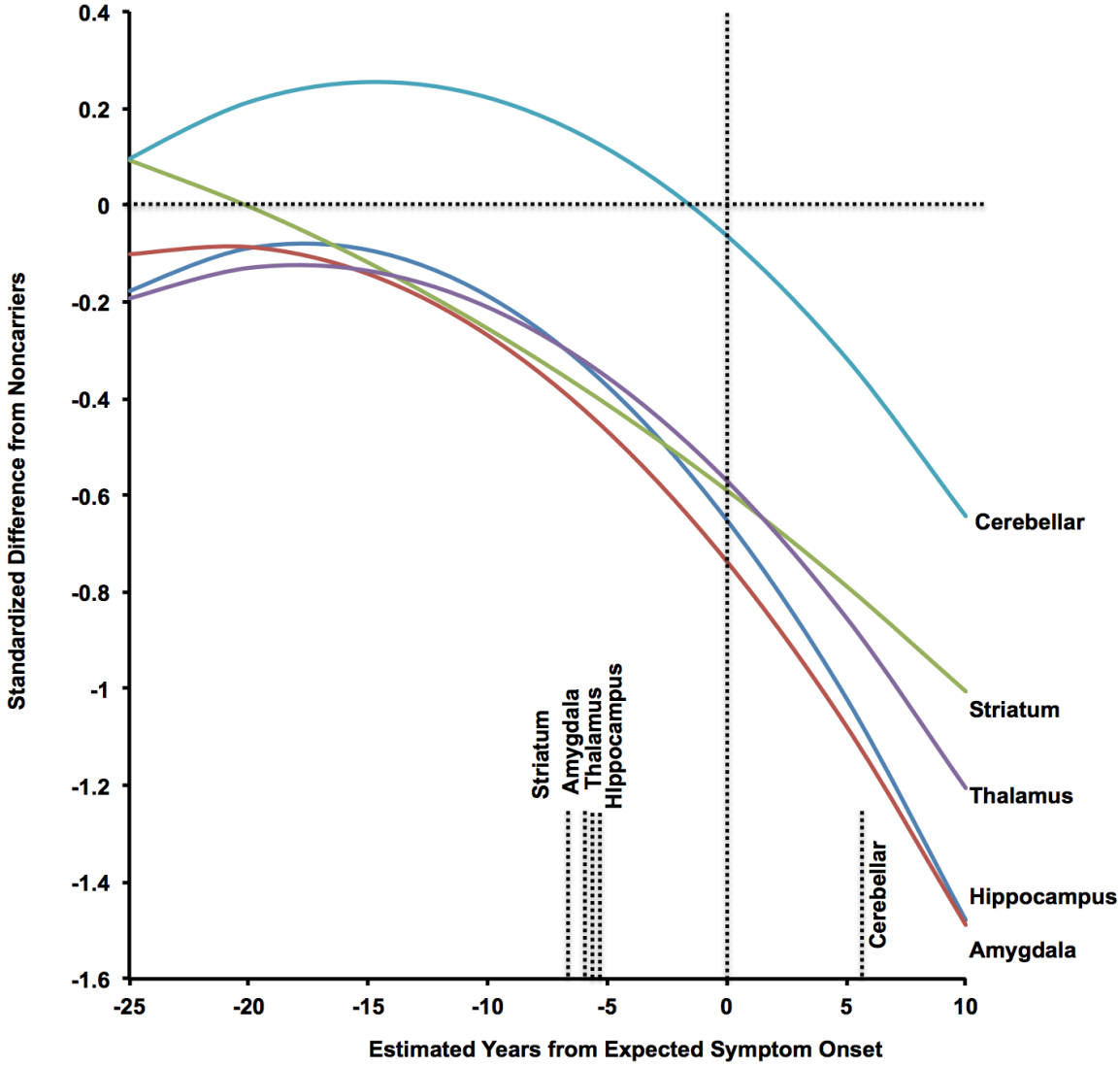

Figure 1-8 Standardized difference between all mutation carriers and non-carriers in subcortical and cerebellar volumetric imaging measures versus estimated years from expected onset of symptoms. Individual data points not plotted to prevent disclosure of genetic status. The time at which the upper 95% confidence interval for each curve crosses zero on the y-axis (i.e., the point at which a significant difference exists between mutation carriers and non-carriers) is shown on the x-axis.

In addition, they found specific patterns of loss in the different mutation types. In line with neuroanatomical loss observed in symptomatic patients, *MAPT* mutation carriers exhibited atrophy in the hippocampus and amygdala (15 years before expected onset), followed by atrophy of the temporal lobe (10 years) and insula (5 years) (Rohrer et al., 2015). *C9orf72* expansion carriers had very early involvement of the thalamus as well as insular and posterior cortical atrophy (25 years before expected onset). Presymptomatic *GRN* mutation carriers also showed early atrophy involving the insular (15 years before expected onset), with temporal and parietal lobe loss evident 10 years before symptom onset. These early differences are likely to explain some of the heterogeneity we see later on in the disease course when symptoms become evident in FTD patients.

1.5.3 Imaging markers of pathology-confirmed FTD

With current and upcoming therapeutic interventions focused on targeting the abnormal proteins underpinning the disease process in FTD, a key challenge is finding neuroanatomical signatures that can accurately predict molecular pathology non-invasively during life. Recent advances in classifying the subtypes of these FTD pathologies have contributed greatly to this goal. In the tauopathies, Pick's disease can be associated with asymmetric atrophy in the frontal cortex, (including orbitofrontal, medial and dorsolateral regions primarily), with involvement of the anterior temporal lobes and insula (Rohrer et al., 2011, 2015). *PSP* is associated classically with midbrain atrophy (Massey et al., 2013), however, despite being relatively specific this is not a sensitive marker. Moreover, in patients with cognitive presentations, *PSP* and *CBD* tauopathies are associated with more posterior frontal cortical atrophy, targeting the supplementary and pre-motor regions, with *CBD* showing a more distributed pattern including the parietal lobes and striatum (Josephs et al., 2008; Whitwell, Jack, Boeve et al., 2010).

TDP-43 proteinopathies can be divided into four distinct subtypes (*Types A–D*), based on the morphology of the inclusions (Seelaar, Rohrer, Pijnenburg, Fox, & van Swieten, 2011). Patients with *TDP-A* pathology (commonly

associated with *GRN* mutations) present with widespread asymmetric frontotemporo-parietal atrophy (Rohrer, Geser et al., 2010; Whitwell, Jack, Parisi et al., 2010), with anterior cingulate and caudate involvement (Rohrer et al., 2011). *TDP-B* patients present with symmetrical frontal lobe as well as insula and anteromedial temporal lobe atrophy, whilst in contrast, *TDP-C* patients show asymmetrical loss in the anteroinferior temporal lobe, consistent with the predominant clinical diagnosis of svPPA (Rohrer & Rosen, 2013). *TDP-D* is by far the rarest variant and associated with mutations in the valosin-containing protein *(VCP)* gene. Preliminary MRI findings have been inconsistent and failed to demonstrate a clear pattern (Kim et al., 2011; Stojkovic et al., 2009; Surampalli et al., 2015). Although rare, investigations into *FUS* pathology cases reveal a pattern of atrophy affecting the orbitofrontal lobe, anteromedial temporal lobe, anterior cingulate, insula and particularly striking caudate atrophy (Figure 1-9) (Josephs et al., 2010).

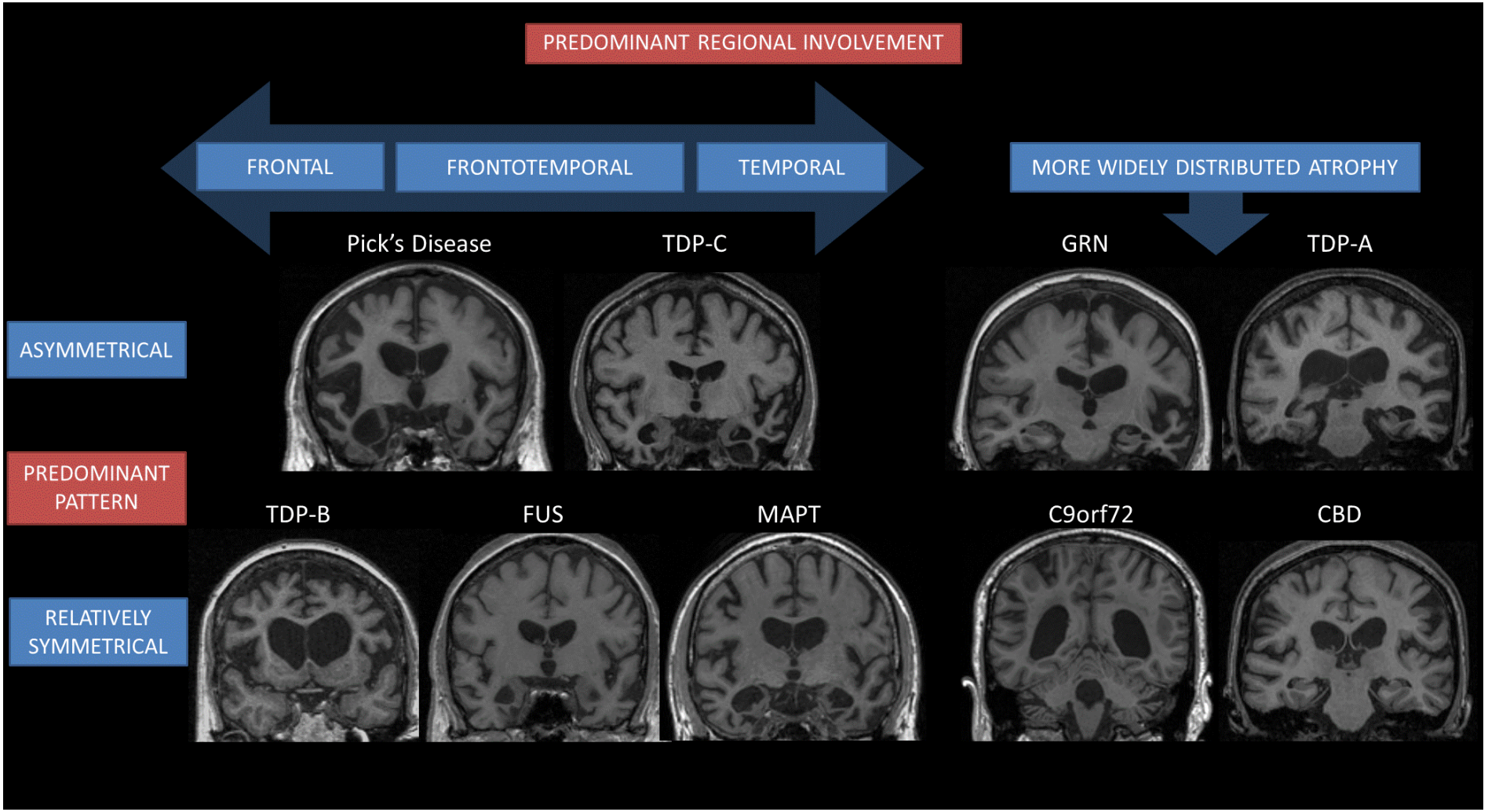

Figure 1-9 Pathological and genetic patterns of region and symmetry of atrophy of FTD patients in the cohort used for this thesis. Adapted from Gordon, Rohrer, & Fox (2016).

In summary, previous studies employing structural imaging have shown promise in capturing informative neuroanatomical signatures in the clinical, genetic (both presymptomatic and symptomatic mutation carriers) and pathological subgroups of FTD. The profiles commonly correspond well with key features of the clinical presentation and despite overlap provide useful tools for differentiating subtypes at the group and in some cases, individual level. However, inconsistencies do exist in the literature and further clarification is required. A major hurdle to consistency of the results is probably partly due to the small study cohorts used. Given the extensive variability included in the FTD spectrum, these small cohorts are likely to be capturing only a subset of this varied patient population as well as including varying levels of disease severity. This makes generalisability of these previous findings problematic and will need further investigation in large well-characterised cohorts.

1.6 The importance of longitudinal markers in FTD: moving towards intervention trials

Another key caveat of previous investigations is that they have been primarily cross-sectional. Despite the clear benefits of longitudinal study design, relatively few studies have investigated how these neuroanatomical profiles change over time and even fewer have reported rates of disease progression across the FTD spectrum. Investigations characterising longitudinal trajectories of change in FTD are increasingly important as we move into therapeutic trials. This is not only to better characterise the evolution of the disease, but also importantly to provide surrogate markers for evaluating the efficacy of treatment interventions and to help inform trial design so that they are statistically powered to detect a clinically relevant change in these markers.

Methods for quantifying brain atrophy rates from serial MRI are already currently employed as biomarker surrogate endpoints in other neurodegenerative conditions including multiple trials in AD (Cash et al., 2014) and more recently HD (Tabrizi et al., 2019). Therefore, validation within the different FTD subgroups is important to ascertain their utility in this patient

population. Several studies have investigated rates of whole-brain loss and ventricular expansion across the clinical syndromes (Chan et al., 2001; Gordon et al., 2010; Knopman et al., 2009; Rohrer et al., 2008, 2012). All have demonstrated significantly increased annual rates of change compared with healthy age-matched controls (Gordon et al., 2010; Knopman et al., 2009; Rohrer et al., 2008, 2012). Reported rates of annualised volume loss substantially vary across studies (Gordon et al., 2016; see Table 3-1) and will need further validation and harmonisation in large well-characterised cohorts. Despite this, there is broad agreement that the PPA patients generally demonstrate more homogeneous rates of global change than the greater heterogeneity evident in bvFTD cohorts, generally in the range of 2–3% annual volume loss (percent of baseline volume). Despite this global consistency in PPA, differences in hemispheric and lobar changes exist. One study suggests that left-hemisphere rates are greater than right hemisphere rates for both svPPA and nfvPPA, with the left temporal lobe progressing fastest in svPPA and left frontal lobe fastest in nfvPPA (Rohrer et al., 2012).

Investigations involving patients with a known genetic mutation show a similar range of global volumetric changes but suggest that *MAPT* mutations are associated with a mean annual rate of ~2.4%, intermediate between those with *GRN* mutations, who exhibit the fastest rate of loss at ~3.5% and *C9orf72* whose volumetric rates have been reported as the lowest at ~1.4% and more in line with AD patients (Mahoney et al., 2012; Spina et al., 2008; Whitwell, Weigand et al., 2011), although there is variability within each group. There is also some evidence for an acceleration in global rates of change in patients with non-tau pathology as disease severity increases (Whitwell et al., 2008); but extensive longitudinal investigations of patients with pathological confirmation is currently lacking.

Distinct neuroanatomical profiles from cross-sectional investigations indicate that improved sensitivity may be achieved with regional analysis over global measures to capture the focal changes we might expect in FTD. However, there are currently very few studies assessing regional longitudinal rates of

change and their utility in the tracking of pathogenic change. One study applying an automated atlas-based segmentation method on serial MRI demonstrated that increased annual atrophy in lateral orbitofrontal grey matter regions clearly differentiated bvFTD from both healthy controls and AD patients, whilst annual reductions in temporal lobe white matter structures were also important in classification of bvFTD subjects (Frings et al., 2014). More recently, investigations of frontal and temporal lobe volume loss have shown significant annualised reductions in bvFTD, svPPA and nfvPPA patients compared with healthy controls, resulting in feasible sample size calculations across all subtypes (Staffaroni et al., 2019). Cortical thickness investigations have also implicated the orbitofrontal gyrus, with significantly faster rates of regional thinning in bvFTD than AD and controls (Staffaroni et al., 2019). Investigations in svPPA suggest improved sensitivity when measuring disease progression with temporal lobar volume change instead of hemispheric or whole-brain measurements (Frings et al., 2014). In addition, a study of progressive apraxia of speech patients (who almost exclusively present with tau pathology at post-mortem), demonstrated that annual rates of atrophy in the precentral and supplementary motor areas were greater than other regional and more global MRI measures (Whitwell et al., 2015).

The utility of these rates of change markers (based on the estimated sample size required to demonstrate a relevant treatment effect) varies across imaging methods and FTD subtypes. Sample sizes are generally smaller in the PPA syndromes due to the heterogeneity of bvFTD and resulting variability of change measures across the subgroup. This highlights the importance of improved stratification in trials and increasingly targeted regional measures of longitudinal change in these focal disease populations. Future confirmation of rates of change in comprehensive well-characterised cohorts is required to harmonise these longitudinal markers. However, it is promising to note that despite variability, many measures do produce feasible samples sizes of 50–100 patients per treatment arm.

1.7 Measuring neuroanatomical volumes using structural MRI

There are multiple methods of extracting regional brain volume information as well as measuring changes in these volumes over time.

1.7.1 Manual segmentation techniques

Manual segmentation by a highly trained individual is considered the gold standard. However, the labour intensive and time-consuming nature of this method limits its feasibility for application in large cohorts or extensive multicentre trials. Inter- and intra-rater reliability also becomes an issue, as does the use of differing training and segmentation protocols. To address this, there has been considerable attention given to developing automated pipelines for the accurate and consistent delineation and extraction of global and regional segmentations.

1.7.2 Automated segmentation techniques

Currently, there are multiple automated or semi-automated segmentation techniques available that label neuroanatomical structures from T1-weighted images. Whilst there are different approaches and pipelines, they generally all involve most or all of the following: brain extraction (identifying brain from non-brain voxels on the scan); tissue classification (identifying grey matter (GM)/white matter (WM) and CSF); intensity correction (to account for inhomogeneity within these images); spatial normalisation or registration (to match to a template or atlas) and finally labelling of regions of interest onto the unsegmented target image. Atlas-based approaches are now some of the most widely used segmentation techniques in clinical image analysis, including Freesurfer (Fischl et al., 2002), FAST (FMRIB's Automated Segmentation Tool) (Zhang, Brady & Smith, 2001) and Geodesic Information Flows (GIF) (Cardoso et al., 2015). These utilise a manually segmented atlas (or more commonly multiple atlases) as the basis for propagating the new labels for segmentation. This is accomplished through image registrations that non-linearly warp the manual label (or labels) to best fit the particular morphology of the new target image. Many

of these atlas-based approaches have demonstrated comparable accuracy to manual methods (Iglesias & Sabuncu, 2015); however, the atlas selection, the registration technique and the protocol for the manual segmentations are important factors in the segmentation performance. The move towards a multiple atlases approach that include a diverse selection of pathological exemplars is important to improve in capturing the anatomical variation likely across the healthy to diseased population spectrum.

1.8 Measuring neuroanatomical volume change longitudinally

Indirect measures

Volumetric change can be derived indirectly by subtracting the volume of the follow-up segmentation from the baseline segmentation. It is important to note that this is highly susceptible to segmentation error as it considers the two images as independent of each other and slight differences in segmentation accuracy may mask or bias the actual underlying physiological change.

Direct measures

A more robust approach for measuring longitudinal volumetric change is to incorporate directly the properties from the image pairs themselves; thus avoiding sole reliance on segmentation accuracy. A number of these will be discussed in Chapter 2.

1.9 Structures often investigated in FTD

Longitudinally, whole-brain, ventricular, frontal and temporal lobe measures of change have been the focus of investigations employing structural T1 imaging (Gordon et al., 2016). Subcortically, several studies have also investigated hippocampal atrophy in bvFTD and svPPA (Chan et al., 2001; Gordon et al., 2010; Knopman et al., 2009; Rohrer et al., 2008, 2012; Staffaroni et al., 2019). Cross-sectionally, involvement of the insula and cingulate cortex is increasingly being explored; as have subcortical regions including the

hippocampus (Barnes et al., 2006; Schroeter et al., 2007), amygdala (Garibotto et al., 2011), caudate, thalamus, hypothalamus (Barnes et al., 2006; Bocchetta, Gordon, et al., 2018; Bocchetta et al., 2015; Landin-Romero et al., 2017; Schroeter et al., 2007), striatum and the basal ganglia (Halabi et al., 2013; Macfarlane et al., 2015). As the pattern of regional involvement varies considerably between clinical, genetic and pathological subtypes there is currently a lack of consensus about which regions might be most informative across the spectrum and within the individual subgroups. With limited longitudinal investigations of more focal cortical and subcortical regions, it will be crucial to expand on cross-sectional work to investigate their potential as disease progression identifiers.

1.10 Knowledge gaps and study aims.

In a disease as complex and varied as FTD, there are several challenges and knowledge gaps I aim to address in this thesis.

First, as a rare condition, findings have been historically based on relatively small cohorts, contributing to the lack of consistent results.

Aim 1: To collate a large, high-quality and well-characterised longitudinal imaging cohort encompassing all key clinical, genetic and pathological subtypes within the FTD spectrum (Chapter 2)

Second, automated segmentation techniques show great promise for measuring disease progression and may provide an exciting avenue to detect disease-related treatment effects. However; it remains untested and unclear as to which of the many available methodologies should be used within and across the FTD spectrum.

Aim 2: To apply and compare widely used whole-brain segmentation techniques and assess their utility as longitudinal biomarkers across the FTD spectrum (Chapter 3)

Third, cross-sectional work suggests that more focal measures could yield promising longitudinal results. However, there continues to be a lack of consensus as to which other neuroanatomical structures might provide the clearest measure of disease progression across the FTD spectrum.

Aim 3: Comprehensively investigate lobar (Chapter 4) and subcortical (Chapter 5) automated measures of longitudinal change across the full FTD spectrum.

Finally, as a highly heritable disease, FTD has great potential as a disease model for presymptomatic interventions as well as providing information about pathological patterns of change *in vivo*. With limited longitudinal investigations, it is unclear whether measures that show promise in symptomatic sporadic patients might also be applicable to genetic and presymptomatic or early intervention cohorts.

Aim 4: To apply global and regional measures shown to perform well in the preceding chapters, to a large genetic cohort to investigate their utility presymptomatically and further investigate differences in the neuroanatomical profile of the key genetic mutations (Chapter 6).

1.11 Key hypotheses of the project

Hypothesis 1: Cross-sectionally, widely used fully automated whole-brain segmentation pipelines will perform well at delineating brain from non-brain regions across the varied FTD spectrum using structural imaging as evaluated by comprehensive visual assessment of segmentation accuracy.

Hypothesis 2: Longitudinally, indirect volume-difference measures derived from these automated segmentations will reveal the impact of segmentation errors by producing significantly different effect size and sample size

calculations based on the mean and standard deviations of the annual rate of global atrophy these methods produce in head-to-head paired comparisons.

Hypothesis 3: Direct methods of measuring whole-brain atrophy will provide more robust biomarkers of longitudinal change across the FTD spectrum than the indirect volume-difference measures, resulting in significantly lower sample sizes required to detect a putative treatment effect.

Hypothesis 4: Adaptations of these direct global longitudinal measures will yield low sample size estimates below 100 participants when applied to more focal regional volumetric measures relevant to FTD (i.e. frontal, temporal, insular, cingulate and subcortical regions such as the thalamus, caudate) providing additional valuable biomarker options for upcoming therapeutic interventions.

Hypothesis 5: Due to different neuroanatomical presentations across the FTD spectrum, the most robust regional measures for detecting volumetric change and resulting lowest sample sizes will differ between the patient subgroups reflecting the different disease processes underlying these presentations. Specific predictions are listed in Chapter 4.1 and 5.1 for the lobar and subcortical structures respectively.

Hypothesis 6: Measures of caudate, thalamus and hippocampal atrophy, which showed the most promise as biomarkers in the FTD symptomatic subgroups, will demonstrate potential for detecting structural change in the preclinical stage, by producing significantly higher longitudinal rates of atrophy in presymptomatic mutation carriers compared with non-carrier controls.

2 Methods

2.1 Introduction

One of the major factors hindering the consistency of FTD neuroimaging findings lies in the limited size and thus reduced representation of the full patient population in previous study cohorts. Given the heterogeneity across the spectrum, large comprehensive cohorts are required to more accurately capture a fuller picture of FTD across its many subgroups and presentations. In addition, image analysis techniques and statistical methodology vary between investigations making clear comparisons between studies difficult. Therefore, a key goal of the current project was to collate a substantial and well-characterised cohort of FTD patients that would be likely to more closely reflect the varied patient population of interest. In addition, all imaging analyses would be applied uniformly to the full spectrum for comparative investigations.

2.2 Collation and characterisation of FTD study cohorts

2.2.1 Retrospective FTD neuroimaging cohort

2.2.1.1 Candidate selection and image review

Participants consisted of individuals who had been recruited through the Cognitive Disorders Clinic at the National Hospital of Neurology and Neurosurgery (NHNN), Queen Square, London, UK. Each had consented to take part in research being conducted at the Dementia Research Centre (DRC, London). Controls were almost exclusively partners, spouses or family/carers of patients involved in this research. The DRC has been conducting volumetric neuroimaging for clinical and research purposes for almost 30 years. Therefore, the initial focus of my work was to utilise this valuable historical resource to locate and collate all available imaging data into a central neuroimaging repository. The key stages involved in creating this retrospective FTD Neuroimaging Cohort are summarised in Figure 2-1.

Candidate identification was the first task. To do this, I reviewed all departmental research publications and PhD theses for study cohort identification. From the identified cohorts, I reviewed all data about the included participants to identify potential candidates, including clinical, genetic and imaging databases. The inclusion criteria were that the individual had a diagnosis of any FTD subgroup, a frontal-lobe syndrome or was denoted as a healthy control at the time of testing. All relevant clinical and research data for these subjects was collated and then reviewed by an experienced neurologist (JR) to confirm the diagnosis. The confirmed candidate list of 748 individuals was the starting point for identifying available neuroimaging data. From this list, I performed a comprehensive search of the different DRC imaging repositories and the PACS (Picture archiving and communication system) imaging database for Queen Square and NHNN scanners. This resulted in 550 individuals who had undergone a clinical or research MR scan. All available T1-weighted and T2-weighted sequences were collated centrally for review.

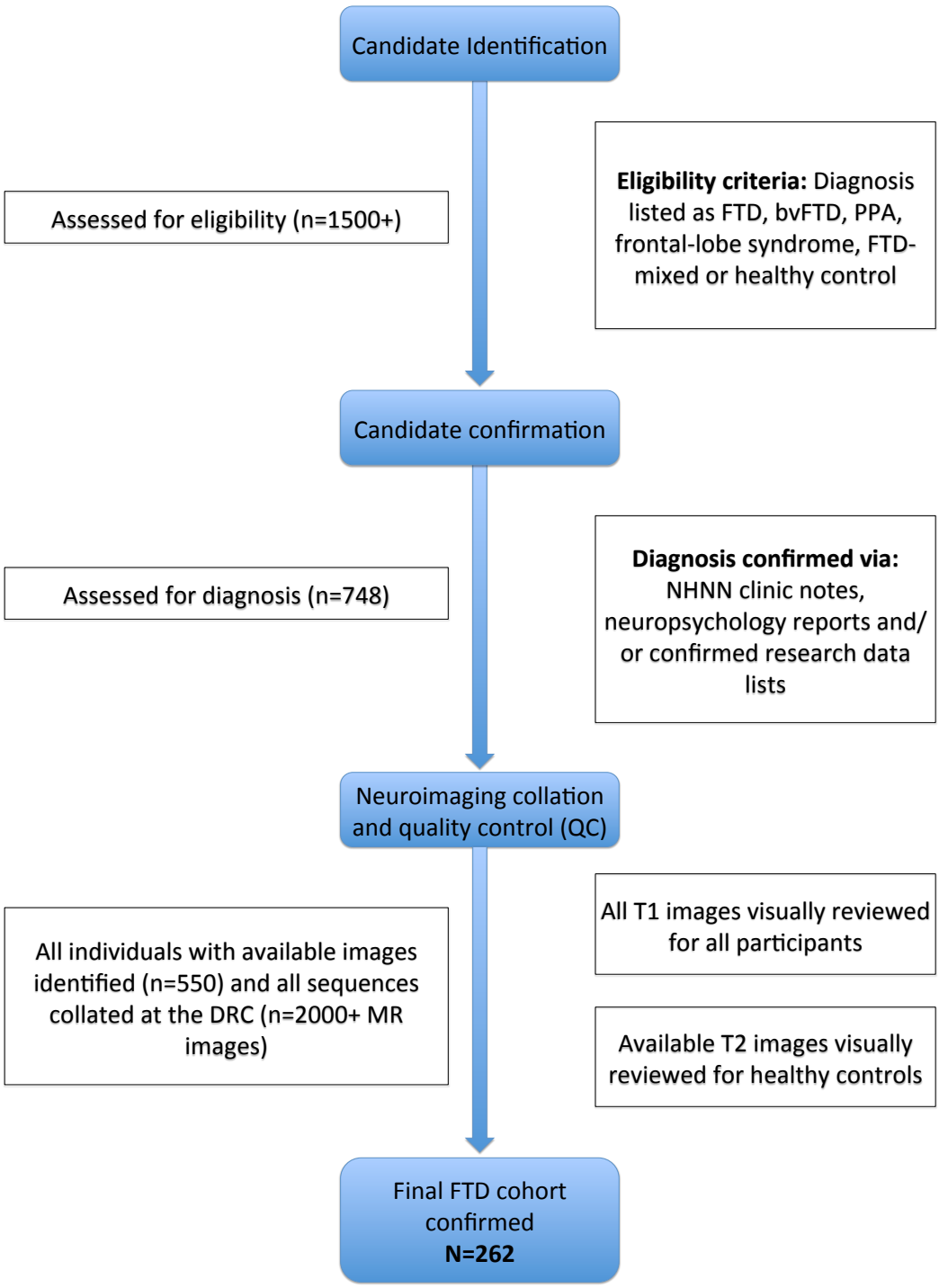

Figure 2-1 Main phases involved in the collation and review of the retrospective FTD neuroimaging cohort the collation and review of the retrospective FTD neuroimaging cohort. Details of the quality control (QC) process given in section 2.3.2.1. Details of the final 262 cohort participants given in **Table 2-1**.

Many of the research participants had been involved in projects at the DRC for several years and undergone multiple imaging sessions across a number different scanners. For these individuals, all available imaging was reviewed simultaneously to identify the same-scanner pair of images that were best matched on quality, positioning and intensity (see Figure 2-2). As new scans were acquired, they were reviewed to determine if they were more suitable for the final database.

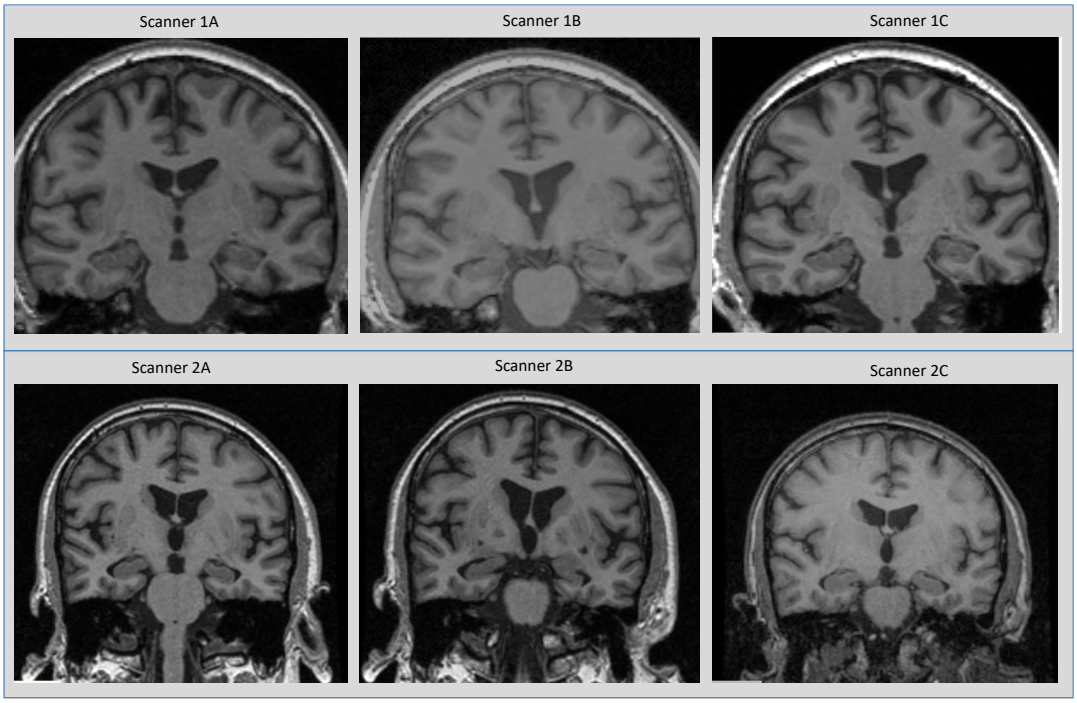

Figure 2-2 Six imaging visits for a single participant performed across two scanners. The initial pair chosen were Scanner 1A and Scanner 1C. Once the participant had acquired an equally good quality matched pair on the 3T Scanner 2, the final pair included in the database was updated to Scanner 2A and Scanner 2B.

The final FTD longitudinal neuroimaging cohort consisted of 262 individuals: 184 FTD patients and 78 healthy older controls. The FTD patients included 66 bvFTD, 45 svPPA, 45 nfvPPA, 7 PPA-NOS and 21 individuals with a diagnosis of lvPPA. Of these 184 clinically diagnosed participants, 34 had a positive genetic result for *MAPT*, *C9orf72* or *GRN* mutations. In addition, several patients had also come to post-mortem and their underlying pathology confirmed by the Queen Square Brain Bank for Neurological Diseases (QSBB), UCL. The pathology subgroup consists of a combination of these individuals and the gene-positive patients. *MAPT* patients are included in the

Tauopathy (Tau) subgroup, whilst the *C9orf72* and *GRN* patients are included in the TDP-43opathy (TDP-43) subgroup informed by known pathogenesis of these mutations. One bvFTD individual also had FUS pathology confirmation at post-mortem.

Baseline demographics are summarised in Table 2-1. These baseline variables were compared using a Kruskal-Wallis test for the continuous variables and a Fisher's exact test for the categorical variables of gender and scanner type. The following tests were run. Controls were compared with the clinical groups, the genetic groups and pathology-confirmed groups in three separate tests. Within patient subgroup comparison were also performed for these three groups. Comparisons between the patient subgroups were not performed as it negated the key underlying assumption of independence of observations given patients' overlapping membership across the clinical, genetic and pathology groups.

Table 2-1 Retrospective FTD cohort demographic summary reporting mean (standard deviation) unless otherwise stated. Patients are split into groups based on their clinical diagnosis, genetic mutation status and those with known Tau or TDP-43 pathology (either pathological confirmation and/or genetic mutation status). Cells in bold were revealed to be significantly different at the $p < 0.05$ using the Kruskal-Wallis test for continuous variable and Fisher's exact test for the categorical variables.

	Healthy controls	**Clinical FTD subgroups**					**Genetic FTD subgroups**			**Pathology FTD subgroups**	
		bvFTD	**svPPA**	**nfvPPA**	**PPA-NOS**	**lvPPA**	***MAPT***	***C9orf72***	***GRN***	**Tau**	**TDP-43**
Number of participants	78	66	45	45	7	21	16	10	8	19	34
Clinical, genetic and pathology mapping	NA	15 *MAPT*; 9 *C9orf72*; 4 *GRN*; 1 Tau; 2 TDP-43 Type A	2 Tau; 10 TDP-43 Type C	1 *C9orf72*; 2 *GRN*; 2 TDP-43 Type A	2 *GRN*	5 AD; 1 *MAPT*	15 bvFTD; 1 lvPPA	9 bvFTD; 1 nfvPPA	4 bvFTD; 2 nfvPPA; 2 PPA-NOS	16 *MAPT*; 3 post-mortem confirmed	10 *C9orf72*; 8 *GRN*; 14 post-mortem confirmed *
Male/Female	34 / 44	**52 / 14**	24 / 21	26 / 19	4 / 3	**15 / 6**	9 / 7	**9 / 1**	3 / 5	12 / 7	21 / 13
% male	44%	**75%**	53%	58%	57%	**71%**	56%	**90%**	38%	63%	49%
Age at baseline	62.3 (11.6)	**62.6 (8.6)**	63.7 (7.8)	65.6 (7.1)	62.6 (7.0)	67.4 (7.4)	**55.4 (7.2)**	62.5 (6.0)	61.2 (7.2)	**56.2 (7.2)**	62.5 (6.4)
Age at symptom onset (years)	NA	56.7 (8.8)	59.0 (7.5)	61.2 (6.8)	60.1 (7.5)	63.3 (7.1)	49.1 (5.8)	55.6 (8.6)	58.0 (6.4)	50.7 (7.0)	57.7 (7.3)
Disease duration (years)	NA	**5.9 (3.9)**	4.7 (1.9)	4.4 (1.9)	**2.4 (1.4)**	4.1 (1.9)	6.3 (3.9)	6.9 (4.4)	3.2 (3.4)	5.6 (3.7)	4.8 (3.4)
Scan interval (years)	1.5 (0.8)	1.4 (0.8)	1.5 (0.8)	1.3 (0.6)	1.6 (0.7)	1.3 (0.7)	1.9 (1.2)	1.0 (0.4)	1.2 (0.4)	1.8 (1.2)	1.5 (0.9)
1.5T / 3.T image pairs	40 / 38	34/ 32	29 / 16	23 / 22	2 / 5	**4 / 17**	6 / 10	4 / 6	5 / 3	10 / 9	21 / 13
Scan acquisition dates	1992 - 2018	1992 - 2017	1993 - 2018	1995 - 2017	2005 - 2015	2005 - 2018	1992 - 2014	1996 - 2015	1995 - 2017	1992 - 2015	1993 - 2017

* Two rarer genetic mutations (TBK1 and SQSTM1) with known TDP-43 associated pathology also included here.

Baseline demographic data reveal several significant differences between the healthy controls and FTD subgroups, as well as between each of the subgroups with the separate clinical, genetic and pathology groups (bolded data in Table 2-1).

Fisher's exact test revealed significant differences in gender distribution for the clinical subgroups (p = 0.001) and genetic subgroups (p = 0.034) when compared with controls. This result was driven by significantly more men in the bvFTD, lvPPA and *C9orf72* subgroups. There were also differences between the clinical subgroups (p = 0.035) with the proportion of males in the bvFTD subgroups higher than in the svPPA and nfvPPA groups. In addition, there were more men in the *C9orf72* than *GRN* subgroups revealed by the genetic subgroup comparison (p = 0.012).

Kruskall-Wallis test revealed significant differences in age at baseline assessment comparing the control group with the genetic (χ^2 (3) = 10.492, p = 0.015) and with the pathology-confirmed subgroups (χ^2 (2) = 10.783, p = 0.005). This was driven by individuals in the *MAPT* and Tau groups being significantly younger. The age of symptom onset analysis also reflected these results with the *MAPT* and Tau patients developing symptoms much earlier than the other patient subgroups. The similarity of the *MAPT* and Tau results is not surprising given 85% of the Tau group consists of *MAPT* patients. This younger presentation and early age of onset in *MAPT* patients is supported by recent work in an extensive genetic cohort of 3,403 FTD patients which demonstrate *MAPT* patients have a significantly earlier age of symptom onset compared with *C9orf72* and *GRN* patients (Moore et al., 2020) and thus the current cohort is likely to reflect the general genetic population well in this regard.

Disease duration was found to differed between the clinical patient subgroups (χ^2 (4) = 12.937, p = 0.012) but not between the genetic and pathology-confirmed patients. This was driven by the longer duration in bvFTD subgroup and the significantly shorter disease duration in the PPA-NOS.

Finally, there were also differences in the ratio of 1.5 T vs. 3 T image acquisition between the clinical subgroups ($p = 0.025$). The lvPPA group is one of the more recent PPA classifications and images for these subgroups had a higher proportion of 3 T acquisitions.

To account for these differences, age, gender and scanner type were included as covariates in all regression analyses.

2.2.2 GENFI longitudinal same-scanner cohort

The GENetic Frontotemporal dementia Initiative (GENFI) is a consortium of 23 research centres across Europe and Canada with expertise in investigating autosomal dominant forms of FTD. It is a natural history study that aims to better understand the aetiology of and methods for detecting different aspects of the disease process in familial FTD. It enrols presymptomatic and symptomatic individuals within a family confirmed to carry an autosomal dominant mutation known to cause FTD. Non-carriers within the same family are also enrolled and used here as the healthy control comparison group. Individuals are not required to know their genetic status to be enrolled and research teams only have access to this data in a blinded manner.

To create the GENFI longitudinal cohort, I initially reviewed the data for the 829 participants who had been enrolled and acquired T1-weighted volumetric imaging to date. I filtered out individuals to keep only those that had at least two imaging assessments on the same scanner, which resulted in 418 individuals. Having visually assessed all available images for each participant (n = 1200+ scans), the best-matched same-scanner pair was chosen, resulting in a final longitudinal GENFI cohort of 382 participants. Table 2-2 provides the subgroup breakdown based on gene and symptomatic status. To more fully investigate potential presymptomatic changes at the most proximal stage to symptom onset, the presymptomatic groups were split into two categories: early presymptomatic (more than 10 years before expected age of symptom onset) and late presymptomatic (less than 10 years to the expected age of

onset). Expected years from symptom onset (EYO) for each individual was calculated separately as the difference between age at assessment and mean age of symptom onset within their family. Non-carriers from all families (and therefore all mutations in the three genes) were combined to produce the non-carrier control group for comparison.

Baseline demographics for the GENFI cohort are summarised in Table 2-2. A Kruskal-Wallis test was applied to compare baseline variable between groups for continuous variables and a Fisher's exact test employed to compare gender. These tests were applied on three levels. The first was to compare all individuals, and thus included the symptomatic, early and late presymptomatic carrier subgroups, as well as the non-carriers. The second was to investigate differences between the presymptomatic groups (early and late separately) and finally to compare the three symptomatic groups.

For the baseline comparisons including all individuals, there were significant differences in age (χ^2 (9) = 154.268, $p < 0.001$) and EYO (χ^2 (9) = 172.180, $p < 0.001$) variables. These were driven by the expected differences between the presymptomatic (early and late) compared with the symptomatic subgroups. In addition, scan interval differed significantly (χ^2 (9) = 26.004, $p = 0.002$) because of the shorter interval of the *C9orf72* symptomatic groups. Fisher's exact test demonstrated significant difference in gender distribution ($p < 0.001$). Although gender was relatively evenly matched across most subgroups, this significant result was driven by the *C9orf72* symptomatic group having substantially more men included.

Comparing the presymptomatic groups, age was significantly different between the late subgroups, driven by the significantly younger *MAPT* subgroup (χ^2 (2) = 16.490, $p < 0.001$) and EYO demonstrated difference between the early presymptomatic (χ^2 (2) = 9.020, $p = 0.011$), driven by fewer years from estimated symptom onset in the *MAPT* subgroup.

In the affected symptomatic group comparison, several variables differed significantly. Age at assessment (χ^2 (2) = 11.362, p = 0.004) and age of symptom onset (AAO) (χ^2 (2) =14.382, $p < 0.001$) were significantly younger in the *MAPT* subgroup than the *C9orf72* and *GRN* patients in line with Moore et al., (2020) seminal paper and reflecting that found in the FTD retrospective cohort discussed above in 2.2.1.

Table 2-2 Summary of demographic data for the GENFI longitudinal cohort, expressed as mean (sd) unless otherwise stated.

Measure	Non-carrier controls	*MAPT*				*C9ORF72*				*GRN*			
		Presymptomatic			Symptomatic	Presymptomatic			Symptomatic	Presymptomatic			Symptomatic
		Early (>10 EYO)	Late (<10 EYO)	Combined		Early (>10 EYO)	Late (<10 EYO)	Combined		Early (>10 EYO)	Late (<10 EYO)	Combined	
Number. of participants	159	17	17	34	11	35	15	50	25	40	41	81	22
Male / Female	70 / 89	5/12	8/9	13 / 21	6 / 5	14/21	4/11	18 / 32	20 / 5	13/27	19/22	32 / 49	8 / 22
% male	44%	29%	47%	38%	55%	40%	27%	36%	80%	33%	46%	40%	36%
Age at baseline (years)	48.2 (13.3)	37.4 (7.3)	47.4 (8.1)	42.3 (9.0)	56.6 (7.2)	39.7 (9.7)	56.6 (8.8)	44.8 (12.1)	65.3 (7.5)	38.9 (8.0)	57.5 (6.7)	48.4 (11.9)	64.3 (6.7)
EYO *	-11.3 (13.5)	-16.5 (4.9)	-3.5 (5.2)	-9.9 (8.3)	4.0 (5.6)	-21.6 (7.9)	-3.3 (4.3)	-16.1 (11.0)	3.3 (6.8)	-21.9 (6.7)	0.6 (6.6)	-11.1 (12.6)	3.4 (6.8)
Age at Symptom onset (years)	NA	NA	NA	NA	50.7 (4.8)	NA	NA	NA	59.5 (9.6)	NA	NA	NA	61.2 (6.7)
Disease duration at baseline	NA	NA	NA	NA	5.8 (5.7)	NA	NA	NA	5.8 (4.7)	NA	NA	NA	3.0 (2.4)
Scan interval (years)	1.5 (0.5)	1.3 (0.4)	1.7 (0.5)	1.5 (0.5)	1.3 (0.5)	1.3 (0.4)	1.5 (0.4)	1.3 (0.4)	1.1 (0.2)	1.4 (0.5)	1.6 (0.5)	1.5 (0.5)	1.4 (0.5)

***** Estimated years from symptom onset (calculated based on the mean age of symptom onset within the family). Negative values represent years until EYO, positive numbers represent years after EYO.

2.3 Neuroimaging

2.3.1 MRI acquisition

2.3.1.1 Retrospective FTD Neuroimaging Cohort

A requirement for inclusion into the cohort was the acquisition of at least two volumetric 3D T1-weighted MR images obtained on the same scanner. Because of the nature of the retrospective study design, images were acquired from several different 1.5 T and 3 T scanners (Table 2-3). The majority of images were acquired on a 1.5 T Signa unit (General Electric Medical Systems, Milwaukee, Wisconsin, USA) employing a spoiled gradient-echo technique between 1996 and 2006. Following a scanner upgrade in 2008, images were then acquired on a 3T Trio Tim scanner and most recently 3T Prisma (Siemens, Erlangen, Germany) employing an MPRAGE sequence.

Each scanner listed in Table 2-3 was considered as independent and scanner type was included as a covariate in analyses to adjust for differences in the acquisition. These scanner classifications (along with a visual assessment of all image pairs) were also used to match longitudinal images to avoid any inherent intensity differences unduly influencing estimates of volumetric changes across time.

Table 2-3 Breakdown of scanner information slit by scanning period and displaying manufacturer, field strength and key acquisition parameters.

Acquisition start (yr)	Scanner strength	Scanner manufacturer	Key acquisition parameters*				
			Sequence name	TI (ms)	TE (ms)	TR (ms)	Voxel size
1992–1999	1.5 Tesla	GE Signa	spoiled gradient echo	650	5	12	1.5×1.5×1.5 mm^3
2000–2004	1.5 Tesla	GE Signa	inversion recovery-prepared spoiled GRASS sequence (IR-SPGR)	650	5	12	1.5×1.5×1.5 mm^3
2005–2008	1.5 Tesla	GE Signa	spoiled gradient echo	650	5	12	1.5×1.5×1.5 mm^3
2009–2014	3 Tesla	Siemens Tim Trio (using a 32 channel coil)	magnetization-prepared rapid gradient echo (MPRAGE)	900	2.9	2200	1.1×1.1×1.1 mm^3
2015–2018	3 Tesla	Siemens MAGNETOM Prisma (using 64 channel coil)	MPRAGE sagittal	850	2.93	2000	1.1×1.1×1.1 mm^3

* TI = Inversion Time, TE = Echo Time and TR = Repetition Time

2.3.1.2 GENFI longitudinal neuroimaging Cohort

Of the 23 sites included in GENFI, 22 provided usable longitudinal data for this cohort employing one of the following five scanners listed in

Table ***2-4***. The protocols had been extensively tested by experienced MR physicists to ensure comparable image acquisition and quality across sites. To further correct for any potential biases in any acquisition differences between FTD subgroups, scanner type was included as a covariate in all regression analyses involving this cohort.

Table 2-4 Breakdown of scanner information for GENFI longitudinal imaging cohort displaying scanner type, image resolution and key acquisition parameters and timings.

Scanner type	Sequence duration	Voxel size (mm^3)	Matrix size	Key acquisition parameters				
				Sequence name	TI (ms)	TE (ms)	TR (ms)	flip angle
Siemens Prisma	5 min 6 s	1.1×1.1×1.1	256×256×208	MPRAGE sagittal, 10% phase oversampling	850	2.93	2000	8°
Siemens Skyra	8 min 32 s	1.1×1.1×1.1	256×256×208	MPRAGE sagittal, iPAT = off	850	2.85	2000	8°
Siemens Trio	5 min 6 s	1.1×1.1×1.1	256×256×208	MPRAGE sagittal, 10% phase oversampling, iPAT = 2	850	2.9	2000	8°
GE Discovery	5 min 6 s	1.1×1.1×1.1	256×256×208	IR-SPGR sagittal, ASSET = 2	400	3	6.6	8°
Philips	4 min 43 s	1.1×1.1×1.1	256×256×208	MPRAGE sagittal, SENSE = 2	933	3.1	2200	8°

2.3.2 Image processing

2.3.2.1 Quality control

To ensure the quality of the core FTD retrospective imaging results, I assessed all 2000+ candidate MR images, reviewing their entire acquisition throughout the coronal, sagittal and axial planes. Images were excluded if excessive motion, imaging artefacts (including substantial susceptibility, flow or intensity

artefacts) or incomplete head coverage precluded accurate delineation of the brain boundary.

In addition, T2-weighted images for 57 controls from the retrospective FTD cohort were available and reviewed for the presence of indicators of WM disease by an experienced neurologist (JR). Individuals with signs of moderate to severe WM damage were excluded from the control cohort (n = 7). For individuals that had multiple repeat images, I chose the best same-scanner pairs for inclusion. I excluded image pairs with less than 6 months interval between baseline and follow-up to help improve the signal-to-noise ratio as suggested by work conducted by Schott et al., (2006).

All GENFI images had previously undergone cross-sectional QC. However, in the course of performing longitudinal QC on all the image pairs side-by-side, I failed an additional 26 for significant quality or intensity differences between paired images.

2.3.3 MRI Analysis: Segmentation techniques

Inconsistent previous neuroimaging results are likely, at least in part, due to variability of segmentation accuracy and methodology. The key differences generally relate to how these automated techniques classify tissue types, which atlases are employed to perform neuroanatomical labelling and accuracy of image matching and propagation of neuroanatomical labelling onto the new target image. These differences can lead to varying anatomical definitions and regions being compared. To investigate the effect of these differences further, I applied six widely used, automated whole-brain segmentation techniques independently to each of the baseline and the repeat scans in our FTD cohort (Table 2-5). Each technique is described in detail below. Default or developer recommended settings and pipelines were used unless otherwise specified.

2.3.3.1 Image registration

Image registration is a critical part of all the automated segmentation techniques and the methods of longitudinal change explored below. In essence, registration is the process of transforming two or more images into spatial alignment. Intra-subject registration, where two or more images from the same subject are aligned, is frequently used to investigate neuroanatomical changes in individuals over time. In contrast, inter-subject registration is used to align images from different individuals. This is increasingly being used to match manually segmented neuroanatomical atlases or tissue templates from a library of individual images on to a particular subject scan in order to transfer this information and label or segment that new individual image.

Both intra- and inter-subject registration algorithms are generally composed of the following different components. These include:

1. a similarity measure, which calculates how well the two images are matched in terms of orientation, intensity and morphology. When images are well-aligned these functions tend to reach local optima (i.e. maxima or minima, depending on whether the function is estimating maximum similarity or minimal difference)
2. a transformation model, which determines how the images are to be transformed to achieve the correct spatially alignment that provides a mathematical mapping between two images or between multiple atlases with labels of interest to the target image of interest.
3. an optimisation process which aids in the assessment of the next estimate of transformation based on maximising/minimising the similarity functions, whilst often including penalty terms to ensure the images are not over-fitted

Registration algorithms can either be linear (where transformation parameters are applied globally to the entire image: Figure 2-3) or non-linear (which allows the registration parameters to be applied on a local scale).

Linear (or rigid) transformations are most suitable for the intra-subject registrations, given relatively minor deformation is likely required to match the images of the same individual. These more constrained models will avoid over-fitting the images and losing the signal of actual volumetric change that may have occurred between image acquisitions. Non-linear transformations are more appropriate for the inter-subject registrations, given the different images in a template library are likely to vary more considerably and require greater flexibility to be matched at the local level to a new candidate image requiring segmentation. Most registration algorithms are iterative until they reach a convergent criterion indicating the registration is no longer improving and a local optimum has been reached.

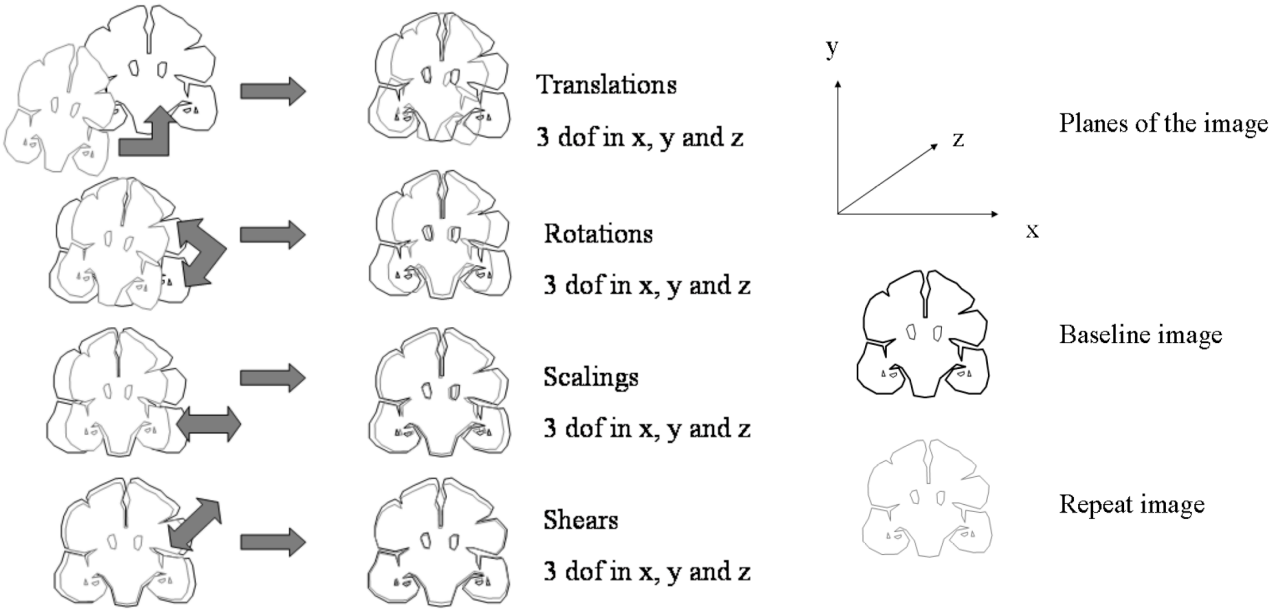

Figure 2-3 Schematic representation of a global brain registration with increasing degrees of freedom (dof).

2.3.3.2 BMAPS – Brain Multi-Atlas Propagation and Segmentation

The Multi-Atlas Propagation and Segmentation (MAPS) (Leung et al., 2011) technique utilises a large template library of semi-automated and manually edited whole-brain segmentations. The new unsegmented target image is first registered to standard space and then compared with each of the segmented template images in standard space, and these exemplars are then ranked based on a global image similarity metric that should indicate high similarity of

morphology to the target image. The labels of the 11 best-matched segmentations are then propagated through registration and then fused to create a consensus labelled whole-brain region onto the new target image. This procedure uses an updated version of STAPLE (Simultaneous truth and performance level estimation: Warfield, Zou, & Wells, 2004) to rank and propagate the best-matched exemplars. All segmentations were done on images in their original acquisition (or native) space (see Figure 2-4).

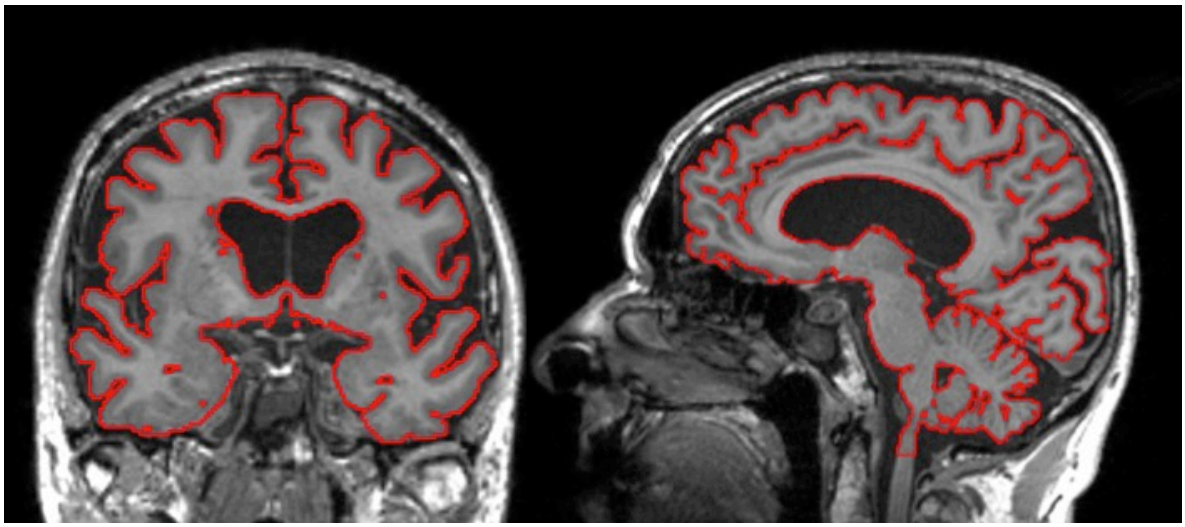

Figure 2-4 BMAPS whole-brain segmentation output from a participant included in the FTD retrospective cohort. The boundary of the BMAPS mask is indicated by the red outline.

2.3.3.3 GIF – Geodesic Information Flows parcellation

The GIF parcellation method (Cardoso et al., 2015) is an automated voxel-wise propagation and label fusion algorithm for labelling GM, WM, and neuroanatomical labels on to volumetric images. As with many of the newer automated segmentation techniques, it utilises a multi-atlas propagation approach. For the purposes of this thesis, the GIF algorithm was coupled initially with the commercially available Neuromorphometric atlas (www.neuromorphometrics.com), containing 30 manually segmented images. However, these 30 images consisted mainly of healthy controls and younger adults. In order to better capture the varied morphology and atrophy patterns evident in FTD, an additional 160 in-house images were labelled and edited to create a comprehensive atlas likely to more accurately reflect the FTD population for improved segmentation accuracy. These 160 individuals were

drawn from multiple patient populations seen at the DRC including bvFTD, svPPA, lvPPA, nfvPPA and AD (both typical amnestic and the posterior cortical atrophy variant – PCA). This included patients with genetic confirmation and covered a wide range of disease severity and atrophy profiles. GENFI participants included in these 160 cases were not included as part of the longitudinal GENFI cohort used for this thesis.

This updated atlas provides a parcellation and volumetric output for 162 discrete neuroanatomical regions (Figure 2-5 and Appendices 1). Following an N4 bias-correction stage, GIF applies a step-wise algorithm to identify and propagate the best-matched atlas exemplars onto the target image at a voxel by voxel basis. Importantly it incorporates a spatially-variant graph to better account for morphological variability between subjects so that at a small local level, it is selecting the most similar subjects from the template to propagate information from, which reduces the risk of registration error at each of these smaller locations.

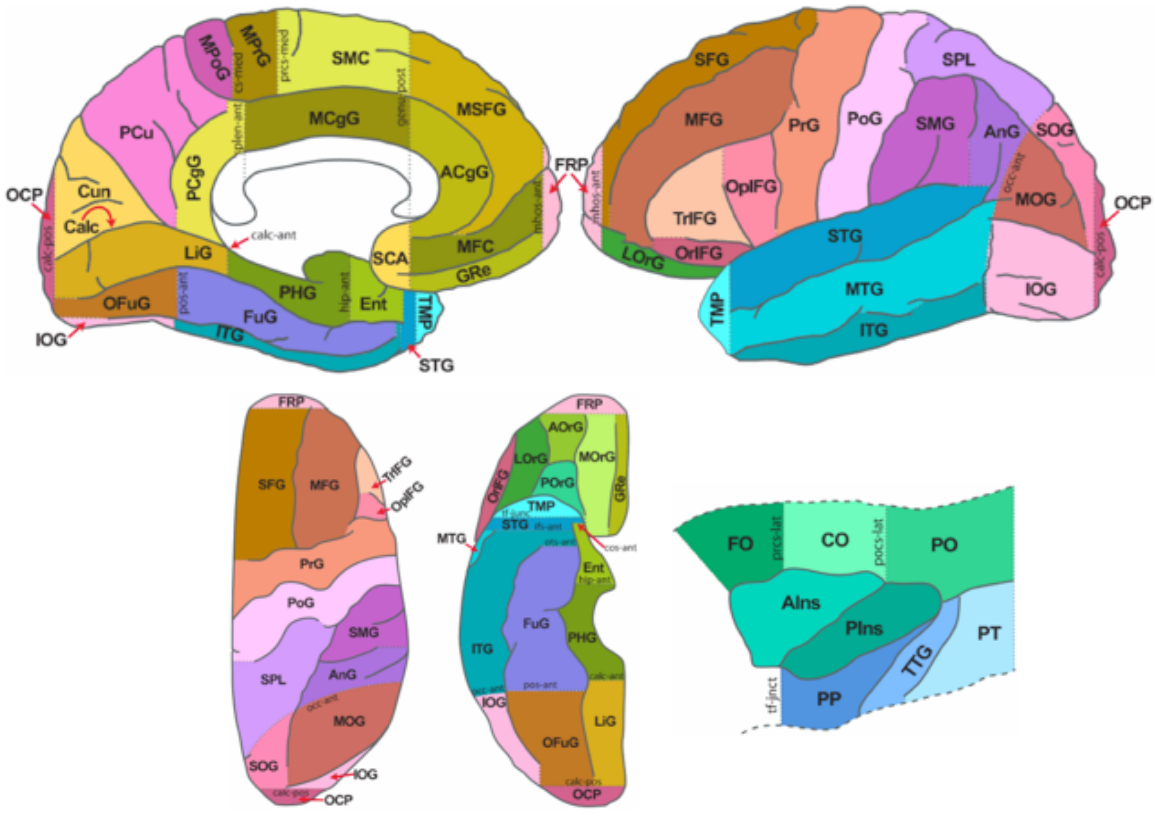

Figure 2-5 Template of the cortical labels from the Neuromorphometric manually labelled atlas used to develop the GIF parcellation.

Appendix 1 provides a list of all 162 labels used in the current dataset and the reference guide for which regions are summed to acquire the global (Chapter 3) and lobar (Chapter 4) regional measures reported in this thesis.

2.3.3.4 SPM Segment

The unified segmentation algorithm (Ashburner & Friston, 2005) was run using Statistical Parametric Mapping software (SPM12) on the MATLAB 14.1 platform (https://mathworks.com/products/matlab.html). This algorithm produces probabilistic maps for different tissue types and bias-corrected versions of the T1-weighted images. Tissue classification is based primarily on voxel intensity with each tissue type modelled as a mixture of Gaussians. It is important to note that in addition to inhomogeneity due to acquisition, intensity Gaussian distributions of tissue types still overlap in T1 images (Figure 2-6). This is because with voxels at 1.5 or 1.1 ml cubed, each voxel is likely to contain more than one tissue type, known as partial volume effects. To overcome this, SPM Segment normalises the T1 images into template space before tissue classification so that additional prior knowledge about the probability of tissue type at each voxel normalised location is included to complement intensity information. The tissues are then classified based on the probability they are GM, WM, CSF, skull and non-brain tissue. The GM and WM segmentations were combined for the whole-brain measure analysis in Chapter 3.

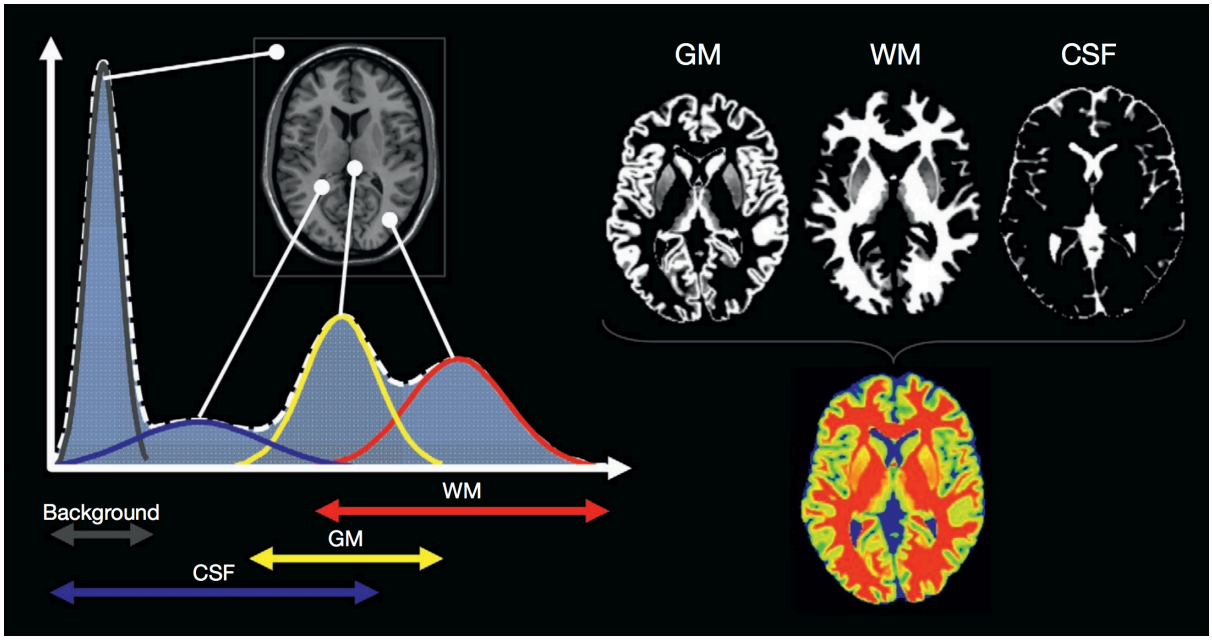

Figure 2-6 Tissue classification in SPM Segment. Left panel demonstrates the intensity distribution of different tissue types to be classified, such as grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF). There is an overlap due to partial volume effects (i.e. a voxel with a combination of tissue types included). Right panel demonstrates an example of the SPM tissue classification output. Reproduced from Kurth, Luders, & Gaser (2015).

2.3.3.5 MALP-EM – Multi-Atlas Label Propagation with Expectation-Maximisation-based refinement

MALP-EM is another freely available multi-atlas propagation method used in image analysis and clinical research (https://github.com/ledigchr/MALPEM). The key stages of the pipeline are illustrated in Figure 2-7. They include an N4 bias correction to remove inhomogeneity, followed by the pincram iterative brain extraction tool (Heckemann et al., 2015). Label propagation is accomplished using the original 30 scan template Neuromorphometric atlas to provide the exemplars and the MAPER (Multi-atlas propagation with enhanced registration (Heckemann et al., 2010)) registration approach to spatially align exemplars to the unsegmented target image. MAPER incorporates tissue probability maps into a non-rigid registration approach to improve spatial matching. The best-matched label maps are averaged into probabilistic priors for the individual structures using the joint label fusion technique developed by Wang and colleagues (Wang et al., 2013). This produces 134 neuroanatomical regions that are refined using an expectation-maximisation (EM) algorithm (Ledig et al., 2015).

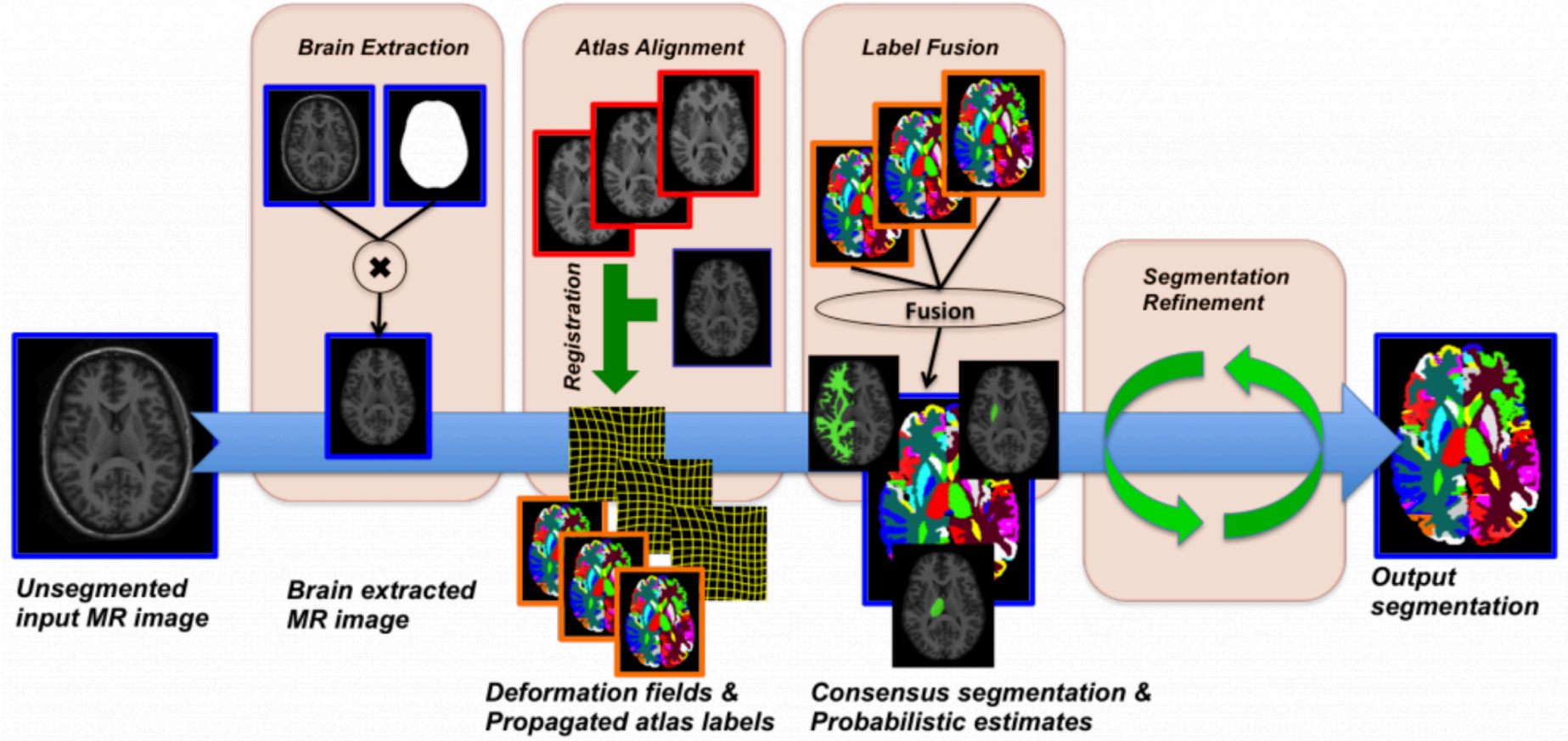

Figure 2-7 Key image processing stages in the MALP-EM segmentation pipeline. The target unsegmented image undergoes brain extraction, a non-linear iterative registration to multiple atlas labels and undergoes label fusion and iterative segmentation refinement until the exit criteria are met and final segmentation outputted. Reproduced from*: https://biomedia.doc.ic.ac.uk/software/malp-em/*

2.3.3.6 Freesurfer (v5.3)

FreeSurfer neuroimaging suite (https://surfer.nmr.mgh.harvard.edu/) is a set of multiple tools used to investigate cortical and subcortical neuroanatomical structures. It is often used for cortical thickness comparisons between groups using the initial surface-based stages of the cortical analysis pipeline. For the volumetric segmentation analyses in Chapter 3, I additionally applied the atlas-based volumetric segmentation pipeline.

The cortical analysis pipeline (Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999) extracts and classifies the GM and WM for a new target image. First, an affine registration to MNI-305 space is performed, followed by intensity normalisation and skull stripping. It then estimates the outer pial (GM to CSF boundary) and the inner white (GM to WM boundary) surface edges (Figure 2-8).

The volumetric segmentation pipeline (Fischl et al., 2002, 2004) was then used to extract cortical GM, WM and subcortical GM segmentations. These regions were combined to produce the whole-brain mask used in Chapter 3. The volume-based labelling computes the probability of the tissue class at each point based on the tissue class that appeared at that location in the manually segmented atlas. This is done in an iterative manner considering information from neighbouring voxels until the probabilities of tissue class no longer change and the segmentation is finalised. This did not include the additional pipeline for segmentation of the cerebellum.

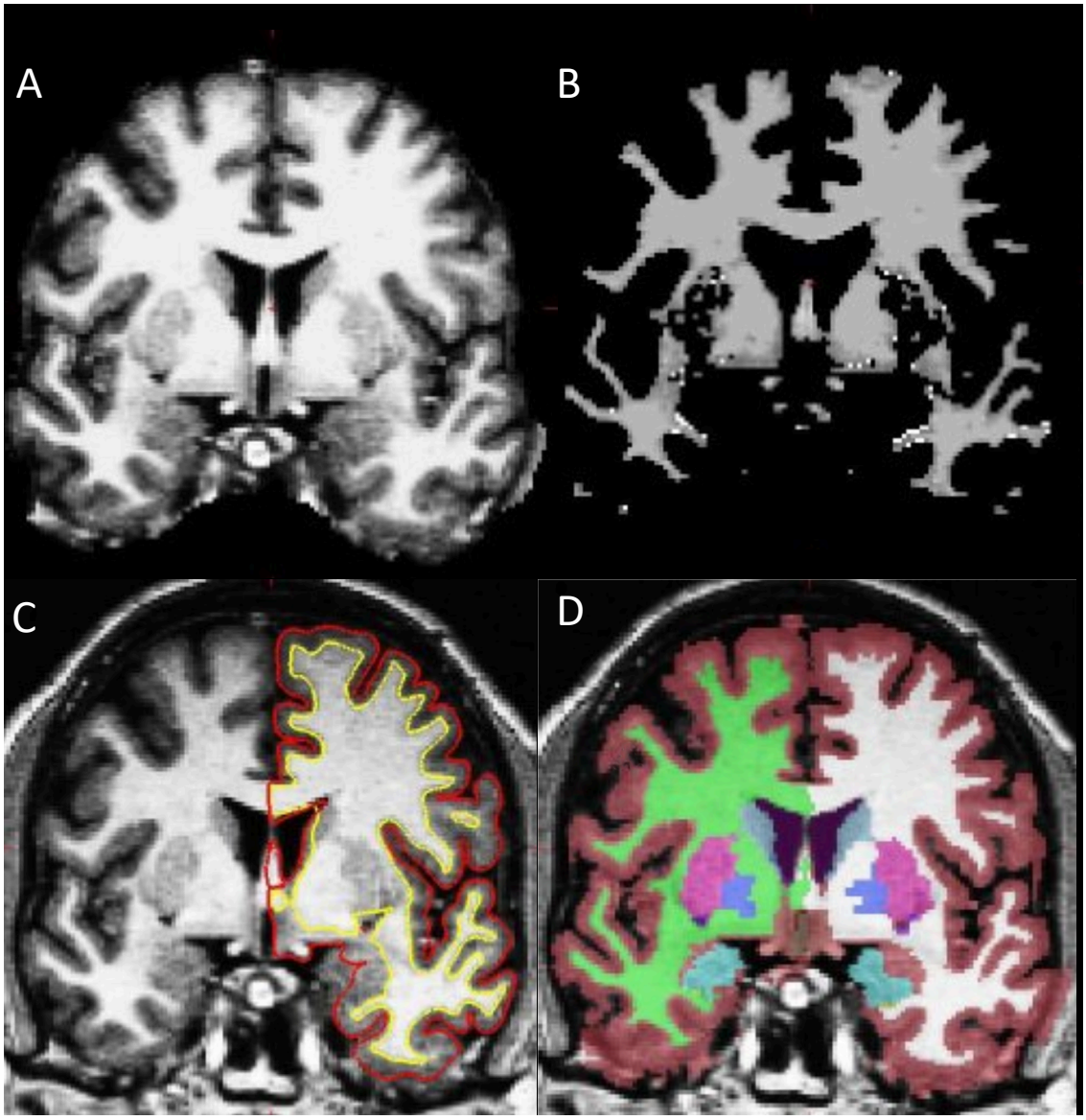

Figure 2-8 Key stages from the cortical analysis and volume-based labelling pipeline. A) skull stripping B) WM segmentation and C) image demonstrating surface between WM and GM (yellow line) and between GM and pial surface (red line), and D) GM and WM volume segmentations, represented as a single tissue class, with hemispheric distinction. Reproduced from: http://surfer.nmr.mgh.harvard.edu/

2.3.3.7 SIENAX – Structural Image Evaluation, using Normalisation, of Atrophy – cross-sectional

SIENAX (Smith et al., 2002, 2007) is another fully automated package for the extraction of whole-brain volumes. It utilises several tools from the FMRIB Software library (FSL: https://fsl.fmrib.ox.ac.uk/) to perform brain segmentation. The first stage of the pipeline is the implementation of the Brain Extraction Tool (BET), which removes all non-brain tissue from the image. The BET default settings were changed using option -B '–f 0.4' because test runs

demonstrated increasing the size of the brain estimation region produced more accurate segmentations. The default option of masking at –f 0.5 was often too restricted and resulted in brain tissue being excluded for many of the subjects. The output from this BET stage includes a stripped brain image and an estimate of the skull surface used as a scaling constraint in the following registration stage.

The second stage employs FLIRT (FMRIB's Linear Image Registration Tool), using the brain and skull images to conduct a constrained registration to the MNI-152 standard brain template space (Mazziotta et al., 1995). This registration and the corresponding image transformation information are used in the final segmentation stage to aid masking of the original image and provide brain tissue volumes that are all relative to a normalised skull size. Initially, the BET and FLIRT stages frequently failed to recognise the correct orientation of the native space input images. As can be seen from Figure 2-9 this resulted in the final segmentation stage applying the algorithm to the wrong orientation and significant misclassification of brain regions.

I attempted several fixes to provide SIENAX with the correct orientation information that corresponded to each input image using the –c (centre of gravity) option because it should inform the pipeline which orientation corresponds to coronal, axial and sagittal. This fixed only a subset of images and resulted in previously passing ones to fail. The only solution I applied that worked consistently for the entire cohort was to transform all original native space images into MNI-152 Talairach template space before running the full pipeline. Because the transformation was done uniformly to all images and the key measurement is within-subject change, I deemed the benefit of the pipeline running successfully across the full cohort outweighed the potential lack of consistency of using native space images for the other segmentation techniques. Finally, FAST (FMRIB's Automated Segmentation Tool) was applied, which produces a binary mask for GM, WM, CSF and background (non-brain). The GM and WM masks were combined using fslmaths to produce the whole-brain masks.

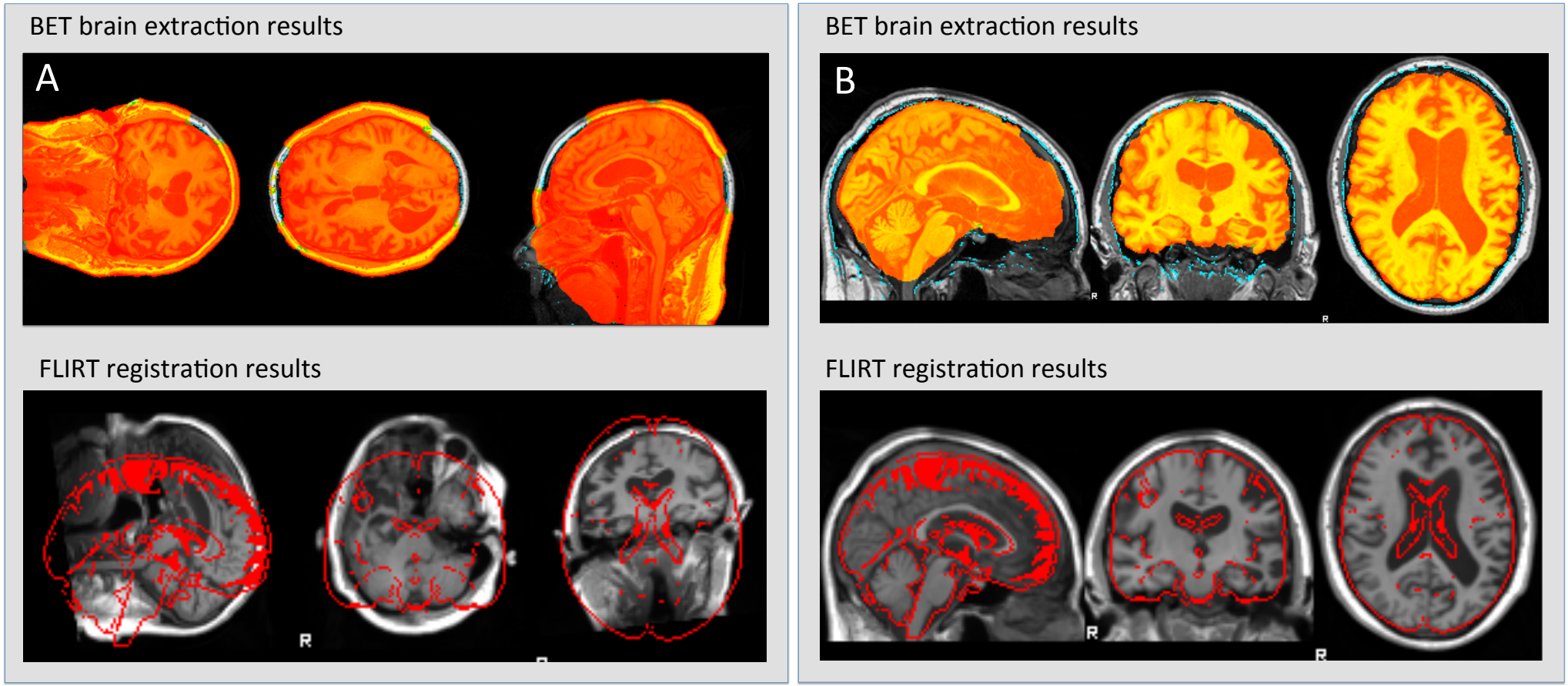

Figure 2-9 SIENAX BET and FLIRT results reported for the same input image. A) in native space and B) following pre-alignment into MNI-125 stereotactic space. The SIENAX pipeline was run with the same –d and –f options for both.

2.3.3.8 STEPS – Similarity and Truth Estimation for Propagated Segmentations

Similar to BMAPS, STEPS utilise a large template library of manually segmented regions. For this thesis, three separate manually segmented template libraries were used for the automated propagation of the hippocampus, caudate and thalamus (Chapters 5 and 6). STEPS initially employs a global similarity metric to optimise the template choices, only including a subset of the database that is reasonably well matched at the level of the image. It then employs a local similarity metric based on the fast locally normalised cross-correlation (LNCC) when comparing the target image with the template examples at the voxel level (Cardoso et al., 2013). This is to address the common issue of global versus local image matching, which is particularly important for smaller structures such as the hippocampus that can have large variations in shape and size.

Table 2-5 Summary of automated segmentation techniques applied in this thesis

Technique	Segmentation method	Intensity normalisation	Anatomical prior knowledge	Pipeline output: Segmented tissues	Binary	Probabilistic
BMAPS	Multi-atlas template propagation based on the best-matched similarity index	N3	Template library of manual or edited brain regions	Whole-brain	Y	N
GIF	Multi-atlas template propagation based on the best-matched similarity index	N4 + integrated into the pipeline	Neuromorphometric atlas (original 30 images + additional 160 in-house images)	162 cortical and subcortical GM and WM neuroanatomical labels	Y	Y
SPM	Intensity profile and location for voxel for tissue classification	Integrated into segmentation pipeline	Tissue-type probability maps	GM, WM and CSF	Y	Y
MALP-EM	Multi-atlas template propagation based on the best-matched similarity index	N4	Neuromorphometric atlas (original 30 images only)	138 cortical and subcortical neuroanatomical labels	Y	Y
SIENAX	FSL tools utilising intensity and template-based prior knowledge of location for tissue classification	Integrated into pipeline	Tissue-type probability maps (from the MNI152 dataset)	GM, WM and CSF	Y	N
Freesurfer	Intensity and location profile classification for the cortical mantle and volume tissue classification	N3	Tissue-type probability maps (from the MNI152 dataset)	GM, WM, CSF and cortical thickness estimates	Y	N
STEPS	Multi-atlas template propagation based on the best-matched similarity index	N4 + integrated	Template library of manually segmented images: Hippocampus (256 images), Caudate (202 images), Thalamus (410 images)	Single structure of the template library used. For this thesis, three template libraries were applied: hippocampus, caudate and thalamus	Y	N

2.3.4 Longitudinal investigations

Indirect measures of longitudinal change were calculated by subtracting the repeat volume from the baseline volume for each of the whole-brain and regional segmentation techniques described above. This difference was expressed as an annual percentage loss from baseline volume:

$$\frac{\frac{a-b}{a}}{interval} * 100$$

However, increasingly investigations in neurodegenerative conditions, such as AD (Anderson et al., 2012; Cash et al., 2015), HD (Tabrizi et al., 2019; Zeun et al., 2019) and multiple sclerosis (Battaglini, Jenkinson, & De Stefano, 2018) have shown that direct measures of change are likely to produce a more robust and reproducible measurement of volumetric change. This has yet to be comprehensively investigated in individuals within and across the FTD spectrum, therefore, I will be comparing several direct measures of longitudinal volumetric change with their indirect counterpart in this thesis (Table 2-6).

2.3.4.1 BSI – Boundary Shift Integral

The BSI (Freeborough & Fox, 1997) was applied to quantify longitudinal volume change between spatially aligned pairs of images. This provides an automated quantification of the volume loss at the boundary of the region of interest (ROI) specified by the input masks. These initial input masks inform the XOR (exclusive OR) region over which the BSI is computed (see Figure 2-11). It assesses the shift (or displacement) of the intensity gradient at the boundary of these aligned structures. For example, the normalised intensity is higher for GM and lower for CSF. As a region atrophies, the intensity profile at the boundary is displaced inward compared with the baseline image boundary and this intensity shift can be used to calculate the corresponding volume change between the two time points (Figure 2-10).

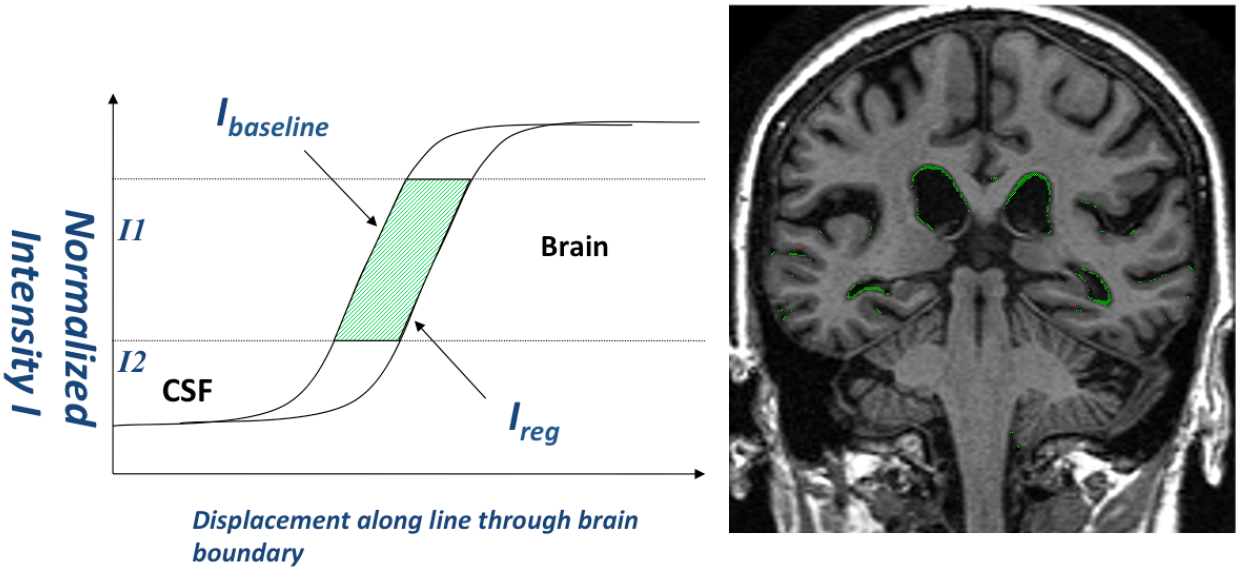

Figure 2-10 Schematic representation of the intensity profile shift between the baseline image ($I_{baseline}$) and the registered repeat image (I_{reg}) used to compute the BSI and corresponding coloured overlay on the registered image.

For all BSI measures, each image pair underwent a symmetric affine rigid 12 degrees-of-freedom (dof) registration to spatially align the baseline and follow-up image. To incorporate scaling and shearing (see Figure 2-3), a 12 dof registration was used to account for any changes in scanner voxel size that can occur over time (Clarkson et al., 2009). The transformation parameters for this registration were averaged to ascertain the midpoint space. This has been shown to avoid bias compared with a unidirectional (non-symmetric) transformation where one image is transformed (and voxels interpolated) whilst the other image remains unchanged. Once transformed into the midpoint space, both images underwent differential bias correction to further match intensity profiles of tissue types across and between the images. Volume change was measured in the average space between the image pairs (Leung, Ridgway, Ourselin, & Fox, 2012). The BSI can be applied using a binary or a probabilistic mask as the initial ROI. Where the mask was binary, I applied the k-means BSI or KBSI (Leung et al., 2010). If the segmentation technique also produced a probabilistic mask (Table 2-5), I applied the generalised BSI or GBSI (Prados et al., 2015) for comparison. The outcome of the BSI represents the volume in millilitres (ml) of brain volume change measured across the scan interval.

For this thesis, the input ROI and resulting XOR regions include whole-brain (Chapters 3 and 6), nine lobar regions (Chapter 4 – see PieBSI below) and subcortical regions derived from STEPS and GIF (Chapters 5 and 6). For the subcortical BSI analyses, I applied a double intensity window to capture the boundary shift at both GM–CSF as well as GM–WM boundaries (Hobbs et al., 2009) on MNI space image and subcortical region pairs (see Figure 5-2). An example of the key BSI stages can be seen in Figure 2-11.

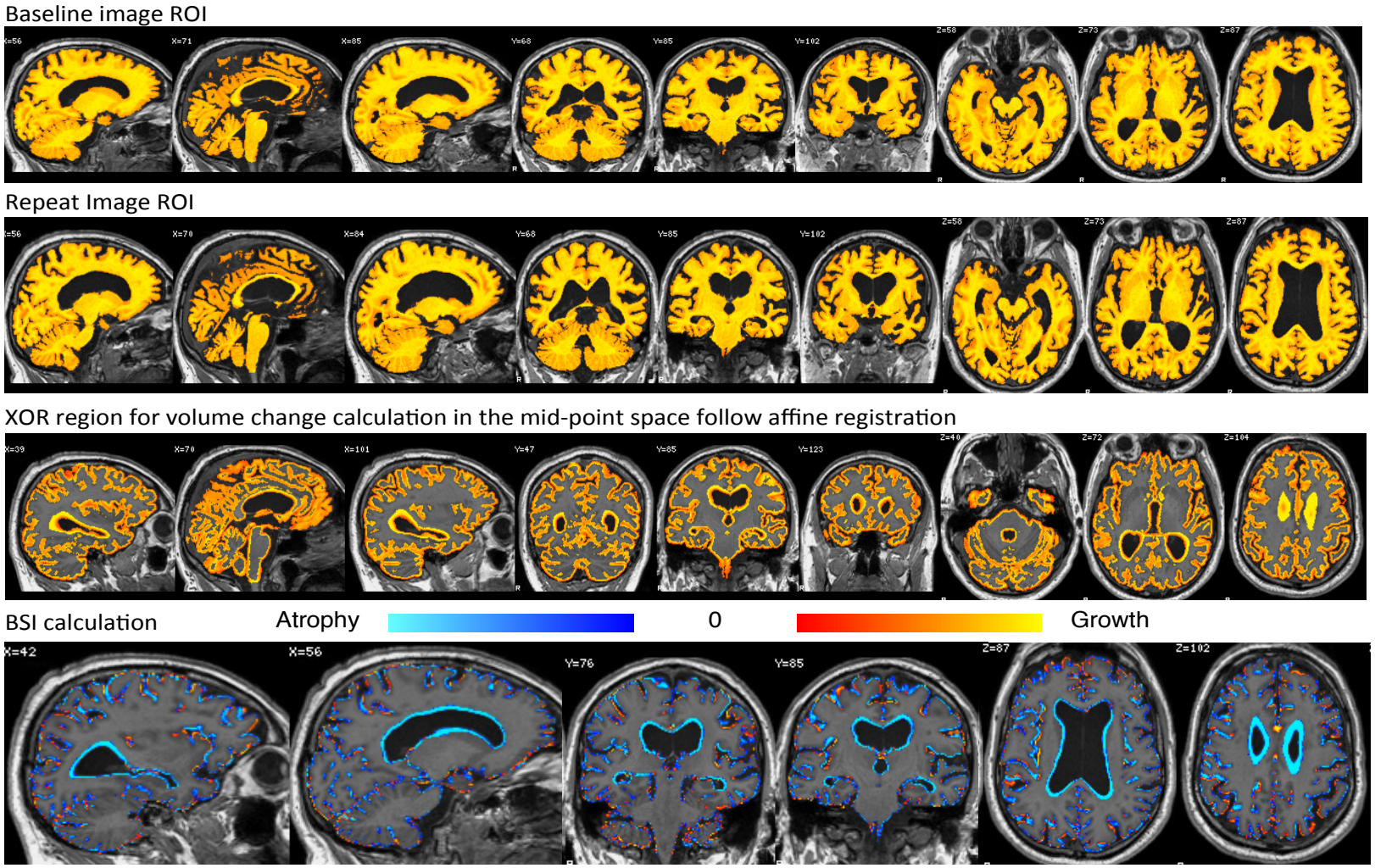

Figure 2-11 Key BSI stages representing the input ROIs, the XOR calculation and final BSI output demonstrating the distribution of atrophy.

2.3.4.2 Pie Boundary Shift Integral (PieBSI)

Measures of lobar volume change were derived using a technique under development at the Centre for Medical Image Computing (CMIC) at UCL. This technique applied the KBSI described previously using the GIF combined whole-brain parcellation as the input region. The individual cortical and subcortical GM GIF parcellation labels were then used to define nine sectors throughout the brain (Figure 2-12). Relevant GIF cortical GM labels were combined and used to define the most-lateral lobar boundaries. The method then connects the most lateral cortical GM with the most medial brain boundary, following the shortest connection via the WM. The ventricular segmentation was used to constrain the ventral boundary the algorithm sought to connect to through the WM in three dimensions. The subcortical GIF GM segmentations were included in the partitioning model as an additional constraint and guide to ensure the boundary of neighbouring sectors matched without overlap, neither both including nor either excluding these subcortical structures for any of the lobar estimations.

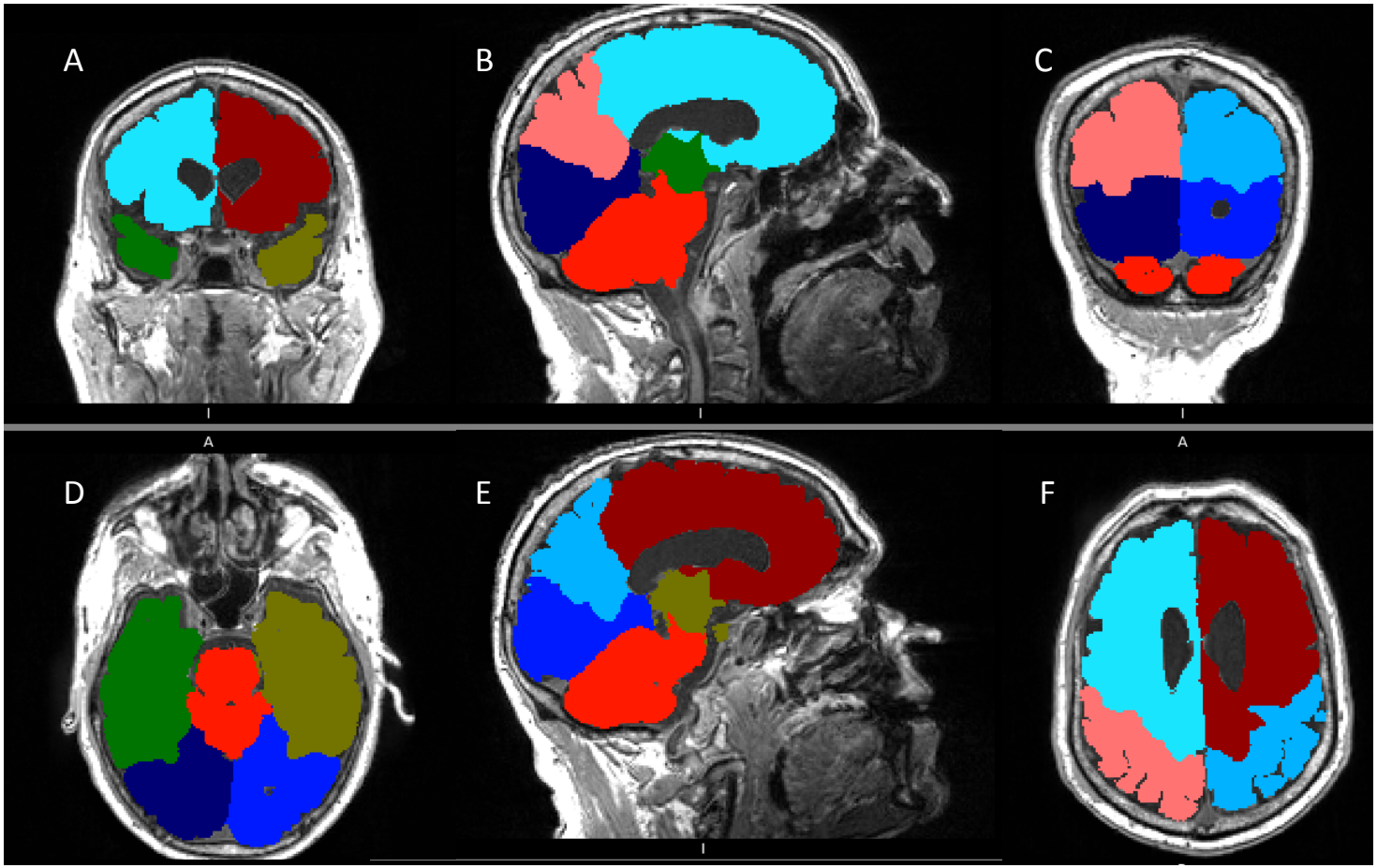

Figure 2-12 Representative image slices demonstrating the nine PieBSI quadrants used to calculate lobar rates of change. Coronal (A) – anterior, (C) – posterior), Sagittal (B) – left, (E) – right) and Axial (D) – inferior, (F) – superior). GIF cortical labels determine the external starting point and the closest ventricular co-ordinates (taking into account subcortical segmentations) inform the internal boundary in a three-dimensional manner for the quadrants. The BSIs within these regions are then outputted separately, the sum of which is equivalent to the whole-brain KBSI

2.3.4.3 Longitudinal SPM

The processing choices and steps used to perform the longitudinal SPM pipeline were informed by the official guidance included in the SPM12 manual (https://www.fil.ion.ucl.ac.uk/spm/doc/spm12_manual.pdf). Based on this, all images underwent a pre-alignment transformation into MNI-125 space to enhance the accuracy of the longitudinal pipeline. These MNI images are loaded in pairs and the interval inputted for the algorithm to discern an annualised deformation field. The paired images are then non-linearly warped and transformed into a midpoint space. The deformation or Jacobian determinants are then annualised and provide a symmetrical measure of how much each voxel had to be displaced from the baseline to the midpoint and from the repeat to the midpoint average. Each encodes the relative volume change that occurred between these images and is reported as a percentage brain volume change (PBVC).

2.3.4.4 SIENA – Structural Image Evaluation, using Normalisation of Atrophy

The final direct longitudinal method applied to measure whole-brain volume change was SIENA (Smith et al., 2002). This technique applies the cross-sectional SIENAX to automatically extract a whole-brain region from baseline and follow-up scans once they have been realigned using a three-step affine registration. These three steps include a BET brain-to-brain registration, followed by a skull-to-skull registration and finally the BET brain-to-brain registration again. These transformations are used to place the pairs into the average midpoint space for a symmetric measure of loss. Change estimation is then performed based on the edge motion at all points reported as boundaries between the brain and non-brain regions. The distance the brain surface has moved between the two time points is then expressed as a PBVC. This output is a percentage of the baseline brain and is not annualised. This method shows strong correspondence with the BSI when measuring whole-brain volume change (Smith et al., 2007). Unlike SIENAX, the input images for the longitudinal pipeline ran successfully in native space because SIENA runs its own initial pre-alignment registration as part of the pipeline.

Table 2-6 Summary of longitudinal methods applied in this thesis

Technique	Input mask	Intensity normalisation	Registration type	Output
KBSI	Any binary ROI	Differential bias correction (DBC)	Linear	ml change between timepoints over ROI
GBSI	Any probabilistic ROI	DBC	Linear	ml change between timepoints over ROI
SIENA	SIENAX whole-brain tissue segmentation	Integrated into pipeline	Linear	PBVC between the two images
Longitudinal SPM	SPM whole-brain tissue segmentations	Integrated into pipeline	Non-linear	Annual PBVC

2.3.5 Software

2.3.5.1 Image reviewing software

MIDAS (Medical Information Display and Analysis System) software (Freeborough, Fox, & Kitney, 1997) was used for the visual assessment of all 2000+ candidate scans described in Section 2.2.1.1 and the original QC of all GENFI images. This software allows visualisation of 3D T1 images in all orientations simultaneously.

NiftyMIDAS, a tool recently created at the CMIC (UCL), was used to review all segmentation outputs and registration results included in this thesis.

2.3.5.2 STATA

All statistical analyses were performed using STATA version 14.1 unless otherwise specified.

3 Automated whole-brain measures of change

3.1 Introduction

As discussed in Chapter 1, there is a large body of research demonstrating the potential of neuroimaging biomarkers in early detection, differential diagnosis and measuring of disease progression in neurodegenerative diseases. In terms of trial design, incorporating these biomarkers as surrogate end-points have been shown to have the potential to detect disease-modifying effects on fewer subjects than standard cognitive tests (Schott et al., 2010; Whitwell et al., 2015). There is also the potential for neuroimaging measures to detect these changes presymptomatically in genetic FTD (Borroni et al., 2008; Cash et al., 2018; Dopper et al., 2014; Lee et al., 2017; Panman et al., 2019; Papma et al., 2017; Popuri et al., 2018; Rohrer et al., 2015), which will be explored in detail in Chapter 6.

As a potential trial biomarker, longitudinal imaging measures should be an indicator of the underlying pathological process, be associated with disease progression and importantly be sensitive to small changes in this rate given that initial treatment effects are likely to be subtle. In practical terms, they should be relatively quick and easy to apply and reliably and accurately measure change over serial images. Importantly, they should be reliable across different centres and image acquisitions because large trials of a rare condition are likely to require multicentre or international collaborations to recruit a sufficient number of individuals. A helpful metric to assess whether a biomarker would be a useful end-point is to calculate the sample size required to detect a treatment effect with sufficient statistical power. These sample sizes are based on the rates of atrophy for the candidate population for a trial (which is usually based on a natural history study); the expectations of the treatment effect (i.e. a 30% slowing in the rate of atrophy) and the trial design. Good biomarkers produce lower sample sizes required to power a trial for shorter intervals, which would ultimately prove time- and cost-effective. Whilst longitudinal neuroimaging investigations are relatively

scarce in FTD, several of these studies have investigated global volumetric change measures and reported associated sample size calculations. As can be seen from Table 3-1, there is substantial variability and inconsistency in these sample size calculations. This is likely due to several factors including small and unrepresentative study cohorts, and overall segmentation accuracy between studies.

The power (β), significance level (α), effect size and sd of the reported sample sizes remained constant so were not required for this calculation. Sample size has an inverse squared relationship with the anticipated treatment effect, so transforming the reported sample sizes to a consistent 30% therapeutic effect can be obtained by multiplying the reported sample sizes by the inverse square of the ratio. Therefore, if deriving 30% down from 40% I multiplied the 40% sample size by $(0.4\times0.4)/(0.3\times0.3) = 1.778$. To go from 20% to 30%, a multiplication of $(0.2\times0.2)/(0.3\times0.3) = 0.444$ was used

For example, Gordon et al., (2010), reports a sample size of 45 svPPA patients required to detect a 40% reduction in whole-brain atrophy. Holding all other factors constant, a sample of 80 ($45\times1.788 = 80.46$) patients would be required to detect a 30% treatment effect.

Table 3-1 Summary of previously published sample size estimates from studies investigating longitudinal global volumetric changes in FTD, split by patient subgroup. The power, β, level of treatment effect, statistical significance level, α, and additional corrections used for the calculation are listed under each publication

Publication and cohort	Measure	bvFTD	svPPA	nfvPPA	lvPPA	FTD-combined	*MAPT*	*C9orf72*	*GRN*
Rohrer et al. (2008)									
β = 90%, α = 0.05									
svPPA (n = 21, 8 FTLD-U)									
30% treatment effect	Whole brain		118						
	Ventricles		89						
Knopman et al. (2009)									
β = 80%, α = 0.05, correcting for									
26% attrition rate bvFTD (n = 34), svPPA (n = 16), nfvPPA (n = 17), lvPPA (n = 9)									
25% treatment effect	Whole brain	165	135	105	81				
	Ventricles	127	58	55	50				
30% treatment effect**	Whole brain	117	96	75	57				
	Ventricles	91	43	41	36				
40% treatment effect	Whole brain	66	54	42	32				
	Ventricles	51	24	23	20				

Publication and cohort	Measure	bvFTD	svPPA	nfvPPA	lvPPA	FTD-combined	*MAPT*	*C9orf72*	*GRN*
Gordon et al. (2010)									
β = 80%, α = 0.05,									
adjusting for control rates and annual attrition of 10%									
bvFTD (n = 11), svPPA (n = 11), nfvPPA (n = 10)									
25% treatment effect	Whole brain	360	115	26					
	Ventricles	495	157	152					
30% treatment effect**	Whole brain	251	80	18					
	Ventricles	343	108	105					
40% treatment effect	Whole brain	141	45	10					
	Ventricles	193	61	59					
Rohrer et al. (2012)									
β = 90%, α = 0.05,									
adjusting for control rates									
svPPA (n = 17), nfvPPA (n = 18)									
30% treatment effect	Whole brain		120	70					
	Ventricles		118	78					

Publication and cohort	Measure	bvFTD	svPPA	nfvPPA*	lvPPA	FTD-combined	*MAPT*	*C9orf72*	*GRN*
Mahoney et al. (2015)									
β = 80%, α = 0.05, adjusting for control rates									
bvFTD (n = 19, including 8 *MAPT* and 4 *C9orf72* mutations)									
20% treatment effect	Whole brain	507							
30% treatment effect	Whole brain	226							
40% treatment effect	Whole brain	127							
Whitwell, Duffy, et al., (2015)									
β = 80%, α = 0.05									
Progressive apraxia of speech (n = 24)									
20% treatment effect	Whole brain			116					
	Ventricles			118					
30% treatment effect**	Whole brain			52					
	Ventricles			52					
Whitwell, Boeve et al. (2015)									
β = 80%, α = 0.05 *MAPT* (n = 21), *C9orf72* (n = 11), *GRN* (n = 11), sporadic FTD (n = 15, 14 bvFTD and 1 FTD-MND)									
20% treatment effect	Whole brain					87	238	135	102
	Ventricles					119	191	56	61
30% treatment effect**	Whole brain					39	106	60	45
	Ventricles					53	85	25	27

Publication and cohort	Measure	bvFTD	svPPA	nfvPPA	lvPPA	FTD-combined	*MAPT*	*C9orf72*	*GRN*
Pankov et al (2016)									
β = 80%, α = 0.05 bvFTD, n = 37 [14 in 1.5 T and 23 in 3 T]), svPPA, n = 49 [29 in 1.5 T and 20 in 3 T]) Sample sizes reported were only for the combined cohort and not reported separately for the bvFTD and svPPA									
20% treatment effect	Whole brain 1.5 T					157			
	Whole brain 3 T					75			
	Lateral Ventricles 1.5 T					332			
	Lateral Ventricles 3 T					176			
30% treatment effect**	Whole brain 1.5 T					70			
	Whole brain 3 T					33			
	Lateral Ventricles 1.5 T					147			
	Lateral Ventricles 3 T					78			
Binney et al. (2017)									
β = 80%, α = 0.05									
bvFTD (n = 44), svPPA (n = 30), or nfvPPA (n = 26)									
20% treatment effect	Whole brain (GM)	521	111	172					
30% treatment effect **	Whole brain (GM)	231	49	76					
40% treatment effect	Whole brain (GM)	131	29	44					

The estimates represent the sample size that would be required per trial arm to detect the stated % treatment effect from measures of rates of change. The power (β) and statistical threshold (α) used to calculate the sample size for each study are listed individually for every publication, as are additional factors such as controlling for normal ageing and expected attrition rate from the putative trial.

*Progressive apraxia of speech (PPAOS) for Whitwell et al. (2015)

** 30% calculated from the reported 20% or 40% treatment effect in each publication.

Segmentation accuracy of volumetric images is fundamental to inform realistic tissue models and measure disease-related structural changes in the brain. Because of the labour-intensive nature of manual delineation of images, there has been considerable attention paid to the development of increasingly sophisticated methods of automated segmentation, as discussed in Chapter 2 (González-Villà et al., 2016; Iglesias & Sabuncu, 2015; Shaikh & Ali, 2019). There are currently many freely available tools that are widely used in clinical research (Ashburner & Friston, 2005; Cardoso et al., 2015; Fischl et al., 2002; Ledig et al., 2015; Leung et al., 2011; Zhang, Brady, & Smith, 2001). Previous comparisons have tended to focus on the most commonly used fully integrated packages such as those included in Freesurfer, SPM and the FSL library (Johnson et al., 2017; Katuwal et al., 2016).

Applying these techniques to both phantom and patient MRIs, comparative studies have shown that whilst these techniques perform well at delineating structural images, there are subtle but important differences in accuracy and reliability. For example, Fellhauer and colleagues (2015) found that when applied to patients with AD, MCI (mild cognitive impairment) and to control data, Freesurfer produced the largest GM volumes alongside the smallest WM volumes; SPM produced the largest WM and FSL calculated the smallest GM from the same image sets. Scan quality was also an issue, with SPM providing the most accurate segmentations when image quality was poor. Comparative studies have also shown significant differences in segmentation accuracy and reliability depending on the software version, operating system and workstation type (Gronenschild et al., 2012); as well as highlighting the importance of visual inspection of automated pipelines (Iscan et al., 2015). They have also shown variability in performance when segmenting patient images compared with controls (Katuwal et al., 2016).

With the majority of these automated methods using atlases and templates derived from healthy brains, it is perhaps unsurprising their accuracy in labelling altered and/or atrophied morphology is affected. As briefly mentioned above, an additional hurdle is scan quality. Patients (particularly those with

behavioural issues) are likely to have difficulty remaining still in the scanner, with increased motion artefact affecting the tissue boundaries and distinctions being an important consideration to take into account.

The current project aims to address the following three issues:

1) Previous longitudinal neuroimaging investigations have included small and variable FTD cohorts that may have limited comparability, generalisability and contributed to inconsistent results.
2) The application of different techniques and analysis methods for measuring atrophy in FTD between studies further complicates comparison and consistent results within each subgroup.
3) There remains a lack of any head-to-head comparison of automated measures of brain volume and atrophy within an FTD population, especially one including all key clinical, genetic and pathology subgroups.

Therefore, the main motivation for this project was to investigate the performance of fully automated segmentation and longitudinal change measures across the full spectrum of FTD disorders and subgroups. More specifically, I was interested in their comparative performance in terms of providing accurate longitudinal measures that could provide potential non-invasive biomarkers in FTD intervention trials.

3.2 Methods

3.2.1 Participants

Participants consisted of the full FTD retrospective cohort described in Chapter 2 (n = 262). Please refer to Table 2-1 and Section 2.2.1 for a full discussion of their demographic characteristics. For the head-to-head comparison, it was important that I ran these with the same subjects and same images included for each technique to avoid any additional bias. To perform this direct head-to-head comparison, a subset cohort was created by removing any individuals who had a single missing value for any of the measures. The summary mean and standard deviation (sd) of all annual rates of change measures for both the full FTD cohort and comparison subset cohort can be found in Table 3-3 and Table 3-4, respectively.

Linear regression analyses confirmed that the comparison subset cohort of 226 subjects did not significantly differ on any of the baseline demographics or the 17 annualised rate of change measures investigated in this chapter.

3.2.2 Image analysis

3.2.2.1 Segmentation

Before application of the segmentation techniques, all images underwent an N4 bias correction for inhomogeneity (Tustison, Cook, & Gee, 2010). I then applied the following six segmentation techniques independently to all 524 images (262 baseline and repeat pairs). SPM12 Segment, BMAPS, MALP-EM (v2.1), Freesurfer (v5.3) and GIF (v3) were all applied to the native space images using default or developer recommended pipelines and options. In addition, I transformed the full set of 524 images into MNI-125 template space using Niftyreg and visually assessed that all had completed the realignment correctly. The SIENAX pipeline was then applied to these pre-aligned images based on previous issues completing the FLIRT and BET stages of the pipeline successfully in native space. Refer to Chapter 2.3.3.7 for additional information on the reasoning for this. A representation of the output for the six segmentation methods can be seen in Figure 3-1.

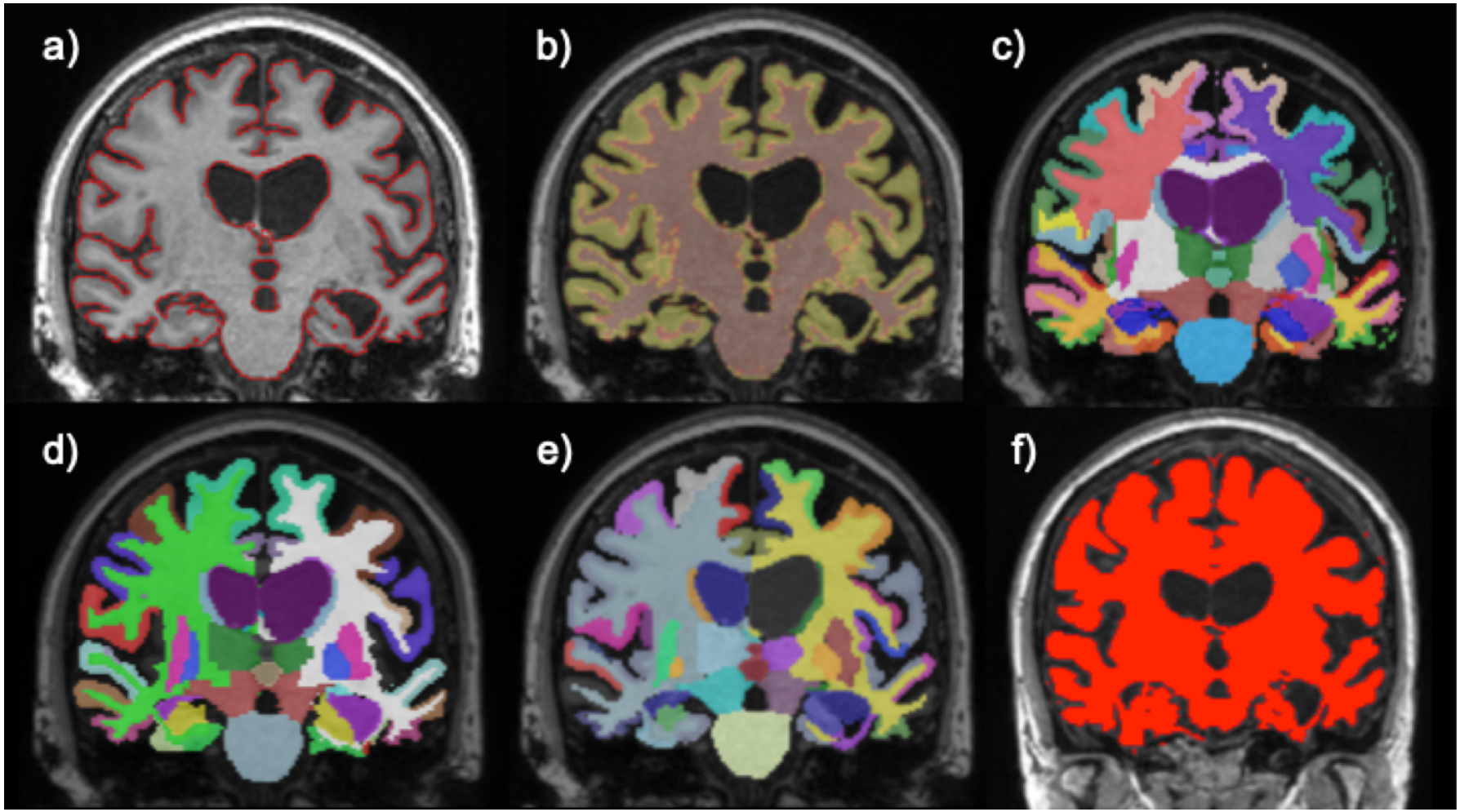

Figure 3-1 Representation of the image label outputs for the six automated whole-brain segmentation techniques: a) BMAPS, b) SPM12, c) GIF, d) Freesurfer, e) MALP-EM, f) SIENAX.

3.2.2.2 Longitudinal rates of change

As discussed in Section 2.3.4, there are direct and indirect methods for measuring longitudinal volume change. For the indirect measure, I summed all appropriate volumes to acquire a single whole-brain volume for the baseline and the repeat image for all 262 pairs. BMAPS provides only a single output for the whole brain, so did not require any additional operation to output the final volume. MALP-EM also outputs a combined tissues volume in its output report. For SPM12, SIENAX and Freesurfer, I combined the GM and WM outputs to obtain the final brain volume, and for GIF I combined the appropriate 162 region volumes. The repeat volume was then subtracted from the baseline volume for each pair and method and expressed as an annual percentage change from baseline. These indirect brain volume (BV) difference measures are referred to in all tables as SPM_BV, BMAPS_BV, MALP-EM_BV, SIENAX_BV, Freesurfer_BV and GIF_BV (and without the underscore in Figure 3-2 because labelling did not allow for it).

$$\left(\frac{\frac{a-b}{a}}{interval} \right) * 100$$

To run the direct BSI methods for calculating change, whole-brain binary mask (for KBSI) and probabilistic masks (for GBSI) were extracted (Table 2-5 summarises which technique provides both). MALP-EM outputs a combined probabilistic mask used for GBSI, which I binarised using fslmaths for application of the KBSI. GM and WM masks were combined for SPM12, SIENAX and Freesurfer and the 144 GIF labels that include cortical GM, WM or subcortical GM of 162 total labels produced (see Appendix 1). The probabilistic masks for SPM and GIF were also binarised.

Each of these masks were then used as the input ROI for the KBSI (all six segmentation techniques) and the GBSI (for GIF, SPM and MALP-EM), which was calculated in midpoint space. The BSI outputs ml difference between the two images. These are expressed as an annual% volume change from

baseline volume. As with the indirect measures labelling, these will be referred to as the segmentation method followed by _KBSI or _GBSI where applicable.

SIENA and the longitudinal SPM pipeline were also applied as direct measures of change to the N4 bias-corrected images in native space. The longitudinal SPM pipeline produces the annualised PBVC and the PBVC produced using SIENA was annualised by dividing by scan interval.

3.2.2.3 Quality control

To ensure the integrity of the imaging data, I visually assessed all 3156 whole-brain segmentations using NiftyMidas. For GIF, SPM and MALP-EM, I also checked the 1578 binarised masks to ensure the conversion had worked. All of these visual assessments were done blinded to patient and diagnosis. NiftyMidas allows the visualisation of images in all three orientations simultaneously and I reviewed the entire segmentation region before the final QC decision. If a segmentation failed for either baseline or repeat image, the indirect longitudinal BV metric for that technique and individual could not be calculated. If the repeat segmentation failed, that pair of images was still used for the BSI because it is robust to segmentation error and includes a dilation and conditional erosion stage in the estimation of the XOR. However, if the baseline failed, then the BSI was not run because the output is reported as a percentage change from baseline and thus would be introducing undue error into the calculation. Because the SIENA and Longitudinal SPM pipelines do not require an input region, they were run on all 262 pairs.

Following the segmentation QC and subsequent running of all BSI and longitudinal pipelines, the longitudinal QC was performed. This involved visually assessing the 2357 registration runs (262 registrations for each of the nine BSI measures except Freesurfer_KBSI given that one baseline region had failed for this measure: see Table 3-2). Visual review and QC were performed using NiftyMidas, which allows switching between the baseline and repeat registered images to assess the efficacy of registration, whilst also providing the ROI for an additional check that the XOR was derived from the

correct input and the BSI overlay, demonstrating where and how much the volume change was being quantified between the two images. The observations (Obs) column shows the final number of passing image pairs included for each of the measures across the subgroups. Reassuringly, many of the measures passed with the full set of data points contributed per subgroup. Only small numbers per measure failed across the subgroups, with the SPM_GBSI incurring the largest loss at just 7 failures out of 78 image pairs from the controls (9%).

Table 3-2 summarises segmentation and registration failures for each of the six automated segmentation techniques. I reviewed all SIENA outputs and reports, and none appeared to be calculated over an inaccurate region or biased in alignment. I also visually assessed the longitudinal SPM deformation fields for the 262 pairs. All deformation fields were applied to the correct regions and no gross errors were evident.

Table 3-2 Overall numbers of segmentations and registrations that failed for each of the six segmentation methods due to incomplete pipeline execution or QC issue

n = 262	**Segmentation**		**KBSI**		**GBSI**	
	Pipeline failed to complete	QC Fail	Pipeline failed to complete	QC Fail	Pipeline failed to complete	QC Fail
SPM	1 Rp	0	4	10	4	9
BMAPS	0	0	0	13	NA	NA
MALP-EM	0	0	2	8	1	12
SIENAX	0	0	1	4	NA	NA
Freesurfer	0	1 Bl, 3 Rp	0	5	NA	NA
GIF	0	0	0	10	0	12

* Bl = baseline image, Rp = repeat image.

Table 3-3 Mean and standard deviation (sd) for annualised rates of change measures for controls and FTD clinical, genetic and pathological subgroups (n = 262)

n = 262	Controls (n = 78)			Clinical subgroups													
				bvFTD (n = 66)			svPPA (n = 45)			nfvPPA (n = 45)			PPA-NOS (n = 7)			lvPPA (n = 21)		
Measure	**Obs***	**Mean**	**sd**	**Obs**	**Mean**	**sd**	**Obs**	**Mean**	**sd**	**Obs**	**Mean**	**sd**	**Obs**	**Mean**	**sd**	**Obs**	**Mean**	**sd**
SPM_BV	78	0.6	(1.0)	66	2.8	(2.4)	44	3.1	(1.8)	45	3.2	(2.3)	7	3.3	(1.0)	21	3.6	(2.7)
SPM_KBSI	72	0.5	(0.7)	62	1.9	(1.5)	43	2.1	(1.0)	44	2.7	(1.4)	6	2.7	(1.0)	20	2.2	(1.4)
SPM_GBSI	71	0.5	(0.8)	63	2.2	(1.5)	43	2.4	(0.9)	44	2.9	(1.4)	7	2.8	(0.8)	20	2.4	(1.5)
Long_SPM_PBVC	78	0.2	(0.6)	66	1.0	(0.9)	45	1.0	(0.7)	45	1.2	(1.0)	7	1.6	(0.6)	21	1.5	(0.7)
BMAPS_BV	78	0.4	(0.9)	66	2.0	(1.8)	45	2.4	(1.2)	45	2.9	(1.8)	7	2.8	(1.0)	21	2.5	(1.6)
BMAPS_KBSI	74	0.5	(0.8)	64	2.0	(1.3)	41	2.2	(0.9)	44	2.7	(1.3)	7	2.6	(0.9)	19	2.4	(0.9)
MALP-EM_BV	78	0.0	(1.8)	66	1.1	(2.5)	45	0.5	(1.9)	45	0.6	(1.8)	7	1.6	(0.7)	21	1.3	(4.3)
MALP-EM_KBSI	75	0.3	(0.5)	63	1.4	(1.0)	45	1.5	(0.6)	42	1.8	(1.0)	7	2.0	(0.7)	20	1.7	(0.6)
MALP-EM_GBSI	72	0.4	(0.7)	62	1.6	(1.1)	45	1.8	(0.7)	44	2.1	(1.1)	7	2.3	(0.7)	19	1.7	(1.1)
SIENAX_BV	78	0.4	(1.1)	66	1.2	(2.3)	45	1.4	(2.3)	45	1.8	(1.8)	7	2.4	(1.1)	21	1.7	(1.6)
SIENAX_KBSI	77	0.5	(0.7)	65	1.6	(1.2)	42	1.9	(0.8)	44	2.5	(1.2)	7	2.6	(0.8)	21	1.9	(1.4)
SIENA_PBVC	78	0.5	(0.7)	66	2.4	(1.9)	45	3.0	(1.0)	45	3.3	(2.0)	7	3.2	(1.2)	21	3.0	(1.3)
Freesurfer_BV	77	0.7	(1.3)	64	2.2	(3.9)	45	2.1	(4.2)	45	2.6	(3.0)	7	2.7	(0.9)	20	0.8	(3.6)
Freesurfer_KBSI	76	0.3	(0.6)	64	1.6	(1.3)	44	2.0	(1.0)	45	2.4	(1.3)	7	2.5	(0.8)	20	1.7	(1.4)
GIF_BV	78	0.2	(0.6)	66	1.1	(1.2)	45	1.1	(0.7)	45	1.4	(0.7)	7	1.7	(0.8)	21	1.2	(1.7)
GIF_KBSI	76	0.5	(0.7)	64	1.7	(1.2)	41	1.9	(0.7)	44	2.3	(1.1)	7	2.4	(0.8)	20	1.9	(1.2)
GIF_GBSI	73	0.4	(0.7)	63	1.7	(1.2)	43	1.9	(0.7)	44	2.3	(1.1)	7	2.4	(0.8)	20	2.0	(0.8)

* Obs = number of passing observations used to produce mean (sd) results per technique.

Table 3-3 continued: Mean and standard deviation (sd) for annualised rates of change measures for controls and FTD clinical, genetic and pathological subgroups (n = 262)

n = 262	Genetic subgroups									Pathology subgroups					
	MAPT (n = 16)			*C9orf72* (n = 10)			*GRN* (n = 8)			Tau (n = 19)			TDP-43 (n = 34)		
Measure	**Obs**	**mean**	**sd**	**Obs**	**mean**	**sd**	**Obs**	**mean**	**sd**	**Obs**	**mean**	**sd**	**Obs**	**mean**	**sd**
SPM_BV	16	2.1	(1.7)	10	2.6	(3.0)	8	4.2	(3.7)	19	2.4	(1.6)	34	3.1	(2.5)
SPM_KBSI	15	1.8	(1.0)	9	1.7	(1.5)	8	3.7	(2.0)	18	1.9	(1.0)	32	2.5	(1.5)
SPM_GBSI	15	2.0	(1.1)	10	1.9	(1.4)	8	3.9	(2.1)	18	2.1	(1.1)	33	2.6	(1.6)
Long_SPM_PBVC	16	1.0	(0.9)	10	1.2	(0.8)	8	0.6	(1.2)	19	1.0	(0.8)	34	0.9	(0.9)
BMAPS_BV	16	1.8	(1.4)	10	2.2	(1.5)	8	3.0	(2.6)	19	2.0	(1.3)	34	2.5	(1.7)
BMAPS_KBSI	16	1.8	(1.0)	10	2.0	(1.3)	8	3.7	(1.8)	19	1.9	(1.0)	33	2.5	(1.4)
MALP-EM_BV	16	0.4	(1.0)	10	0.7	(1.7)	8	-0.7	(4.0)	19	0.6	(1.1)	34	0.4	(2.2)
MALP-EM_KBSI	16	1.2	(0.7)	10	1.2	(0.7)	8	2.6	(1.1)	19	1.2	(0.7)	33	1.7	(0.9)
MALP-EM_GBSI	16	1.4	(0.9)	10	1.4	(0.9)	8	3.0	(1.3)	19	1.5	(0.9)	33	1.9	(1.1)
SIENAX_BV	16	0.8	(2.9)	10	1.5	(1.5)	8	1.5	(1.9)	19	1.0	(2.7)	34	1.5	(1.6)
SIENAX_KBSI	16	1.6	(1.0)	10	1.6	(1.2)	8	3.2	(1.6)	19	1.7	(1.0)	34	2.1	(1.2)
SIENA_PBVC	16	2.2	(1.2)	10	2.3	(1.1)	8	4.4	(2.4)	19	2.4	(1.2)	34	2.9	(1.8)
Freesurfer_BV	16	1.2	(1.7)	10	1.7	(3.1)	8	3.7	(1.7)	19	1.7	(2.3)	34	2.6	(3.4)
Freesurfer_KBSI	16	1.6	(0.9)	10	1.6	(1.1)	8	2.8	(2.0)	19	1.7	(1.0)	34	2.1	(1.4)
GIF_BV	16	1.0	(0.6)	10	1.4	(1.0)	8	1.2	(1.8)	19	1.1	(0.6)	34	1.3	(1.1)
GIF_KBSI	16	1.6	(0.8)	10	1.6	(1.1)	8	3.1	(1.5)	19	1.6	(0.8)	33	2.1	(1.1)
GIF_GBSI	15	1.7	(0.8)	10	1.6	(1.1)	8	3.1	(1.4)	18	1.7	(0.8)	34	2.1	(1.1)

Table 3-4 Mean and standard deviation (sd) of annualised rates of change measures for controls and the clinical, genetic and pathology FTD subgroups for the comparison subset FTD cohort (n = 226)

n = 226	Controls (n = 66 of 78)			Clinical subgroups													
				bvFTD (n = 56 of 66)			svPPA (n = 38 of 45)			nfvPPA (n = 42 of 45)			PPA-NOS (n = 6 of 7)			lvPPA (n = 18 of 21)		
Measure	mean	(sd)	% dropped *	mean	(sd)	% dropped	mean	(sd)	% dropped	mean	(sd)	% dropped	mean	(sd)	% dropped	mean	(sd)	% dropped
SPM_BV	0.5	(1.0)	15	2.8	(2.1)	15	2.9	(1.8)	14	3.2	(2.3)	7	3.2	(1.1)	14	3.2	(2.7)	14
SPM_KBSI	0.4	(0.6)	8	2.0	(1.4)	10	2.2	(1.0)	12	2.7	(1.4)	5	2.7	(1.0)	0	2.2	(1.4)	10
SPM_GBSI	0.4	(0.7)	7	2.2	(1.5)	11	2.4	(0.9)	12	2.8	(1.4)	5	2.8	(0.9)	14	2.3	(1.5)	10
Long_SPM_PBVC	0.2	(0.5)	15	1.1	(0.9)	15	0.9	(0.7)	16	1.2	(1.0)	7	1.5	(0.6)	14	2.6	(0.7)	14
BMAPS_BV	0.4	(0.8)	15	2.2	(1.7)	15	2.3	(1.2)	16	2.7	(1.8)	7	2.9	(1.0)	14	2.7	(1.6)	14
BMAPS_KBSI	0.4	(0.7)	11	2.0	(1.3)	13	2.3	(0.9)	7	2.6	(1.3)	5	2.7	(0.9)	14	2.4	(0.9)	5
MALP-EM_BV	0.2	(1.0)	15	1.0	(2.5)	15	0.4	(1.9)	16	0.5	(1.8)	7	1.7	(0.7)	14	1.8	(4.3)	14
MALP-EM_KBSI	0.3	(0.4)	12	1.4	(1.0)	11	1.5	(0.6)	16	1.8	(1.0)	0	2.0	(0.8)	14	1.7	(0.6)	10
MALP-EM_GBSI	0.3	(0.6)	8	1.6	(1.1)	10	1.8	(0.7)	16	2.1	(1.1)	5	2.3	(0.8)	14	1.7	(1.1)	5
SIENAX_BV	0.2	(1.0)	15	1.2	(2.4)	15	1.6	(2.3)	16	1.8	(1.8)	7	2.5	(1.1)	14	1.7	(1.6)	14
SIENAX_KBSI	0.4	(0.7)	14	1.7	(1.1)	14	1.9	(0.8)	10	2.4	(1.2)	5	2.7	(0.9)	14	1.9	(1.4)	14
SIENA_PBVC	0.5	(0.6)	15	2.6	(1.9)	15	3.0	(1.0)	16	3.3	(2.0)	7	3.2	(1.3)	14	3.2	(1.3)	14
Freesurfer_BV	0.6	(1.2)	14	2.6	(3.9)	13	1.6	(4.2)	16	2.6	(3.0)	7	2.8	(1.0)	14	1.3	(3.6)	10
Freesurfer_KBSI	0.3	(0.5)	13	1.7	(1.3)	13	2.0	(1.0)	14	2.4	(1.3)	7	2.6	(0.9)	14	1.8	(1.4)	10
GIF_BV	0.2	(0.6)	15	1.1	(1.2)	15	1.1	(0.7)	16	1.4	(0.7)	7	1.8	(0.9)	14	1.4	(1.7)	14
GIF_KBSI	0.4	(0.7)	13	1.7	(1.1)	13	1.9	(0.7)	7	2.2	(1.1)	5	2.4	(0.9)	14	1.9	(1.2)	10
GIF_GBSI	0.4	(0.7)	10	1.7	(1.2)	11	1.9	(0.7)	12	2.2	(1.1)	5	2.4	(0.9)	14	2.1	(0.8)	10

* Percentage of data points dropped from the original full retrospective cohort for each measure to ensure the same set of images used for comparing longitudinal techniques.

Table 3-4 continued: Mean and standard deviation (sd) of annualised rates of change measures for controls and the clinical, genetic and pathology FTD subgroups for the comparison subset FTD cohort (n = 226)

n = 226	Genetic subgroups									Pathology subgroups					
	MAPT (n = 14 of 16)			*C9orf72* (n = 9 of 10)			*GRN* (n = 8 of 8)			Tau (n = 17 of 19)			TDP-43 (n = 32 of 34)		
Measure	**mean**	**(sd)**	**% dropped**	**mean**	**(sd)**	**% dropped**	**mean**	**(sd)**	**% dropped**	**mean**	**(sd)**	**% dropped**	**mean**	**(sd)**	**% dropped**
SPM_BV	2.5	(1.4)	12.5	2.1	(3.0)	10.0	4.2	(3.7)	0.0	2.7	(1.4)	10.5	3	(2.5)	5.9
SPM_KBSI	1.9	(0.9)	6.7	1.7	(1.5)	0.0	3.7	(2.0)	0.0	2	(0.9)	5.6	2.5	(1.5)	0.0
SPM_GBSI	2.1	(1.0)	6.7	1.7	(1.4)	10.0	3.9	(2.1)	0.0	2.2	(1.0)	5.6	2.6	(1.6)	3.0
Long_SPM_PBVC	1.1	(0.8)	12.5	1.2	(0.8)	10.0	0.6	(1.2)	0.0	1.1	(0.7)	10.5	0.9	(0.9)	5.9
BMAPS_BV	2.1	(1.0)	12.5	1.9	(1.5)	10.0	3	(2.6)	0.0	2.3	(1.0)	10.5	2.5	(1.6)	5.9
BMAPS_KBSI	2	(1.0)	12.5	1.8	(1.3)	10.0	3.7	(1.8)	0.0	2.1	(1.0)	10.5	2.5	(1.4)	3.0
MALP-EM_BV	0.5	(1.0)	12.5	0.7	(1.7)	10.0	-0.7	(4.0)	0.0	0.7	(1.1)	10.5	0.4	(2.3)	5.9
MALP-EM_KBSI	1.3	(0.7)	12.5	1.2	(0.7)	10.0	2.6	(1.1)	0.0	1.3	(0.6)	10.5	1.7	(0.9)	3.0
MALP-EM_GBSI	1.5	(0.9)	12.5	1.3	(0.9)	10.0	3	(1.3)	0.0	1.6	(0.8)	10.5	1.9	(1.1)	3.0
SIENAX_BV	0.8	(2.9)	12.5	1.3	(1.5)	10.0	1.5	(1.9)	0.0	1	(2.7)	10.5	1.5	(1.7)	5.9
SIENAX_KBSI	1.7	(1.0)	12.5	1.7	(1.2)	10.0	3.2	(1.6)	0.0	1.8	(0.9)	10.5	2.2	(1.2)	5.9
SIENA_PBVC	2.4	(1.2)	12.5	2.1	(1.1)	10.0	4.4	(2.4)	0.0	2.5	(1.2)	10.5	2.9	(1.8)	5.9
Freesurfer_BV	1.6	(1.7)	12.5	2.5	(3.1)	10.0	3.7	(1.7)	0.0	2.1	(2.0)	10.5	2.9	(3.2)	5.9
Freesurfer_KBSI	1.8	(0.9)	12.5	1.6	(1.1)	10.0	2.8	(2.0)	0.0	1.9	(0.9)	10.5	2.1	(1.4)	5.9
GIF_BV	1.1	(0.6)	12.5	1.3	(1.0)	10.0	1.2	(1.8)	0.0	1.2	(0.5)	10.5	1.3	(1.1)	5.9
GIF_KBSI	1.7	(0.8)	12.5	1.5	(1.1)	10.0	3.1	(1.5)	0.0	1.8	(0.8)	10.5	2.1	(1.2)	3.0
GIF_GBSI	1.7	(0.8)	6.7	1.5	(1.1)	10.0	3.1	(1.4)	0.0	1.8	(0.8)	5.6	2.1	(1.2)	5.9

3.3 Statistical rationale and analysis

3.3.1 Assumptions underlying clinical trial design

For the key sample size calculations in this and the following three chapters, it is assumed that the proposed clinical trial will be of a disease-modifying treatment, which means that the effect of treatment is assumed to slow the rate of disease progression. It is further assumed that a completely effective treatment (stopping disease progression entirely) would slow the atrophy rate to that seen in controls. This assumption was made because it would not be reasonable to expect that any treatment could slow the rate of decline in FTD patients to less than the rate of decline seen in controls, so the sample size calculation is corrected for the corresponding control rate of change in each measure. It was assumed that the trial would have 1:1 randomisation into control and active treatment groups.

3.3.2 Sample size calculations

Sample sizes are estimated for a hypothetical trial design where two scans are taken, one at the beginning of the trial and one at the end to produce a single measure of brain volume change as the outcome, with a simple t-test to compare means.

The sample size per arm is:

$$n = \frac{2\sigma^2}{\delta^2} f(\alpha, \beta)$$

where

σ is the standard deviation of the annualised rate of change in the outcome measure in the patient subgroups,

δ is the treatment effect (which consists of the difference in the rate of change for treated and untreated at the follow-up visit),

α is the significance level ($p = 0.05$),

1 – β is the power to detect a treatment effect (80%),

$$f(\alpha,\beta) = \left(z_{1-\alpha/2} + z_{1-\beta}\right)^2.$$

For example, α = 0.05 (conventional 5% significance level), β = 0.2 (power 80%), gives $f(\alpha,\beta)$ = 7.85.

It is assumed that treatment would lead to a 30% reduction in disease progression (as measured using annualised atrophy rate) in those on the treatment arm versus those on placebo, and this, in turn, would reduce the difference in atrophy rate compared with controls by 30%.

Therefore, the assumed treatment effect is:

$$\delta = 0.30(\mu_1 - \mu_0)$$

where

μ_1 is the mean atrophy rate in the patient group,

μ_0 is the mean atrophy rate in the control group.

3.3.3 Confidence interval calculation

Bias-corrected and accelerated (BCa) confidence intervals (CIs) for the sample size and ratios of sample sizes were calculated using bootstrapping sampling with replacement, with 2000 bootstrap samples stratified by diagnosis, genetic status and pathology patient subgroups (Carpenter & Bithell, 2000). The upper and lower limits for the confidence intervals for the sample size and ratios of sample sizes were calculated from the upper and lower limits of the confidence interval for the effect size and natural log of the ratio of effect sizes, respectively. This was done to take into account uncertainty in the estimates of both σ and δ, and because it was anticipated that the bootstrap distribution for the effect size and log of the ratio of effect sizes would be approximately normal and therefore confidence intervals calculated on this scale would have good coverage.

The effect size is a measure of the difference between patients and controls, relative to the standard deviation of each of the 17 whole-brain rates of change outcomes reported in this chapter.

$$ES = \frac{\mu_1 - \mu_0}{\sigma}$$

The effect size was converted into estimates for the sample size (per arm) by noting that:

$$n = \frac{2\sigma^2}{\delta^2} f(\alpha, \beta) = \frac{2\sigma^2}{(0.30(\mu_1 - \mu_0))^2} f(\alpha, \beta) = 2\frac{1}{(0.30ES)^2} f(\alpha, \beta)$$

From the above formula, it can be seen that the sample depends on significance level (α), power (1 – β), size of the relative reduction in atrophy compared with controls (30% used in this thesis) and the effect size. This was applied to the full FTD retrospective cohort (n = 262).

The ratio of effect sizes can be converted into the ratio of sample size by noting that:

$$\frac{n_1}{n_2} = \frac{2\sigma_1^2/\delta_1^2 f(\alpha, \beta)}{2\sigma_2^2/\delta_2^2 f(\alpha, \beta)} = \frac{2\sigma_1^2/(0.30(\mu_{1,1} - \mu_{1,0}))^2 f(\alpha, \beta)}{2\sigma_2^2/(0.30(\mu_{2,1} - \mu_{2,0}))^2 f(\alpha, \beta)}$$
$$= \frac{\sigma_1^2/(\mu_{1,1} - \mu_{1,0})^2}{\sigma_2^2/(\mu_{2,1} - \mu_{2,0})^2} = \frac{1/ES_1^2}{1/ES_2^2} = \frac{ES_2^2}{ES_1^2},$$

where

$\mu_{i,1}$ is the mean atrophy rate in the patient group for the ith technique (from 1 to 17 in this chapter),

$\mu_{i,0}$ is the mean atrophy rate in the control group for the ith technique (from 1 to 17),

σ_i is the standard deviation of atrophy rate in the patient group for the ith technique,

n_i is the sample size for a trial using the ith technique as an outcome measure

ES_i is the effect size for the ith technique.

From the above formula, it can be seen that the ratio of sample sizes does not depend on significance level (α), power (1 – β) or size of the relative reduction in atrophy compared with controls (30% in the above formula). It is a function of the relative effect sizes when using different techniques to measure the atrophy rate. This ratio of sample size calculation was applied to the comparison subset FTD cohort (n = 226) to perform a direct head-to-head paired comparison of each longitudinal measure with each of the other 16 measures separately within each of the 10 FTD clinical, genetic and pathology subgroups.

Linear regression analyses were also performed for each measure comparing all FTD subgroups with controls, correcting for age, gender and scanner type.

3.4 Results

3.4.1 Annual rates of whole-brain change and regression results

Mean and sd of annualised whole-brain changes for each of the 17 longitudinal measures are given in Table 3-3 for the full retrospective cohort (n = 262) and Table 3-4 for the comparison subset (n = 226). Figure 3-2 shows these mean rates of change with 95% confidence intervals for controls and patients across all longitudinal measures split by FTD subgroup (n = 262). Note the extended x-axis for the *GRN* subgroup to accommodate the higher annual rates and wider BCa confidence intervals. These demonstrate that there was generally good group separation demonstrated for all longitudinal measures, with the exception for the indirect measures MALP-EM_BV and Freesufer_BV for the majority of clinical, genetic and pathological subgroups. Because of the larger cohort numbers in the clinical groups, this was more evident than the genetic or pathology subgroups, which presented a wider range of confidence intervals for all techniques.

Control rates of change (blue in Figure 3-2) were relatively homogeneous across the measures with a mean annualised rate of change of 0.4% (0.8). The upper and lower limits of the range of values extended from MALP-EM_BV with 0.0% (1.8) to Freesurfer_BV with 0.7% (1.3). The high standard deviations relative to mean change in these measures speak to the increased noise inherent not only generally in indirect measures compared with direct measures but also the segmentation accuracy itself for these two techniques specifically. Applying the direct BSI measures to both segmentation methods substantially improved the mean-to-sd ratio, providing a more robust and consistent measure of change.

In contrast to the controls, annual rates of change derived from the 17 methods showed more heterogeneity within and between each patient subgroup. The full summary of rates can be reviewed in Table 3-3 and are most clearly demonstrated by Figure 3-2. In the clinical subgroups, mean rate of annual

atrophy for the bvFTD subgroup was the lowest overall at 1.7 (1.7)% and demonstrated the highest sd relative to mean. SvPPA mean rates were 1.9 (1.2)%, nfvPPA showed a mean rate of 2.3 (1.5)%, PPA-NOS showed a 2.4 (0.9)% mean annual change and the lvPPA subgroup showed a mean rate of 2.0 (1.6)%. When considering the range of values provided by each measure, MALP-EM_BV consistently provided the lowest annual rates of change, with either SPM_BV or SIENA_PBVC reporting the highest raw mean rates of change.

For the genetic subgroups, *MAPT* mean annual rate derived across all methods was 1.5 (1.2)%, *C9orf72* mean rate was 1.7 (1.4)% and *GRN* showed a 3.0 (2.0)% mean annual change. Mean rates within the pathology subgroups reflected the contributions of the genetic subgroup results with the tau subgroup presenting with a mean rate of 1.6 (1.2)%, similar to the *MAPT* results. TDP-43 reported a higher mean with 2.0 (1.6)%, reflecting the combined contribution of the *C9orf72* and *GRN* individuals along with the sporadic patients who had come to post-mortem. Again, MALP-EM_BV universally underestimated the annual rates of change and either SPM_BV or SIENA_PBVC consistently provided the highest mean rate of change.

Overall, the measures with the lowest mean-to-sd ratio were the indirect MALP-EM_BV, Freesurfer_BV and SIENAX_BV measures, which were under 0.5 across the majority of patient subgroups. Measures with the highest mean-to-sd ratio varied across subgroup but were all direct measures including BMAPS_KBSI, GIF_GBSI, SPM_GBSI and MALP-EM_GBSI, ranging from 1.5:1 to 3.3:1.

Table 3-5 reports the linear regression analysis results with 95% CI and accompanying regression coefficient (Coef.) demonstrating the additional increase (adjusted mean difference) in the annual rate of atrophy in the FTD subgroups as measured by each method compared with controls, when correcting for age, gender and scanner type. These analyses revealed significant differences between all patient subgroups and controls for the

majority of measures. Again, given the higher variability of segmentation accuracy and lower mean rates relative to sd, the MALP-EM_BV, SIENAX_BV and Freesurfer_BV comparisons did not consistently reach $p < 0.001$ across all subgroups. In contrast, all indirect measures of annual volume change for SPM, BMAPS and GIF segmentations were significantly different from control rates at the $p < 0.001$ level, except SPM_BV ($p < 0.005$ for *C9orf72* subgroup only). All direct BSI and PBVC measures reached $p < 0.001$ significance level across all patient subgroups, except for Long_SPM_PBVC ($p < 0.073$ for the *GRN* subgroup only).

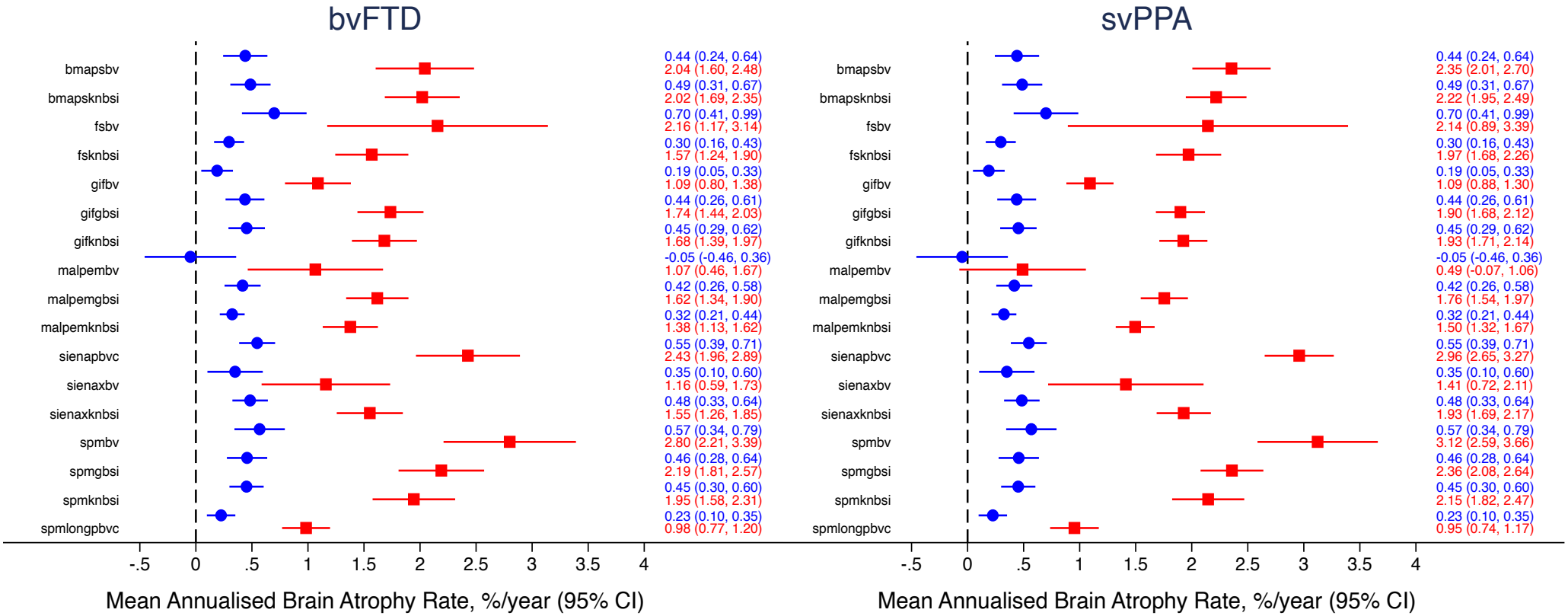

Figure 3-2 Mean and 95% confidence interval for annual whole-brain atrophy rate for all longitudinal measures for controls (blue) and each FTD subgroup

- fsbv = Freesurfer_BV

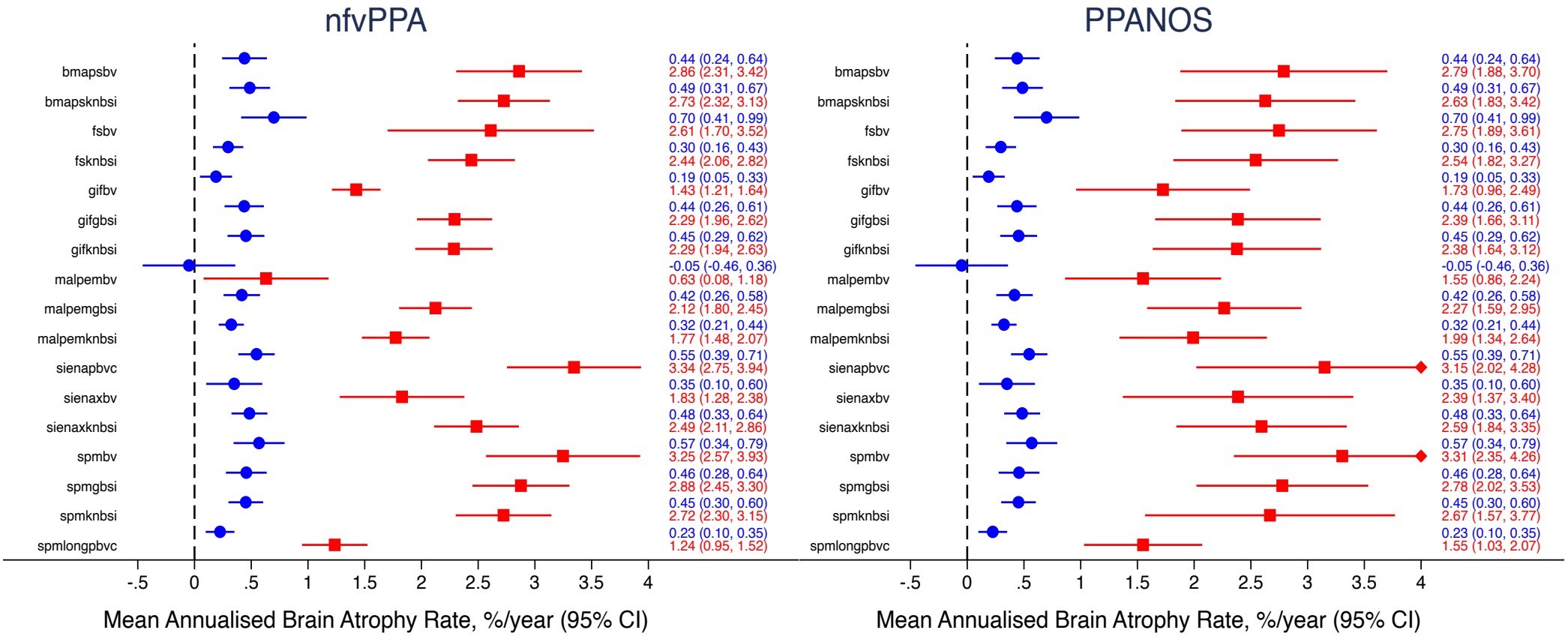

Figure 3-2 continued: Mean and 95% confidence interval for annual whole-brain atrophy rate for all longitudinal measures for controls (blue) and each FTD subgroup.

♦ represents a capped value where the 95% limit lies outside that chosen for the x-axis range. The actual value is listed on the right of the graph.

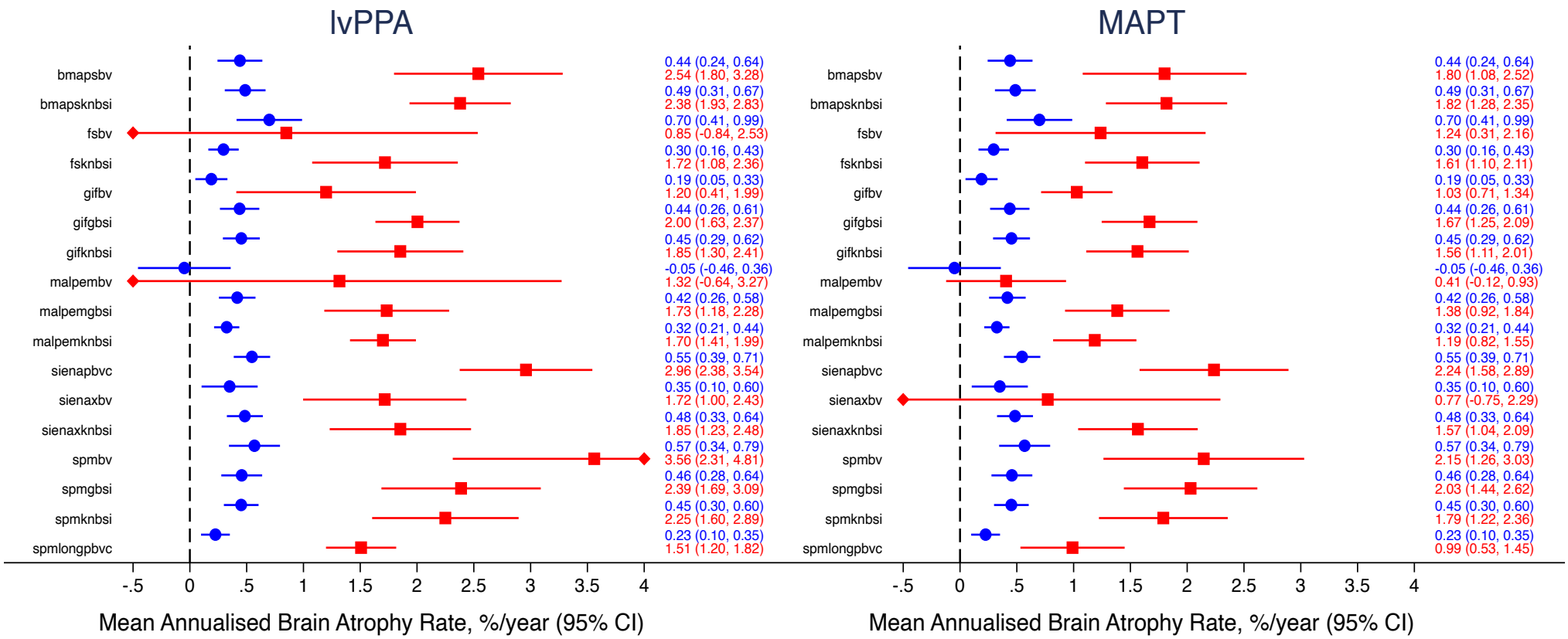

Figure 3-2 continued: Mean and 95% confidence interval for annual whole-brain atrophy rate for all longitudinal measures for controls (blue) and each FTD subgroup.

♦ represents a capped value where the 95% limit lies outside that chosen for the x-axis range. The actual value is listed on the right of the graph.

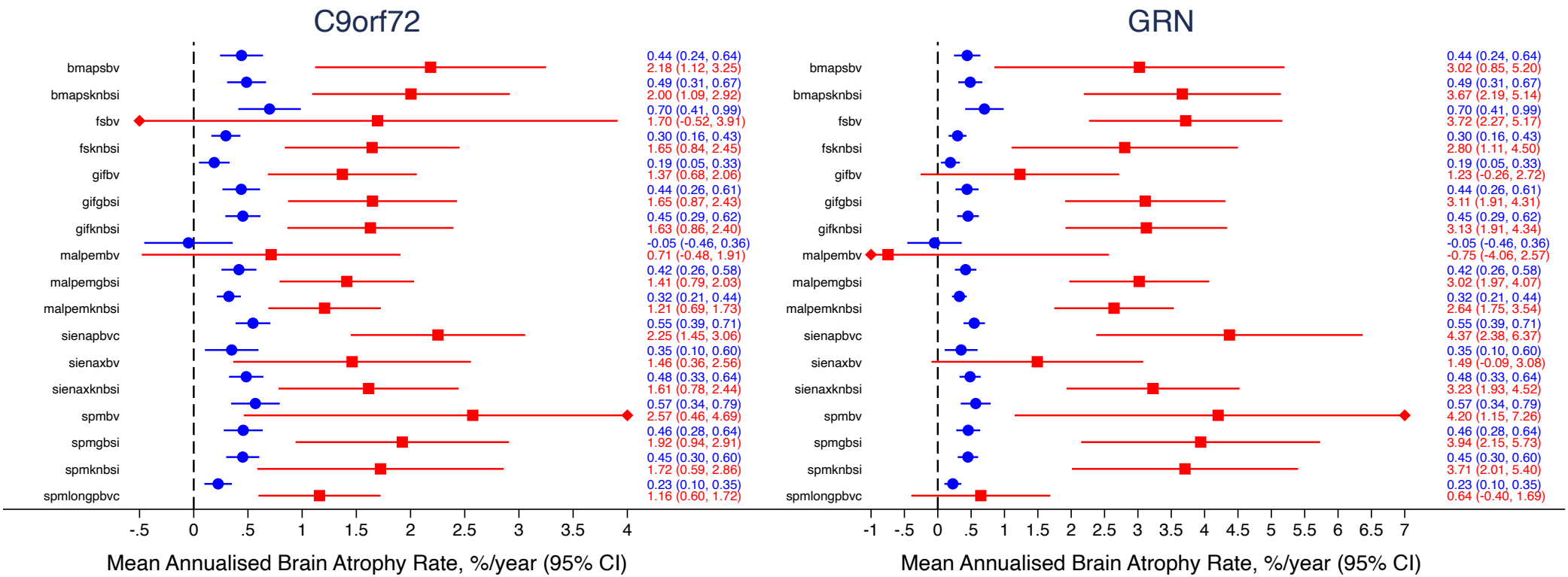

Figure 3-2 continued: Mean and 95% confidence interval for annual whole-brain atrophy rate for all longitudinal measures for controls (blue) and each FTD subgroup.

♦ represents a capped value where the 95% limit lies outside that chosen for the x-axis range. The actual value is listed on the right of the graph.

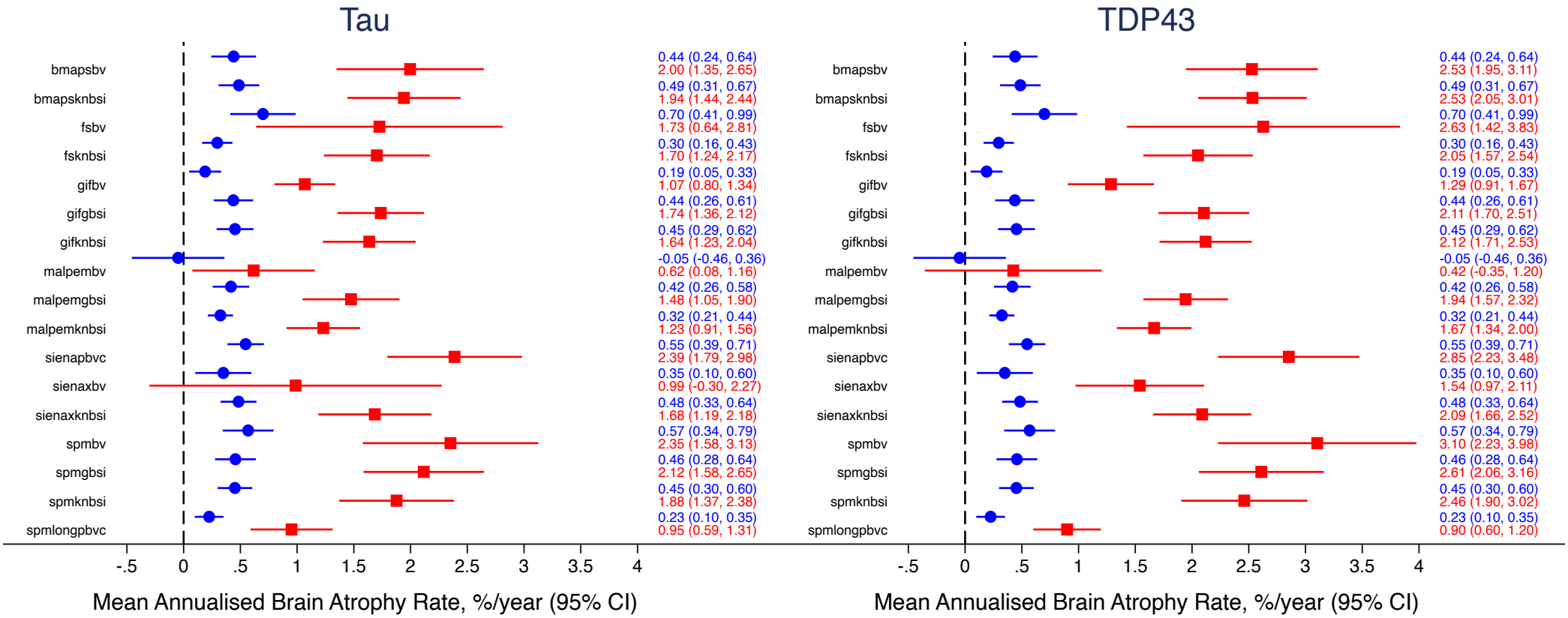

Figure 3-2 continued: Mean and 95% confidence interval for annual whole-brain atrophy rate for all longitudinal measures for controls (blue) and each FTD subgroup.

Table 3-5 Linear regression analysis results reporting the regression coefficient [95% CI] for clinical, genetic and pathology FTD subgroups for all longitudinal annual whole-brain measures of change. *

n = 262	Clinical Subgroups									
	bvFTD		svPPA		nfvPPA		PPA-NOS		lvPPA	
Measure	Coef. [95% CI]	*p*-value	Coef. [95% CI]	*p*-value	Coef. [95% CI]	*p*-value	Coef. [95% CI]	*p*-value	Coef. [95% CI]	*p*-value
SPM_BV	2.2 [1.6–2.9]	**<0.001**	2.6 [1.8–3.3]	**<0.001**	2.6 [1.9–3.4]	**<0.001**	2.7 [1.2–4.2]	*0.001*	3.0 [2.0–3.9]	**<0.001**
SPM_KBSI	1.5 [1.1–1.9]	**<0.001**	1.7 [1.2–2.1]	**<0.001**	2.3 [1.8–2.7]	**<0.001**	2.2 [1.2–3.2]	**<0.001**	1.9 [1.2–2.5]	**<0.001**
SPM_GBSI	1.8 [1.4–2.1]	**<0.001**	1.9 [1.4–2.4]	**<0.001**	2.4 [2.0–2.9]	**<0.001**	2.4 [1.4–3.3]	**<0.001**	2.0 [1.4–2.6]	**<0.001**
Long_SPM_PBVC	0.8 [0.5–1.0]	**<0.001**	0.8 [0.5–1.0]	**<0.001**	1.0 [0.7–1.3]	**<0.001**	1.2 [0.7–1.8]	**<0.001**	1.1 [0.8–1.5]	**<0.001**
BMAPS_BV	1.6 [1.2–2.1]	**<0.001**	1.9 [1.4–2.4]	**<0.001**	2.4 [1.9–3.0]	**<0.001**	2.4 [1.3–3.5]	**<0.001**	2.2 [1.5–3.0]	**<0.001**
BMAPS_KBSI	1.6 [1.2–2.0]	**<0.001**	1.7 [1.3–2.2]	**<0.001**	2.3 [1.8–2.7]	**<0.001**	2.2 [1.3–3.0]	**<0.001**	2.0 [1.4–2.5]	**<0.001**
MALP-EM_BV	1.1 [0.3–1.8]	*0.008*	0.5 [-0.3–1.4]	0.216	0.7 [-0.2–1.5]	0.118	1.6 [-0.2–3.4]	0.083	1.4 [0.3–2.5]	*0.016*
MALP-EM_KBSI	1.1 [0.8–1.4]	**<0.001**	1.2 [0.9–1.5]	**<0.001**	1.5 [1.2–1.8]	**<0.001**	1.7 [1.1–2.3]	**<0.001**	1.4 [1.1–1.8]	**<0.001**
MALP-EM_GBSI	1.3 [1.0–1.6]	**<0.001**	1.4 [1.0–1.7]	**<0.001**	1.7 [1.4–2.1]	**<0.001**	1.9 [1.2–2.9]	**<0.001**	1.4 [0.9–1.9]	**<0.001**
SIENAX_BV	0.9 [0.3–1.5]	*0.006*	1.1 [0.4–1.8]	*0.002*	1.5 [0.8–2.2]	**<0.001**	2.0 [0.6–3.5]	*0.007*	1.4 [0.4–2.3]	*0.004*
SIENAX_KBSI	1.1 [0.7–1.4]	**<0.001**	1.5 [1.1–1.9]	**<0.001**	2.0 [1.6–2.4]	**<0.001**	2.1 [1.3–2.9]	**<0.001**	1.3 [0.8–1.8]	**<0.001**
SIENA_PBVC	2.0 [1.5–2.5]	**<0.001**	2.4 [1.9–2.9]	**<0.001**	2.8 [2.3–3.4]	**<0.001**	2.7 [1.6–3.8]	**<0.001**	2.6 [1.9–3.3]	**<0.001**
Freesurfer_BV	1.2 [0.2–2.3]	*0.024*	1.5 [0.4–2.7]	*0.010*	1.8 [0.6–3.0]	*0.002*	1.7 [-0.7–4.1]	0.171	-0.3 [-1.9–1.2]	0.669
Freesurfer_KBSI	1.3 [0.9–1.7]	**<0.001**	1.7 [1.3–2.1]	**<0.001**	2.2 [1.8–2.6]	**<0.001**	2.3 [1.4–3.1]	**<0.001**	1.5 [0.9–2.0]	**<0.001**
GIF_BV	0.8 [0.5–1.2]	**<0.001**	0.9 [0.5–1.3]	**<0.001**	1.2 [0.9–1.6]	**<0.001**	1.5 [0.8–2.3]	**<0.001**	1.0 [0.5–1.5]	**<0.001**
GIF_KBSI	1.3 [0.9–1.6]	**<0.001**	1.5 [1.1–1.9]	**<0.001**	1.8 [1.5–2.2]	**<0.001**	1.9 [1.2–2.7]	**<0.001**	1.4 [0.9–1.9]	**<0.001**
GIF_GBSI	1.3 [1.0–1.7]	**<0.001**	1.5 [1.1–1.8]	**<0.001**	1.9 [1.5–2.2]	**<0.001**	2.0 [1.2–2.7]	**<0.001**	1.6 [1.1–2.1]	**<0.001**

* All analyses adjusted for age, gender and scanner type: ***p* < 0.001 bolded**, *p < 0.05 italicised*

Table 3-5 continued: Linear regression analysis results with 95% CI and regression coefficient for clinical, genetic and pathology FTD subgroups for all longitudinal annual whole-brain measures of change. *

n = 262	Genetic Subgroups						Pathology Subgroups			
	MAPT		*C9ORF72*		*GRN*		Tau		TDP-43	
Measure	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**
SPM_BV	1.8 [0.9–2.7]	**<0.001**	2.1 [1.0–3.3]	*0.005*	3.6 [2.4–4.9]	**<0.001**	2.1 [1.3–2.9]	**<0.001**	2.6 [1.9–3.3]	**<0.001**
SPM_KBSI	1.4 [0.9–2.0]	**<0.001**	1.3 [0.6–2.0]	*0.001*	3.3 [2.6–4.0]	**<0.001**	1.6 [1.0–2.1]	**<0.001**	2.1 [1.6–2.5]	**<0.001**
SPM_GBSI	1.7 [1.1–2.3]	**<0.001**	1.5 [0.8–2.2]	**<0.001**	3.5 [2.7–4.3]	**<0.001**	1.8 [1.2–2.4]	**<0.001**	2.2 [1.8–2.7]	**<0.001**
Long_SPM_PBVC	0.9 [0.5–1.3]	**<0.001**	0.9 [0.4–1.3]	**<0.001**	0.4 [0.0–0.9]	0.073	0.8 [0.5–1.2]	**<0.001**	0.7 [0.4–0.9]	**<0.001**
BMAPS_BV	1.6 [0.9–2.3]	**<0.001**	1.8 [1.0–2.6]	**<0.001**	2.6 [1.7–3.5]	**<0.001**	1.8 [1.2–2.4]	**<0.001**	2.1 [1.6–2.6]	**<0.001**
BMAPS_KBSI	1.4 [0.9–2.0]	**<0.001**	1.5 [0.9–2.2]	**<0.001**	3.2 [2.5–3.9]	**<0.001**	1.6 [1.1–2.2]	**<0.001**	2.1 [1.7–2.5]	**<0.001**
MALP-EM_BV	0.5 [-0.6–1.6]	0.389	0.8 [-0.5–2.2]	0.224	-0.8 [-2.2–0.6]	0.270	0.6 [-0.4–1.6]	0.230	0.4 [-0.4–1.2]	0.305
MALP-EM_KBSI	0.9 [0.6–1.3]	**<0.001**	0.9 [0.5–1.4]	**<0.001**	2.3 [1.9–2.8]	**<0.001**	1.0 [0.7–1.3]	**<0.001**	1.4 [1.1–1.7]	**<0.001**
MALP-EM_GBSI	1.0 [0.6–1.5]	**<0.001**	1.0 [0.5–1.6]	**<0.001**	2.6 [2.1–3.2]	**<0.001**	1.2 [0.7–1.6]	**<0.001**	1.6 [1.3–2.0]	**<0.001**
SIENAX_BV	0.4 [-0.5–1.3]	0.363	0.9 [-0.1–1.9]	0.092	1.3 [0.2–2.4]	*0.021*	0.7 [-0.1–1.5]	0.101	1.2 [0.6–1.8]	**<0.001**
SIENAX_KBSI	1.1 [0.7–1.6]	**<0.001**	1.1 [0.5–1.7]	**<0.001**	2.8 [2.2–3.4]	**<0.001**	1.3 [0.8–1.8]	**<0.001**	1.6 [1.3–2.0]	**<0.001**
SIENA_PBVC	1.9 [1.3–2.4]	**<0.001**	1.8 [1.1–2.5]	**<0.001**	3.8 [3.2–4.6]	**<0.001**	2.0 [1.4–2.6]	**<0.001**	2.3 [1.9–2.8]	**<0.001**
Freesurfer_BV	0.6 [-0.3–1.4]	0.179	0.6 [-0.4–1.7]	**0.206**	3.2 [2.1–4.3]	**<0.001**	1.2 [0.1–2.3]	*0.040*	2.0 [1.1–2.9]	**<0.001**
Freesurfer_KBSI	1.4 [0.9–1.9]	**<0.001**	1.4 [0.8–2.0]	**<0.001**	2.5 [1.8–3.1]	**<0.001**	1.5 [1.0–1.9]	**<0.001**	1.8 [1.4–2.2]	**<0.001**
GIF_BV	0.9 [0.4–1.3]	**<0.001**	1.2 [0.6–1.7]	**<0.001**	1.1 [0.5–1.6]	*0.001*	0.9 [0.5–1.3]	**<0.001**	1.1 [0.8–1.4]	**<0.001**
GIF_KBSI	1.2 [0.8–1.7]	**<0.001**	1.2 [0.6–1.7]	**<0.001**	2.7 [2.1–3.3]	**<0.001**	1.3 [0.9–1.8]	**<0.001**	1.7 [1.4–2.1]	**<0.001**
GIF_GBSI	1.3 [0.8–1.8]	**<0.001**	1.2 [0.6–1.8]	**<0.001**	2.7 [2.1–3.3]	**<0.001**	1.4 [0.9–1.9]	**<0.001**	1.7 [1.3–2.1]	**<0.001**

* All analyses adjusted for age, gender and scanner type: ***p* < 0.001 bolded**, *p < 0.05 italicised*

3.4.2 Effect size and sample size results

Table 3-6 reports the effect size calculations for each of the 17 techniques per patient subgroup. The 95% BCa confidence intervals are also included. Effect sizes over one are presented in bold because these are within the range considered as representing a large effect as based on Cohen's (1988) original formulation. For these bolded results, at least 84% of the control group atrophy rates were below the average rate in the corresponding FTD patient subgroup for that measure, suggesting good group separation. The more robust direct measures of change almost all produced effect sizes greater than one across the clinical, genetic and pathology subgroups with only a few exceptions. In line with the annual rates of change used to calculate these, the SPM_BV, BMAPS_BV and GIF_BV were the indirect measures that performed the best with higher effect sizes in general compared with the other indirect measures. Fewer measures for the more heterogeneous bvFTD subgroup produced effect sizes greater than one, with only 8 of the 17 techniques compared with a much higher number for the more homogeneous PPA subgroups. This is particularly true for the PPA-NOS subgroup, with high effect sizes across all 17 longitudinal measures, ranging from 1.9 [1.0–3.1] for SIENAX_BV to 2.9 [2.0–3.8] for Freesurfer_KBSI.

Table 3-7 reports the sample size estimates required to detect a 30% reduction in annual atrophy rate relative to control rates, assuming the trial will have 80% statistical power and a 5% significance level to detect the treatment effect. These were calculated based on results from the full cohort of 262 subjects to include the maximum amount of passing data points for the estimation. The colour-coded comparison results are based on the subset cohort of 226 subjects who all had a complete dataset across measures with no missing values for any of the techniques to allow a direct head-to-head comparison. The results in blue and bold highlight the longitudinal technique that performed the best, i.e. resulted in the lowest sample size estimate for each of the patient subgroups. Results highlighted in green and underlined did not significantly differ from this lowest estimate. Those in black produced significantly higher

sample sizes ($p < 0.05$) than the best longitudinal technique but were significantly lower than those highlighted in red, which produced the highest estimates or were unable to be calculated using the bootstrapping methodology because multiple samples for those methods produced annual rates of change that did not significantly differ from the control group.

The best performing longitudinal technique differed between the clinical, genetic and pathological subgroups; however, in all cases, they were not statistically significantly different from several other longitudinal measures (denoted by those in green and underlined in Table 3-7). Sample size estimates were low and feasible across all patient subgroups with PPA-NOS providing the lowest at 25 [95% CI (11–44)] patients required per treatment arm using SPM_GBSI. All 17 measures provided consistently low estimates for this patient subgroup.

The methods providing the lowest effect size and highest sample size estimates, or where the upper 95% CI limit could not be calculated using the bootstrapping methodology, were almost exclusively indirect measures, particularly using MALP-EM, SIENAX and Freesurfer segmentations. In addition, the direct measures Long_SPM_PBVC failed to determine an upper limit estimate for lvPPA, *GRN* and TDP-43 subgroups and SIENA_PBVC also produced an infinite upper limit for lvPPA. Overall, the BSI measures provided the lowest estimate or were not significantly different from the lowest estimate and the indirect measures for BMAPS, SPM and GIF provided significantly lower estimates than the other indirect measures.

Table 3-6 Effect size calculations with 95% BCa confidence intervals for clinical, genetic and pathology FTD subgroups *

Measure	Clinical Subgroups									
	bvFTD		svPPA		nfvPPA		PPA-NOS		lvPPA	
n=262	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)
SPM_BV	0.9	[0.7–1.2]	**1.4**	[1.0–1.9]	**1.2**	[0.8–1.6]	**2.6**	[1.6–4.0]	**1.1**	[0.3–1.6]
SPM_KBSI	**1.0**	[0.8–1.7]	**1.6**	[1.0–2.1]	**1.6**	[1.3–2.1]	**2.1**	[1.6–3.2]	**1.3**	[0.3–2.5]
SPM_GBSI	**1.1**	[0.8–1.4]	**2.1**	[1.6–2.6]	**1.7**	[1.3–2.3]	**2.8**	[1.9–4.0]	**1.3**	[0.3–2.4]
Long_SPM_PBVC	0.9	[0.6–1.1]	**1.0**	[0.6–1.4]	**1.1**	[0.5–1.5]	**2.4**	[1.4–3.4]	**1.9**	[1.3–2.5]
BMAPS_BV	0.9	[0.6–1.2]	**1.6**	[1.2–2.1]	**1.3**	[1.0–1.8]	**2.4**	[1.6–3.9]	**1.3**	[0.8–1.7]
BMAPS_KBSI	**1.1**	[0.9–1.4]	**2.0**	[1.5–2.5]	**1.7**	[1.3–2.1]	**2.5**	[1.7–3.5]	**2.0**	[1.4–2.7]
MALP-EM_BV	0.5	[0.1–0.8]	0.3	[-0.2–0.8]	0.4	[-0.1–0.9]	**2.1**	[1.2–3.5]	0.3	[-0.3–0.6]
MALP-EM_KBSI	**1.1**	[0.8 –1.3]	**2.0**	[1.5–2.6]	**1.5**	[1.0–2.0]	**2.4**	[1.7–3.5]	**2.2**	[1.7–2.8]
MALP-EM_GBSI	**1.1**	[0.8–1.4]	**1.9**	[1.4–2.4]	**1.6**	[1.2–2.1]	**2.5**	[1.6–3.7]	**1.2**	[0.1–2.3]
SIENAX_BV	0.3	[0.0–0.7]	0.5	[0.0–0.9]	0.8	[0.4–1.2]	**1.9**	[1.0–3.1]	0.9	[0.2–1.4]
SIENAX_KBSI	0.9	[0.6–1.2]	**1.9**	[1.3–2.5]	**1.6**	[1.2–2.1]	**2.6**	[1.7–3.8]	**1.0**	[0.1–1.8]
SIENA_PBVC	**1.0**	[0.7–1.2]	**2.4**	[1.9–2.8]	**1.4**	[1.0–1.8]	**2.1**	[1.5–3.0]	**1.9**	[1.2–2.5]
Freesurfer_BV	0.4	[0.1–0.6]	0.3	[-0.1–0.7]	0.6	[0.2–1.0]	**2.2**	[0.8–4.3]	0.0	[-0.5–0.5]
Freesurfer_KBSI	**1.0**	[0.7–1.2]	**1.8**	[1.3–2.2]	**1.7**	[1.3–2.1]	**2.9**	[2.0–3.8]	**1.0**	[0.2–1.9]
GIF_BV	0.8	[0.5–1.0]	**1.3**	[0.7–1.9]	**1.7**	[1.2–2.3]	**1.9**	[1.0–2.7]	0.6	[0.0–1.0]
GIF_KBSI	**1.1**	[0.8–1.3]	**2.2**	[1.6–2.7]	**1.6**	[1.2–2.1]	**2.4**	[1.6–3.7]	**1.2**	[0.1–2.4]
GIF_GBSI	**1.1**	[0.8–1.4]	**2.1**	[1.5–2.6]	**1.7**	[1.3–2.2]	**2.5**	[1.6–3.8]	**2.0**	[1.5–2.5]

* Effect sizes of one or above are in **bold** to indicate results demonstrate a high degree of group separation from controls for that subgroup and measure

Table 3-6 continued: Effect size calculations with 95% BCa confidence intervals for clinical, genetic and pathology FTD subgroups *

Measure	Genetic Subgroups						Pathology Subgroups			
	MAPT		*C9orf72*		*GRN*		Tau		TDP-43	
n=262	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)
SPM_BV	1.0	[0.3–1.6]	0.7	[0.1–1.3]	1.0	[0.6–1.4]	**1.1**	[0.5–1.9]	**1.0**	[0.6–1.4]
SPM_KBSI	**1.3**	[0.6–1.9]	0.9	[0.2–1.5]	**1.6**	[1.0–2.4]	**1.4**	[0.8–2.0]	**1.3**	[0.9–1.7]
SPM_GBSI	**1.5**	[0.9–2.2]	**1.1**	[0.5–1.8]	**1.6**	[1.0–2.4]	**1.6**	[1.0–2.3]	**1.4**	[0.9–1.8]
Long_SPM_PBVC	0.9	[0.2–1.5]	**1.2**	[0.5–2.0]	0.3	[-0.7–1.7]	1.0	[0.4–1.6]	0.8	[0.2–1.4]
BMAPS_BV	**1.0**	[-0.1–1.8]	**1.2**	[0.6–1.8]	1.0	[0.4–1.6]	**1.2**	[0.3–1.9]	**1.3**	[0.9–1.7]
BMAPS_KBSI	**1.3**	[0.7–1.9]	**1.2**	[0.5–2.0]	**1.8**	[1.0–2.6]	**1.4**	[0.8–2.0]	**1.5**	[1.1–2.0]
MALP-EM_BV	0.5	[-0.3–1.5]	0.5	[-0.4–2.0]	-0.2	[-1.0–0.8]	0.6	[-0.2–1.3]	0.2	[-0.2–0.8]
MALP-EM_KBSI	**1.3**	[0.8–1.8]	**1.2**	[0.6–1.9]	**2.2**	[1.3–3.3]	**1.3**	[0.9–1.9]	**1.4**	[1.0–1.8]
MALP-EM_GBSI	**1.1**	[0.6–1.7]	**1.1**	[0.6–1.9]	**2.1**	[1.2–3.2]	**1.2**	[0.7–1.8]	**1.5**	[1.0–1.9]
SIENAX_BV	0.1	[-0.5–1.3]	0.7	[-0.2–1.6]	0.6	[-0.3–1.8]	0.2	[-0.3–1.3]	0.7	[0.3–1.1]
SIENAX_KBSI	**1.1**	[0.3–1.6]	1.0	[0.3–1.6]	**1.8**	[0.8–3.1]	**1.2**	[0.5–1.7]	**1.3**	[0.9–1.7]
SIENA_PBVC	**1.4**	[0.7–2.0]	**1.5**	[0.7–2.8]	**1.6**	[0.8–2.4]	**1.5**	[0.8–2.2]	**1.3**	[0.9–1.7]
Freesurfer_BV	0.3	[-0.4–1.1]	0.3	[-0.6–1.5]	**1.7**	[1.0–2.9]	0.5	[-0.1–0.9]	0.6	[0.2–0.9]
Freesurfer_KBSI	**1.4**	[0.7–2.0]	**1.2**	[0.5–1.9]	**1.2**	[0.5–2.4]	**1.5**	[0.8–2.1]	**1.3**	[0.9–1.6]
GIF_BV	**1.4**	[0.9–2.0]	**1.2**	[0.5–2.4]	0.6	[-0.4–2.6]	**1.6**	[1.0–2.2]	**1.0**	[0.4–1.6]
GIF_KBSI	**1.3**	[0.7–1.9]	**1.1**	[0.4–2.0]	**1.8**	[1.1–2.8]	**1.4**	[0.8–2.0]	**1.5**	[1.0–1.9]
GIF_GBSI	**1.6**	[1.1–2.4]	**1.1**	[0.4–2.0]	**1.9**	[1.0–2.9]	**1.7**	[1.2–2.4]	**1.5**	[1.0–1.9]

* Effect sizes of one or above are in **bold** to indicate results demonstrate a high degree of group separation from controls for that subgroup and measure

Table 3-7 Sample size calculations to detect a 30% treatment effect, with 80% statistical power and a 5% significance level, with 95% BCa CIs. Results highlighted in **blue** produced the lowest sample size for each patient subgroup. Those in green and underlined did not significantly differ from the lowest estimate. Those in black produced significantly higher sample sizes than these lowest values but were significantly lower than those highlighted in red, which produced the highest estimates or were unable to be calculated using the bootstrapping methodology because multiple samples failed to find a significant difference between patient and control groups.

Measure	Clinical Subgroups									
	bvFTD (n=66)		svPPA (n=45)		nfvPPA (n=45)		PPA-NOS (n=7)		lvPPA (n=21)	
	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)
SPM_BV	135	[87–239]	48	[28–85]	123	[67–319]	27	[11–78]	146	[71–1914]
SPM_KBSI	138	[92–231]	42	[26–79]	66	[41–132]	38	[15–61]	103	[29–1767]
SPM_GBSI	121	[80–200]	32	[21–60]	60	[36–123]	**25**	**[11–44]**	105	[29–2657]
Long_SPM_PBVC	179	[106–340]	161	[76–704]	152	[76–973]	38	[17–96]	49	[29–101]
BMAPS_BV	209	[127–423]	64	[40–116]	76	[43–155]	31	[1–67]	105	[59–240]
BMAPS_KBSI	145	[88–291]	38	[23–80]	59	[40–111]	30	[12–54]	42	[25–85]
MALP-EM_BV	1643	[479– 4.1e+05]	23782	[502–∞]	5071	[475–∞]	38	[11–189]	1726	[487–∞]
MALP-EM_KBSI	133	[89–240]	34	[21–64]	70	[44–134]	34	[13–60]	**35**	**[23–60]**
MALP-EM_GBSI	**120**	**[78–208]**	37	[23–71]	62	[40–117]	29	[11–80]	131	[34–8830]
SIENAX_BV	962	[282–∞]	223	[126–516]	250	[115–1797]	44	[10–169]	233	[93–3301]
SIENAX_KBSI	152	[95–318]	40	[24–86]	64	[42–118]	26	[9–65]	174	[52–15855]
SIENA_PBVC	135	[86–225]	**28**	**[20–44]**	75	[45–140]	41	[18–72]	49	[28–111]
Freesurfer_BV	613	[296–3605]	2210	[286–∞]	420	[167–3522]	37	[9–255]	100303	[597–∞]
Freesurfer_KBSI	130	[84–234]	50	[30–96]	60	[39–102]	25	[13–46]	162	[46–3513]
GIF_BV	267	[151–830]	80	[40–346]	59	[33–116]	51	[18–165]	516	[153–4.2e+05]
GIF_KBSI	130	[82–235]	36	[23–67]	62	[39–115]	32	[11–62]	125	[30–9331]
GIF_GBSI	127	[79–226]	35	[22–68]	**58**	**[37–105]**	31	[10–61]	45	[28–82]

Table 3-7 continued: Sample size calculations to detect a 30% treatment effect, with 80% statistical power and a 5% significance level, with 95% BCa CIs.

Measure	Genetic Subgroups						Pathology Subgroups			
	MAPT (n=16)		*C9orf72* (n=10)		*GRN* (n=8)		Tau (n=19)		TDP-43 (n=34)	
	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)
SPM_BV	88	[38–337]	483	[128–∞]	167	[90–482]	68	[28–235]	164	[87–424]
SPM_KBSI	70	[37–155]	222	[82–2551]	66	[32–170]	63	[33–126]	99	[57–198]
SPM_GBSI	60	[28–135]	171	[64–919]	65	[32–176]	57	[27–117]	92	[52–190]
Long_SPM_PBVC	120	[50–572]	130	[45–946]	1443	[64–∞]	104	[44–502]	285	[103–7249]
BMAPS_BV	65	[34–147]	127	[49–536]	172	[66–1256]	55	[28–113]	106	[59–264]
BMAPS_KBSI	62	[27–133]	126	[46–817]	52	[25–165]	58	[27–122]	**75**	**[43–141]**
MALP-EM_BV	2975	[135–∞]	2602	[68–∞]	3021	[494–∞]	1010	[163–∞]	25752	[767–∞]
MALP-EM_KBSI	74	[36–218]	123	[53–425]	**35**	**[17–98]**	66	[32–170]	78	[48–136]
MALP-EM_GBSI	76	[33–242]	134	[58–460]	38	[17–111]	72	[34–203]	77	[45–138]
SIENAX_BV	5121	[63–∞]	366	[70–∞]	401	[53–∞]	2127	[57–∞]	296	[127–1205]
SIENAX_KBSI	80	[43–167]	151	[52–1427]	53	[19–202]	78	[44–165]	87	[53–179]
SIENA_PBVC	69	[30–368]	**77**	**[25–378]**	65	[28–223]	58	[26–253]	95	[57–201]
Freesurfer_BV	267	[61–∞]	121	[44–1133]	52	[18–137]	289	[113–6176]	311	[139–1788]
Freesurfer_KBSI	**55**	**[25–112]**	143	[61–1105]	113	[29–744]	53	[27–111]	107	[65–204]
GIF_BV	56	[30–131]	137	[35–1069]	501	[29–∞]	**45**	**[24–102]**	183	[79–1334]
GIF_KBSI	59	[25–123]	156	[55–1573]	49	[22–145]	55	[25–115]	79	[46–153]
GIF_GBSI	55	[25–113]	152	[55–1302]	48	[20–154]	50	[24–102]	76	[44–148]

3.5 Discussion

The current study applied six automated whole-brain segmentation techniques widely used in clinical research to a large comprehensive longitudinal FTD cohort and control comparison group. Indirect and direct measures of longitudinal change were calculated using segmentation volume subtraction, and the BSI and longitudinal PBVC pipelines, respectively. The annual rate of whole-brain volumetric change, expressed as a percentage of baseline volume, was calculated for these 17 longitudinal measures and compared within each of the 10 clinical, genetic and pathology FTD subgroups. Overall, segmentation accuracy using the six automated segmentation techniques was good (particularly for BMAPS, GIF and SPM techniques). The automated segmentation techniques rely on the accurate registration of atlases or templates and the unsegmented MR image to determine the spatial transformation of the correct labels onto the target image. On cross-sectional QC, all segmentation techniques performed well at doing this, with only 4 of the 3144 segmentations failing to label the brain region to a good approximation. These failed segmentations were produced by Freesurfer and failed due to considerable exclusion of the temporal lobes. This segmentation issue has previously been reported for Freesurfer in a large HD study (Johnson et al., 2017). Although differences in GM/dura inclusion and minor segmentation errors were noted across all techniques, manual editing was not performed to avoid any bias to the subset of images that would be chosen for such correction. In addition, the aim was to assess fully automated pipelines with the view that they could be applied by any site irrespective of whether an experienced image segmentor was available.

Despite an overall high level of segmentation quality cross-sectionally, the level of accuracy becomes much clearer when we look at their performance across both images in the pair. The variability in the indirect measures derived solely from these segmentation volumes demonstrates that even the smallest amount of additional dura included or brain matter excluded can have significant effects on the ability to detect underlying volumetric changes longitudinally. MALP-EM_BV consistently produced the lowest rate of annual

volumetric change (Table 3-3). This was because both baseline and repeat images had the largest overall volumes and had overestimated the GM boundary to include substantial regions of dura on both. This has also previously been reported in a population of HD patients and controls (Johnson et al., 2017). It is encouraging to note that the application of the BSI, even to the segmentations that were poorer quality, significantly reduced noise in the longitudinal measure as demonstrated by substantial increases in the mean-to-sd ratio. This resulted in increased effect sizes and considerably reduced sample size estimates to detect a treatment effect, with clear narrowing of the 95% BCa confidence intervals for these estimates. This means that ultimately the BSI needs to have a reasonable, but not necessarily very precise segmentation, in order to estimate where the atrophy is occurring to provide a robust measure of anatomical change between scans.

The number and scope of FTD patients included in the current cohort and the ability to perform a direct head-to-head comparison of techniques is a key strength of this study. This aimed to address the issue that, to date, considerable variability and inconsistency of rates of change results have likely arisen from small cohort numbers and application of techniques that could not be directly compared across studies. The direct comparison performed here showed significant differences in the performance of the techniques within each of the 10 FTD subgroups. As previously mentioned, it is possible that initial positive therapeutic effects may translate to small changes in the rate of atrophy. Given the variability of performance, it becomes clear that the correct choice of longitudinal technique will be crucial to improve the chance of detecting any such disease-modifying effect. For a trial enrolling any of the patient subgroups, application of the BSI (GBSI where possible) to BMAPS, GIF or SPM segmentations overall produced the fewest subjects needed to detect a treatment effect with subtle differences between groups (Table 3-7). Reassuringly, these were low across all subgroups, ranging from 25 [95% CI: 11–44] for PPA-NOS to 120 [78–208] for the more heterogeneous bvFTD group. In the genetic subgroups, *GRN* produced the lowest sample size with 35 [17–98] patients per treatment and *C9orf72* providing the highest but still

feasible estimate of 77 [25–378] patients required for enrolment per arm using the best method.

The current study reported sample sizes for whole-brain change that are relatively consistent with several of the previously reported studies in the clinical subgroups (Table 3-1). This was only for studies that also had larger patient cohorts and who also employed the direct BSI method. Compared with the studies investigating smaller cohorts, the current sample sizes are considerably lower across the clinical subgroups for the best performing method.

To date, only one study has reported global rates of change in the different genetic populations (Whitwell, Boeve et al., 2015). This team applied the BSI to SPM segmentations to derive sample size estimates. Reassuringly, the current study produced significantly lower estimates for *MAPT* and equivalent sample sizes for the *C9orf72* and *GRN* subgroups using this direct measure of change. Despite equivalent subject numbers, there were clear differences between the age of subjects included and disease duration compared with the current study, which potentially masked or overemphasised comparative differences in the rate of change measures. The issue of whole-brain atrophy time course and staging in FTD remains a matter that requires further investigation. Encouragingly, they report that rates of change at the subject level demonstrated trajectories that were approximately linear over the period of follow-up, but it is possible the angle of that linear trajectory may vary over the course of the disease at the group level. Overall, both this and the current study demonstrate the utility of direct automated longitudinal measures of whole-brain volume change as potential non-invasive markers for upcoming genetic FTD trials. This will be explored further in Chapter 6 using the GENFI genetic FTD cohort.

A key aim of this study was to assess the utility of these fully automated measures as potential biomarkers for secondary trial end-points. A key method of evaluation biomarker performance is the calculation of sample sizes, which

have performed well across all segmentation techniques and subgroups, with clear improvements when applying the direct BSI or PBVC measures over the indirect BV change method. When evaluating biomarker choice, the issue of ease of application is also an important one. This will be particularly relevant in multicentre trials of a rare disease such as FTD, where some recruitment sites may not have experienced image analysts to run these techniques and troubleshoot when pipelines fail to complete. One of the SPM segmentations failed to complete, and some additional options were required to ensure MALP-EM brain extraction produced the correct output, but it was the SIENAX segmentation pipeline that required substantial troubleshooting to resolve. The implemented solution in this particular dataset was to transform all images into MNI-125 space. This consideration of the ability to reliably, easily and repeatedly apply these techniques across multiple sites, scanners, acquisitions and available expertise suggests that SPM, GIF and BMAPS stand out as preferable automated segmentation techniques in this condition. However, the longitudinal SIENA pipeline did perform in native space without any errors or advanced settings and generally provided robust measures of change as well. The potential issue is that any analysis would have to wait until both images were acquired, which presents logistic hurdles in terms of QC and data analysis at baseline.

Another important issue in clinical trials is the attrition rate. Here, I provide sample sizes correcting for control rates of change but have not included a putative attrition rate because it has yet to be established what the retention rate in a large FTD trial might be. This, of course, would depend on the severity of patients included, the treatment type, likely adverse effects, the number of visits required (Grill & Karlawish, 2010; Grill et al., 2015), so the current results are likely to underestimate the final recruitment numbers required once accounting for any patient drop-out. An additional important caveat to this study is that I undertook QC of the images before inclusion into the cohort. This naturally biases the sample size estimates to be lower given a trial would recruit individuals who may not provide a usable scan that is adequate for analysis. Again, there are no published data on the prevalence of this pass

rate in FTD; however, UCL has conducted extensive neuroimaging in its longitudinal investigation of FTD (LIFTD) observational study, where approximately 16% of scans failed QC at baseline before analysis (based on almost 300 patients enrolled). Patients often returned to repeat this imaging assessment and acquired a successful baseline scan, but even taking that upper 16% adjustment into account, fully automated measures of whole-brain atrophy provide promising biomarkers for upcoming FTD trials with small and feasible sample sizes required per treatment arm across all clinical, genetic and pathology subgroups.

An additional caveat in the current study relates to the number of data points included per technique, as demonstrated in the longitudinal QC results. For the 2,344 registered pairs that I QCed, 84 registrations failed across all techniques, resulting in an overall loss of 3.3% of data points (Table 3-3). In contrast, none of the SIENA_PBVC and Long_SPM_PBVC pairs were failed for QC after reviewing the output images and reports. With the BSI QC, it is easy to toggle between the registered images, which makes issues like geometric distortion between scans and misregistration much easier to detect visually. It is much clearer how these issues are distorting the quantification of underlying volumetric change. It is important to note that I have had over 10 years' experience providing QC for registered T1 images produced from application of the BSI, so there is a possibility I am more attuned to how registration and quality differences impact the reliability of such measures. Subtle signs of geometric distortion when viewing the Long_SPM_PBVC deformation maps or SIENAX_PBVC outputs were less obvious than the method for reviewing registered pairs of images for the BSI measure, so there is the possibility that these measures included several noisier data points that were removed from the BSI measures. Despite this, it is reassuring that only 3.6% of data was excluded across all BSI measures given previous work by Hua et al., (2013) demonstrated that selective data exclusion results in a bias towards lower sample sizes. Conversely, the inclusion of scans with poorer quality such as those with motion artefacts has shown to produce higher sample sizes (Iscan et al., 2015) for fully automated pipelines. Given that only

one BSI pair failed due to motion and multiple pairs with motion artefacts were included, there is a high degree of confidence that the estimates reported here are reliable.

The results of this study confirm that fully automated methods for measuring longitudinal whole-brain atrophy perform well across all clinical, genetic and pathology subgroups. Whilst the best performing techniques vary across the different subgroups, there are multiple techniques for each that provide small and feasible sample sizes to detect a disease-modifying effect on atrophy rate. The direct head-to-head comparison of techniques in this large FTD cohort demonstrates the importance of the choice of technique depending on the patient population being enrolled or investigated, which will be of value in informing biomarker choice in the future. In general, direct measures of change including application of the BSI to SPM, GIF or BMAPS segmentation outperformed the other direct and all indirect measures of change. Many of these segmentation and longitudinal pipelines worked 'out of the box', requiring little or no additional optimisation, making them ideal for application in multicentre FTD trials, which may include sites that do not have advanced image analysis expertise.

4 Automated measures of lobar volume change in FTD

The previous chapter confirmed that fully automated methods of extracting whole-brain volume change provided sensitive measures of atrophy across the clinical, genetic and pathology-confirmed FTD subgroups, producing small feasible sample sizes required to detect a 30% treatment effect in a putative intervention trial in FTD. Having established that GIF provided consistently good segmentations and was one of the indirect longitudinal measures that outperformed many of the other indirect methods in Chapter 3, it was chosen as the segmentation technique to investigate lobar rates of change in the current chapter. An additional advantage of GIF is that it provides probabilistic masks to enable both GBSI and KBSI methods to be run and not only provides a brain mask as used in the previous chapter but extracts multiple neuroanatomical labels that can be combined into more focal regions such as lobar and subcortical structures (Appendix 1). Because the direct BSI longitudinal measures typically outperformed the other direct methods and consistently outperformed all indirect measures, the newly developed Pie Boundary Shift Integral (PieBSI) was chosen as the direct longitudinal technique to investigate lobar volume change measures.

4.1 Introduction

Characterised as a focal disease, it is possible that more targeted regional measurements across the FTD spectrum could provide a more robust and specific signal of change than the global measures investigated in Chapter 3. Cross-sectional and longitudinal neuroimaging findings of areas particularly vulnerable to volume loss suggest several candidate regions predominantly within the frontal, temporal, insular and cingulate cortices.

For bvFTD, the most prominent atrophy occurs primarily in the frontal and temporal lobes, particularly involving the prefrontal cortex, anterior temporal lobes, along with volume loss in the insula and anterior cingulate (Krueger et al., 2010; Schroeter et al., 2007). As discussed in the Introduction, this

frontotemporal-insula-anterior-cingulate involvement has been proposed to represent a structural and functional network that appears to be particularly vulnerable in bvFTD, although with the caveat that these structures may reflect varying combinations of the four distinct neuroanatomical clusters of bvFTD patients proposed by Whitwell and colleagues (2009). Therefore, I hypothesise that both left and right direct longitudinal measures of frontal and temporal lobe will perform well at detecting regional atrophy, providing the lowest sample size estimates for this subgroup.

SvPPA patients present with (often left-sided) asymmetric temporal lobe atrophy (Gorno-Tempini et al., 2004; Rohrer et al., 2009), although a right-dominant variant has also been reported (Chan et al., 2009; Josephs et al., 2009). As the disease progresses, atrophy extends anteriorly to include orbitofrontal, inferior frontal, insular and anterior cingulate cortices, and posteriorly to temporoparietal regions with increasing contralateral involvement (Schroeter et al., 2007). Despite the characteristic asymmetric presentation, cross-sectional data from multiple stages of the disease have shown bilateral volume reduction in the temporal lobes, even in the earliest stages of the disease (Bocchetta, Iglesias, Russell, et al., 2019). Longitudinally, temporal lobe rates of change are significantly higher than controls, with a suggestion of faster rates in the right temporal region, due to the significant atrophy that has already occurred on the left by the time that diagnosis occurs (Rohrer et al., 2008). Therefore, I hypothesis measures of both temporal lobes will provide a good signal of change as well as measures of frontal and parietal longitudinal volume loss. The occipital and infratentorial regions will not provide a strong atrophy measure, resulting in small effect size and larger sample size calculations.

In nfvPPA, reported atrophy profiles are also left-hemisphere dominant, including regions in the IFG, DLPFC, superior temporal gyrus and insula. As the disease develops, this pattern extends ipsilaterally to more distributed frontal, temporal lobe and anterior parietal regions as well as analogous regions in the right hemisphere (Gorno-Tempini et al., 2004; Rogalski et al.,

2011). Longitudinal measures of the left temporal region followed by left frontal and left parietal lobar change are hypothesised to provide the most robust biomarkers in this clinical population.

In contrast to svPPA and nfvPPA patients, measures of symmetric and more posterior regions may prove useful for lvPPA because they have previously been shown to develop a profile of volume loss in the temporoparietal, posterior cingulate and precuneus (Marshall et al., 2018). I hypothesis that direct measures of bilateral temporal and parietal regions will perform best at measuring longitudinal change and produce the lowest sample sizes in this population.

The distribution of neuroanatomical loss in PPA-NOS has yet to be investigated with no current literature on most robust regional measures available. However, given the prominent language deficits, the left temporal region is hypothesised to provide the most valuable biomarker of change in this population, providing the lowest sample size estimates for putative trials.

For the genetic subgroups, bilateral temporal measures are hypothesised to perform best for *MAPT* patients, a more distributed selection of bilateral temporal, frontal and parietal measures in *C9orf72* and a strongly asymmetric pattern in *GRN*, with measures of left frontal, left temporal and left parietal regions providing the best potential biomarkers and lowest sample sizes estimates.

Figure 4-1 depicts these characteristic patterns of GM loss in the clinical and genetic subgroups of FTD.

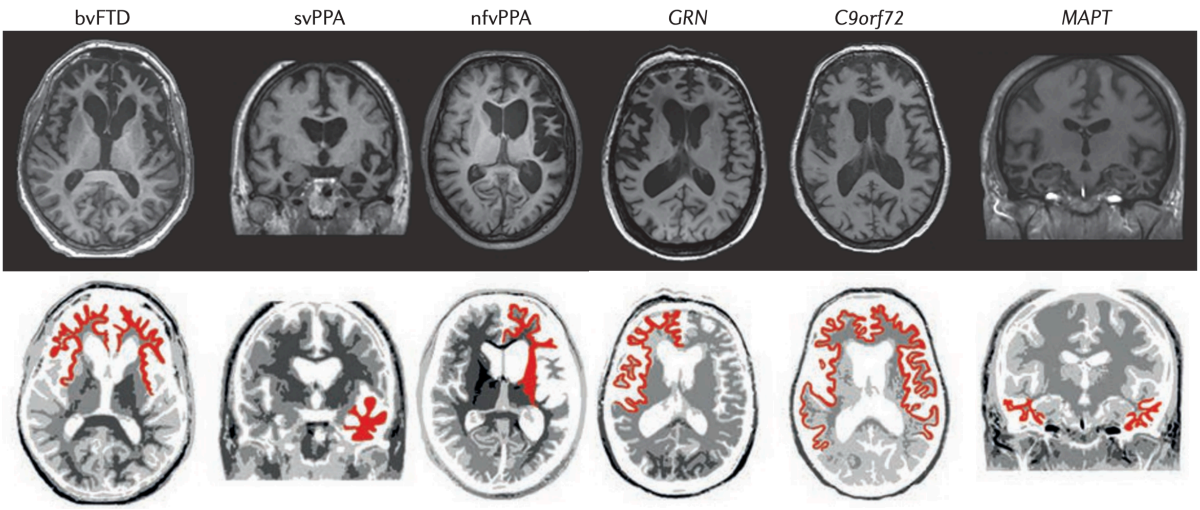

Figure 4-1 Characteristic patterns of GM atrophy (highlighted in red) in different clinical and genetic subtypes of FTD. Modified from Meeter et al., (2017).

Based on reported patterns for pathology-confirmed patients, measures of frontal and temporal volume change are likely to perform best for the Tau group, and more distributed and posterior measures of lobar change on the left may perform best in the TDP-43 group.

As with the longitudinal investigations of whole-brain change, there are only a limited number of longitudinal studies that investigated rates of lobar volume change and even fewer with accompanying sample size calculations. Table 4-1 provides a summary of these previous studies and sample size estimates. To allow for improved comparison, I back transformed a 30% treatment effect result from the published 20% or 40% reported in each paper if not available as described in Chapter 3.1.

Table 4-1 Summary of previously published sample size estimates from studies investigating neuroimaging measures of lobar volume change in FTD

Publication and cohort	Measure	bvFTD	svPPA	nfvPPA	lvPPA	FTD-combined	*MAPT*	*C9orf72*	*GRN*
Rohrer et al. (2008)									
β = 90%, α = 0.05									
svPPA (n = 20, 9 FTLD-U)									
30% treatment effect	Left temporal lobe		60						
	Right temporal lobe		55						
Whitwell, Duffy et al. (2015)									
β = 80%, α = 0.05									
Progressive apraxia of speech (n = 24)									
20% treatment effect	Superior Frontal lobe			163					
	Superior Parietal lobe			224					
30% treatment effect **	Superior Frontal lobe			72					
	Superior Parietal lobe			99					
Whitwell, Boeve et al. (2015)									
β = 80%, α = 0.05									
MAPT (n = 21), *C9orf72* (n = 11), *GRN* (n = 11), sporadic FTD (n = 15, 14 bvFTD and 1 FTD-ALS)									
Method: BSI for whole-brain and ventricles, TBM in SPM5 for lobar measures									
20% treatment effect	Frontal lobe					140	462	274	112
	Temporal lobe					197	343	123	51
	Parietal lobe					249	723	190	89
	Occipital lobe					668	3144	258	180

Publication and cohort	Measure	bvFTD	svPPA	nfvPPA	lvPPA	FTD-combined	*MAPT*	*C9orf72*	*GRN*
Whitwell, Boeve et al. (2015) continued									
30% treatment effect **	Frontal lobe					62	205	122	50
	Temporal lobe					87	152	55	23
	Parietal lobe					111	321	84	40
	Occipital lobe					297	1396	115	80
Pankov et al. (2016)									
β = 80%, α = 0.05									
bvFTD, n = 37 [14 in 1.5 T and 23 in 3 T]), svPPA, n = 49 [29 in 1.5 T and 20 in 3 T])									
Method									
20% treatment effect	Frontal lobe 1.5 T					319			
	Frontal lobe 3 T					170			
	Temporal lobe 1.5 T					161			
	Temporal lobe 3 T					76			
	Frontal and Temporal 1.5 T					145			
	Frontal and Temporal 1.5 T					83			
30% treatment effect **	Frontal lobe 1.5 T					142			
	Frontal lobe 3 T					75			
	Temporal lobe 1.5 T					71			
	Temporal lobe 3 T					34			
	Frontal and Temporal 1.5 T					64			
	Frontal and Temporal 3 T					37			

Publication and cohort	Measure	bvFTD	svPPA	nfvPPA	lvPPA	FTD-combined	*MAPT*	*C9orf72*	*GRN*
Pankov et al. (2016) continued									
40% treatment effect	Frontal lobe 1.5 T					81			
	Frontal lobe 3 T					44			
	Temporal lobe 1.5 T					41			
	Temporal lobe 3 T					20			
	Frontal and Temporal 1.5 T					37			
	Frontal and Temporal 3 T					22			
Binney et al. (2017)									
β = 80%, α = 0.05									
bvFTD (n = 44), svPPA (n = 30), or nfvPPA (n = 26)									
20% treatment effect	Frontal lobe	593	346	191					
	Temporal lobe	564	73	170					
	Frontal/Temporal	755	94	507					
30% treatment effect **	Frontal lobe	263	154	85					
	Temporal lobe	250	32	75					
	Frontal/Temporal	335	42	225					
40% treatment effect	Frontal lobe	149	88	49					
	Temporal lobe	142	19	44					
	Frontal/Temporal	190	25	128					

Publication and cohort	Measure	bvFTD	svPPA	nfvPPA	lvPPA	FTD-combined	*MAPT*	*C9orf72*	*GRN*
Staffaroni et al. (2019)									
β = 80%, α = 0.05									
bvFTD (n = 77), svPPA (n = 45), nfvPPA (n = 39)									
30% treatment effect, with 20% attrition**	Right Frontal	297	196	649					
	Left Frontal	290	212	539					
	Right Temporal	427	132	976					
	Left Temporal	398	126	709					
40% treatment effect, with 20% attrition	Right Frontal	167	110	365					
	Left Frontal	163	119	303					
	Right Temporal	240	74	549					
	Left Temporal	224	71	399					

The estimates represent the sample size that would be required per trial arm to detect the stated % treatment effect from measures of rates of change. The power (β) and statistical threshold (α) used to calculate the sample size for each study are listed individually for every publication, as are additional factors such as controlling for normal ageing and expected attrition rate from the putative trial. *Progressive apraxia of speech (PPAOS) for Whitwell et al. (2015) ** 30% calculated from the reported 20% or 40% treatment effect in each publication.

4.2 Methods

4.2.1 Participants

Participants consisted of the full retrospective neuroimaging cohort (n = 262) described in detail in Section 2.2 and Table 2-1.

4.2.2 Image analysis

All images had previously undergone segmentation using the GIF methodology as described in Methods 2.3.3.3. The full set of GIF segmentations had undergone visual QC as described in Chapter 3. All 524 segmentations passed overall global QC with minor segmentation errors noted for some regions and structures. These were most often mis-segmentation of the temporal lobes generally resulting in minor errors with the hippocampus and amygdala delineation. Some inclusion of the dura/meninges was also evident, particularly in the interhemispheric fissure. Three measures of longitudinal change were investigated. The two indirect measures were derived from GIF (a GM and a WM+GM lobar measure) and the direct PieBSI method.

Indirect lobar measure

The corresponding GIF labels were combined to provide 25 distinct regional measures for each scan. The first 12 were the left and right cortical GM segmentations for each of the frontal, temporal, insula, cingulate, parietal and occipital lobes to investigate GM cortical volume loss across the FTD spectrum (referred to as 'GIF GM volume'). As cortical GM constitutes a relatively small percentage of the whole brain, segmentation error may have a disproportionate effect on the longitudinal indirect measures derived from these automated segmentations. To investigate this effect, as well as to more closely reflect the lobar sectors produced by the direct longitudinal PieBSI measure, the remaining 13 GIF measures included the combined GM and WM labels to provide lobar segmentations (referred to as 'GIF lobar volume'). All have left and right measures reported separately, except for the infratentorial region, which was the full combined GM and WM of the

cerebellum and vermis segmentations because these are not split by hemisphere in the segmentation pipeline. The subcortical GM segmentations were not included in the lobar measures because they traversed the boundary between two or more lobar delineations. A notable benefit of using the GIF segmentation volumes to measure change was that it provided measures of insular and cingulate atrophy not currently available separately with the PieBSI. It is important to note that the insula and cingulate regions are subsumed into the temporal and the frontal and parietal regions, respectively, for the PieBSI so there is this difference in the GIF vs. PieBSI lobes. However, because the insular and cingulate cortices are important structures in FTD, indirect measures of these regions are investigated separately. All 25 indirect longitudinal measures are reported as an annual percentage volume change compared with baseline volume.

Direct lobar measure

To derive a direct longitudinal measure of lobar volume change, the PieBSI (a methodology recently developed and currently undergoing final validation at the CMIC) was applied to all 262 pairs of scans. The main procedural steps involve each image pair undergoing a symmetric affine 12 dof registration to transform both images into midpoint space, followed by computation of the same KBSI pipeline as described in Section 2.3.4.1. The input mask to inform the XOR region over which the KBSI is estimated was the combined GIF whole-brain binary mask. The regions that were combined to form the lobar regions in the indirect measures, as well as the ventricular regions, were used to define nine sectors throughout the brain. The partitioning model uses the subcortical GM segmentations as additional constraints to ensure no neighbouring lobar boundaries overlap or fail to meet. The KBSI quantification of volume change between the two scans within each of the nine sectors is then extracted as a left and right result for frontal, temporal, parietal and occipital lobes as well as a single infratentorial region result that again is not partitioned by hemisphere. As with the whole-brain KBSI, the output is ml change between scans, which was then annualised using the scan interval and presented as the annual percentage

change from baseline volume. See Section 2.3.4.2 for more information and Figure 2-12 for an example of the nine partitioned sectors from one of the current participants.

Two PieBSI registrations failed to complete the full pipeline (one control and one bvFTD participant), so their final registered images, partitioned sections and volume outputs were not available to review and were excluded from all of the analyses in this chapter. To ensure the integrity of the data for the remaining PieBSI registrations and sectors, I visually assessed the 260 aligned image pairs, BSI overlays and the nine sector segmentations for each. All visual reviews were done blinded to patient and diagnosis. Of the 260 registered image pairs and lobar segmentations, 258 passed. One pair of images failed for excessive geometric distortion unrelated to FTD after alignment that affected multiple lobar BSI sections and one for excessive motion that distorted the boundary used to quantify the volume loss across multiple sectors.

4.2.3 Statistical analysis

The statistical analysis implemented in the previous chapter was also applied here to investigate the utility of regional measures in the same way the global measures were evaluated. Regression analysis was performed correcting for age, gender and scanner type, based on the mean rates of change in each lobar measure, within each FTD subtype separately. Effect size and sample size calculations were performed, as described in detail in Section 3.3. Confidence interval calculations using the bootstrapping method with 2000 randomly sampled replications was also applied as outlined in Section 3.3.3. The head-to-head ratio of sample size paired comparison analysis was also applied to the direct PieBSI measures as outlined in Section 3.3.3.

4.3 Results

Mean (sd) rates of annual lobar volume change for the 9 direct and 25 indirect (12 GM and 13 Lobar) longitudinal measures are summarised in Table 4-2 and Table 4-3.

Figure **4-2** presents these mean lobar atrophy rates with 95% confidence intervals for the controls (in blue) compared with each FTD subgroup separately. Note the extended x-axis required for *C9orf72* and *GRN*, which required a slight extension of the lower limit from –2 to –3 to better accommodate the data. Rates of annual change were consistently low for the healthy controls across all lobes for the direct PieBSI measure, ranging from 0.1 (0.2)% to 0.2 (0.4)% across the lobes. The indirect GIF volume longitudinal measures were more variable with rates ranging from –0.1 (2.5)% for the infratentorial region to 0.7 (2.3)% for the right insula GM change. The standard deviations relative to the mean were substantially larger in the indirect than the direct measures. This was also true in general across the patient subgroups, except for several GIF frontal and temporal GM and lobar measures performing comparatively well in the range of 1.5:1–3:1.

Annual rates of change across all the FTD subgroups were notably higher than the controls for both direct and indirect measures, with a clear anterior-to-posterior gradient. In terms of the frontal and temporal PieBSI measures, these ranged from 0.7 (0.4)% for the right frontal lobe to 3.8 (1.7)% for the left temporal lobe. Both extremes of this range came from the PPA-NOS group, which demonstrated a marked left-dominant atrophy pattern.

4.3.1 Rates of change by subgroup

4.3.1.1 Clinical subgroups

For the clinical subgroups, an overall anterior–posterior gradient was evident in rates of change, with an additional asymmetric left-side dominance across the PPA subgroups. The bvFTD patients presented with a symmetric pattern of atrophy with all four frontal and temporal measures

demonstrating a 1.2% volume loss across each. For svPPA, left temporal lobe atrophy was the highest at 2.3% change and right frontal the lowest at 0.8% for these more anterior measures. NfvPPA also demonstrated higher rates on the left for both frontal and temporal lobes at 1.9% and 1.8%, respectively, but were also relatively high on the right with 1.5% and 1.2%, respectively. PPA-NOS presented with a strong left-dominant pattern of atrophy across frontal, temporal, parietal and occipital lobes, with the highest rates in the left temporal lobe at 3.8% annual change. Finally, the lvPPA also demonstrated a left-dominant anterior profile of atrophy with 1.4% annual volume loss for both frontal and temporal lobes. Although to a lesser degree, some more posterior regional measures also showed evidence of annual atrophy in the clinical subgroups, particularly in the left parietal lobe. This ranged from 0.7% for the bvFTD subgroup, to 1.9% for the PPA-NOS subgroup. The PPA-NOS and lvPPA subgroups also presented with an annual rate of left occipital volume loss at 1.4% and 1.2%, respectively. All other occipital and infratentorial measures across the clinical subgroups were under 1%.

4.3.1.2 Genetic and pathology subgroups

The genetic and pathology subgroups presented with generally characteristic atrophy patterns. The *MAPT* and Tau groups demonstrated the highest rates of change in the temporal lobe at a 1.2% annual loss on the left for both. *C9orf72* patients had lower, more variable and disperse rates of change across the frontal, temporal and parietal lobes ranging from 0.6% for the right temporal lobe to 1.1% for the left frontal lobe when measured using the PieBSI. *GRN* patients had the fastest rates with a clear left-sided dominance, including 2.1% for the left parietal, 2.4% for the left temporal and 2.6% annual volume loss for the left frontal-lobe PieBSI measures. The TDP-43 subgroup also demonstrated a distributed pattern of atrophy across all four lobes, with an average of 45% higher rates on the left than the corresponding right lobe.

Across the clinical subgroups, the indirect GIF volume-difference measures also demonstrated a similar anterior-to-posterior gradient in rates of volume change as well as the left-dominant profile seen in svPPA, nfvPPA, PPA-NOS and *GRN* subgroups evident in the PieBSI results and clearly evident in
Figure **4-2**. However, the genetic and pathology subgroups, and in particular the *C9orf72* and *GRN* subgroups, showed much greater variability due to segmentation error.

As previously mentioned, an important benefit of the indirect GIF segmentation atrophy measures is the ability to additionally investigate the insula and cingulate. For the GM and full cortices measures, both the insula and cingulate demonstrated elevated rates of annual change across the FTD spectrum compared with controls, particularly within the left insula. Left GM insula change ranged from an average 3.0 (5.0)% change in the bvFTD patients to 7.7 (2.5)% in the PPA-NOS subgroup. The full GM+WM insula cortex measure showed a similar left-dominant pattern ranging from 1.2 (2.4)% change for bvFTD to 4.3 (2.2)% for the PPA-NOS patients. In contrast, the cingulate measures demonstrated a varied pattern of asymmetry. Rates of change were higher on the right for bvFTD, *MAPT*, *C9orf72* and Tau subgroups. In both the GM and full GM/WM lobar measure for the cingulate, rates were higher in the left for all the PPA subgroups as well as the *GRN* and TDP-43 patients.

Overall, the GIF combined GM/WM lobar measures improved the mean-to-sd ratio compared with the GM only measures across the clinical subgroups. However, these improvements were generally small, with the GM measures performing comparably and providing a similar profile of atrophy across all FTD subgroups. Given this similarity and for the sake of conciseness and readability, only the GM+WM lobar GIF volume changes are presented in Figure **4-2** alongside the PiedBSI annual atrophy rates.

Table 4-2 Mean (sd) annual rates of lobar PieBSI and GIF volume difference expressed as a percentage loss from baseline volume split by FTD clinical subgroups and the controls.

Measure	Controls (n = 77)		Clinical subgroups									
			bvFTD (n = 64)		svPPA (n = 44)		nfvPPA (n = 45)		PPA-NOS (n = 7)		lvPPA (n = 21)	
PieBSI (n = 258)	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**
Frontal_left	0.1	(0.2)	1.2	(1.3)	1.2	(0.9)	1.9	(1.3)	2.1	(1.3)	1.4	(1.0)
Frontal_right	0.1	(0.2)	1.2	(1.3)	0.8	(0.7)	1.5	(1.0)	0.7	(0.4)	1.0	(0.9)
Temporal_left	0.2	(0.3)	1.2	(1.0)	2.3	(0.8)	1.8	(1.1)	3.8	(1.7)	2.0	(0.9)
Temporal_right	0.1	(0.3)	1.2	(1.2)	1.6	(0.7)	1.2	(0.8)	0.8	(0.5)	1.4	(0.8)
Parietal_left	0.2	(0.3)	0.7	(0.6)	1.0	(0.5)	1.3	(0.9)	1.9	(1.6)	1.6	(0.8)
Parietal_right	0.1	(0.3)	0.7	(0.8)	0.6	(0.4)	0.9	(0.6)	0.7	(0.6)	1.2	(0.8)
Occipital_left	0.2	(0.4)	0.7	(0.6)	0.9	(0.6)	0.8	(0.6)	1.4	(0.6)	1.2	(0.9)
Occipital_right	0.2	(0.4)	0.7	(0.6)	0.6	(0.6)	0.7	(0.5)	0.5	(0.3)	0.8	(0.7)
Infratentorial	0.1	(0.3)	0.5	(0.5)	0.3	(0.5)	0.6	(0.6)	0.5	(0.4)	0.5	(0.7)

Table 4-2 continued: Mean (sd) annual rates of lobar PieBSI and GIF volume difference expressed as a percentage loss from baseline volume split by FTD clinical subgroups and the controls.

Measure	Controls (n = 78)		Clinical subgroups									
			bvFTD (n = 66)		svPPA (n = 45)		nfvPPA (n = 45)		PPA-NOS (n = 7)		lvPPA (n = 21)	
GIF GM volume difference (n = 262)	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**
Frontal_GM_left	–0.1	(1.6)	1.9	(2.3)	2.3	(1.8)	2.8	(3.0)	4.4	(2.8)	3.0	(4.8)
Frontal_GM_right	–0.2	(1.5)	1.8	(2.4)	0.9	(2.1)	2.3	(3.1)	1.0	(1.2)	2.7	(6.1)
Temporal_GM_left	0.2	(1.1)	2.2	(2.9)	2.9	(2.6)	2.6	(1.9)	5.3	(2.4)	4.5	(3.3)
Temporal_GM_right	0.1	(1.2)	2.0	(2.7)	3.9	(2.2)	1.3	(2.4)	1.8	(1.2)	3.7	(3.0)
Insula_GM_left	0.3	(2.6)	3.0	(5.0)	4.9	(4.2)	3.9	(4.4)	7.7	(2.5)	4.1	(4.7)
Insula_GM_right	0.7	(2.3)	2.5	(6.4)	4.5	(3.5)	3.8	(4.3)	1.6	(2.0)	3.9	(5.5)
Cingulate_GM_left	0.2	(2.0)	1.3	(3.9)	2.5	(2.8)	2.5	(3.4)	3.4	(2.0)	2.1	(4.9)
Cingulate_GM_right	0.1	(1.7)	1.8	(4.3)	1.5	(2.3)	1.2	(3.0)	0.8	(1.6)	1.1	(4.3)
Parietal_GM_left	0.1	(1.2)	1.4	(3.0)	1.5	(3.5)	2.0	(2.6)	3.7	(3.3)	1.9	(7.1)
Parietal_GM_right	0.3	(1.6)	1.3	(2.8)	0.5	(3.0)	1.6	(2.1)	1.1	(1.6)	2.1	(3.2)
Occipital_GM_left	0.3	(2.9)	1.2	(2.9)	0.5	(2.9)	1.0	(3.0)	1.4	(1.8)	–1.1	(9.7)
Occipital_GM_right	0.5	(3.0)	1.1	(3.1)	0.6	(2.7)	0.3	(3.1)	–0.3	(2.0)	–0.3	(3.5)

Table 4-2 continued: Mean (sd) annual rates of lobar PieBSI and GIF volume difference expressed as a percentage loss from baseline volume split by FTD clinical subgroups and the controls.

Measure	Controls (n = 78)		Clinical subgroups									
			bvFTD (n = 66)		svPPA (n = 45)		nfvPPA (n = 45)		PPA-NOS (n = 7)		lvPPA (n = 21)	
GIF lobar volume difference (n = 262)	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**
Frontal_left	0.4	(1.1)	1.8	(2.0)	1.9	(1.3)	2.9	(2.3)	4.2	(2.5)	1.5	(2.8)
Frontal_right	0.3	(1.4)	1.9	(2.3)	0.8	(1.4)	2.6	(1.7)	1.1	(0.6)	1.0	(2.5)
Temporal_left	0.2	(1.0)	1.5	(2.0)	2.8	(2.3)	1.9	(1.7)	5.4	(1.6)	4.0	(3.0)
Temporal_right	0.1	(0.8)	1.4	(1.8)	2.8	(1.9)	1.0	(1.7)	1.7	(0.7)	2.4	(1.0)
Insula_left	0.3	(1.5)	1.2	(2.4)	2.0	(1.6)	2.0	(1.8)	4.3	(2.2)	1.7	(1.9)
Insula_right	0.2	(1.7)	1.1	(2.5)	1.4	(2.3)	1.4	(2.0)	1.2	(1.1)	1.1	(2.4)
Cingulate_left	0.2	(1.5)	1.4	(3.2)	2.3	(2.5)	2.4	(3.1)	3.6	(2.0)	1.6	(3.9)
Cingulate_right	0.1	(1.3)	1.7	(3.7)	1.4	(1.8)	1.1	(2.5)	1.1	(1.5)	0.9	(3.1)
Parietal_left	0.2	(1.0)	1.0	(1.6)	1.4	(1.4)	1.6	(1.8)	2.5	(2.3)	2.9	(2.6)
Parietal_right	0.2	(0.9)	1.1	(1.5)	0.4	(1.6)	1.0	(1.8)	0.6	(0.8)	1.3	(1.1)
Occipital_left	0.2	(1.4)	0.4	(1.9)	–0.1	(1.6)	0.2	(2.2)	0.6	(0.9)	1.9	(4.1)
Occipital_right	0.1	(1.4)	0.2	(1.9)	0.1	(2.1)	–0.2	(2.3)	–0.5	(1.1)	0.4	(2.2)
Infratentorial_diff	–0.1	(2.5)	0.1	(3.4)	–1.3	(3.3)	0.7	(2.5)	0.0	(0.7)	0.1	(4.2)

Table 4-3 Mean (sd) annual rates of lobar PieBSI and GIF volume difference expressed as a percentage loss from baseline volume split by FTD genetic and pathology subgroups.

Measure	Genetic subgroups						Pathology subgroups			
	MAPT (n = 16)		*C9orf72* (n = 10)		*GRN* (n = 8)		Tau (n = 19)		TDP-43 (n = 34)	
PieBSI (n = 258)	**mean**	**sd**	**mean**		**mean**	**sd**	**mean**	**sd**	**mean**	**sd**
Frontal_left	0.9	(0.5)	1.1	(1.1)	2.6	(2.1)	1.1	(0.7)	1.5	(1.4)
Frontal_right	0.8	(0.5)	0.9	(0.8)	1.7	(1.7)	0.9	(0.5)	1.0	(1.0)
Temporal_left	1.2	(0.6)	0.9	(0.8)	2.4	(1.8)	1.2	(0.7)	1.9	(1.3)
Temporal_right	1.1	(0.6)	0.6	(0.6)	1.2	(1.4)	1.0	(0.6)	1.3	(1.0)
Parietal_left	0.5	(0.6)	0.8	(0.9)	2.1	(1.7)	0.5	(0.4)	1.2	(1.2)
Parietal_right	0.5	(0.5)	0.7	(0.8)	1.3	(1.3)	0.4	(0.3)	0.8	(0.9)
Occipital_left	0.6	(0.8)	0.7	(0.9)	1.4	(0.6)	0.6	(0.6)	1.0	(0.7)
Occipital_right	0.5	(0.5)	0.5	(0.7)	1.0	(0.7)	0.5	(0.5)	0.7	(0.6)
Infratentorial	0.4	(0.2)	0.5	(0.8)	0.6	(0.7)	0.4	(0.2)	0.5	(0.6)

Table 4-3 continued: Mean (sd) annual rates of lobar PieBSI and GIF volume difference expressed as a percentage loss from baseline volume split by FTD genetic and pathology subgroups.

Measure	Genetic subgroups						Pathology subgroups			
	MAPT (n = 16)		*C9orf72* (n = 10)		*GRN* (n = 8)		Tau (n = 19)		TDP-43 (n = 34)	
GIF GM volume difference (n = 262)	**mean**	**sd**	**mean**		**mean**	**sd**	**mean**	**sd**	**mean**	**sd**
Frontal_GM_left	2.6	(1.5)	2.8	(1.9)	3.3	(4.2)	2.8	(1.6)	3.0	(2.5)
Frontal_GM_right	2.1	(1.5)	2.3	(2.6)	0.7	(2.4)	1.9	(1.6)	1.6	(2.1)
Temporal_GM_left	1.9	(1.6)	1.9	(2.3)	2.7	(4.8)	1.6	(1.6)	2.0	(2.8)
Temporal_GM_right	2.1	(1.1)	1.8	(2.7)	-0.4	(1.7)	2.3	(1.3)	2.0	(2.9)
Insula_GM_left	4.0	(2.6)	4.0	(2.8)	3.9	(7.2)	4.8	(3.4)	3.8	(4.2)
Insula_GM_right	4.4	(2.6)	3.1	(2.7)	–3.6	(10.9)	4.4	(2.9)	2.3	(6.8)
Cingulate_GM_left	1.2	(1.8)	–0.7	(3.0)	2.5	(5.9)	1.1	(2.2)	1.7	(3.8)
Cingulate_GM_right	2.8	(3.4)	3.5	(5.6)	0.0	(4.5)	3.0	(3.3)	2.0	(4.1)
Parietal_GM_left	1.6	(1.8)	2.6	(2.8)	3.6	(3.6)	1.2	(1.9)	2.8	(2.7)
Parietal_GM_right	1.3	(1.8)	2.1	(2.3)	2.3	(4.8)	1.0	(1.7)	1.6	(2.8)
Occipital_GM_left	0.9	(1.8)	1.0	(1.9)	3.5	(6.3)	0.5	(1.8)	0.9	(3.9)
Occipital_GM_right	0.5	(1.6)	1.0	(1.9)	1.4	(5.9)	–0.1	(2.1)	0.1	(3.5)

Table 4-3 continued: Mean (sd) annual rates of lobar PieBSI and GIF volume difference expressed as a percentage loss from baseline volume split by FTD genetic and pathology subgroups.

GIF lobar volume difference (n = 262)	Genetic subgroups						Pathology subgroups			
	MAPT (n = 16)		*C9orf72* (n = 10)		*GRN* (n = 8)		Tau (n = 19)		TDP-43 (n = 34)	
Frontal_left	1.8	(1.4)	2.6	(1.9)	3.6	(3.8)	2.2	(1.7)	2.7	(2.2)
Frontal_right	1.6	(1.2)	2.1	(1.7)	1.7	(2.3)	1.7	(1.3)	1.6	(1.7)
Temporal_left	1.5	(1.1)	1.6	(2.4)	1.7	(4.0)	1.4	(1.1)	1.9	(2.4)
Temporal_right	1.4	(1.1)	1.6	(1.7)	-0.9	(1.7)	1.5	(1.1)	1.8	(2.4)
Insula_left	1.5	(1.7)	2.0	(1.8)	2.5	(2.0)	1.7	(1.8)	2.1	(1.7)
Insula_right	1.1	(1.4)	1.6	(2.4)	-0.2	(3.0)	0.8	(2.0)	1.3	(2.3)
Cingulate_left	1.3	(1.4)	-0.3	(2.4)	2.2	(4.8)	1.2	(2.1)	1.6	(3.1)
Cingulate_right	2.7	(3.3)	3.1	(4.9)	-0.5	(4.3)	2.7	(3.1)	1.7	(3.7)
Parietal_left	0.7	(1.1)	1.9	(1.9)	1.9	(3.5)	0.7	(0.9)	1.7	(2.2)
Parietal_right	0.7	(0.8)	1.6	(1.4)	0.6	(2.1)	0.6	(0.6)	1.0	(1.7)
Occipital_left	0.4	(1.7)	1.4	(1.2)	-1.5	(2.6)	0.3	(1.8)	0.0	(1.9)
Occipital_right	-0.1	(1.6)	0.5	(2.0)	-2.4	(3.3)	-0.4	(1.4)	-0.4	(2.4)
Infratentorial_diff	-0.6	(2.7)	0.1	(3.3)	-0.1	(3.0)	-0.4	(2.6)	-0.1	(2.7)

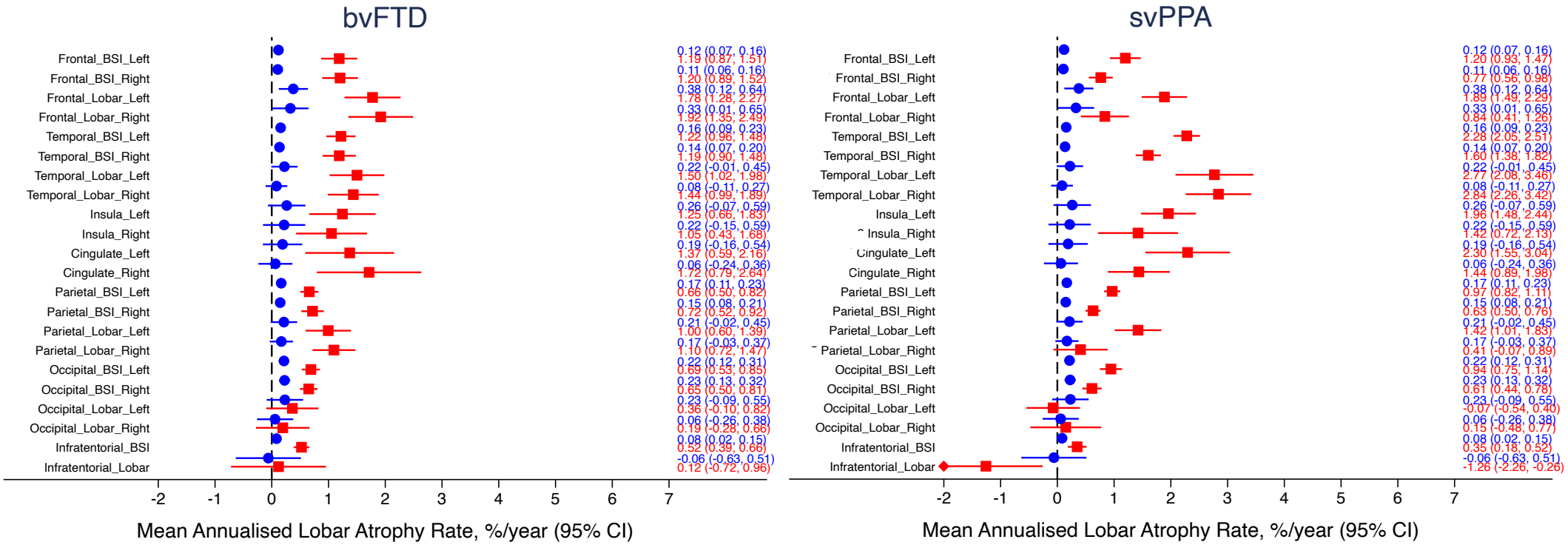

Figure 4-2 Mean annual rates of lobar volume change with 95% confidence intervals for controls (in blue) and each FTD subgroup (red) for all direct and indirect longitudinal measures.

♦ represents a capped value where the 95% limit lies outside that chosen for the x-axis range. The actual value is listed on the right of the graph.

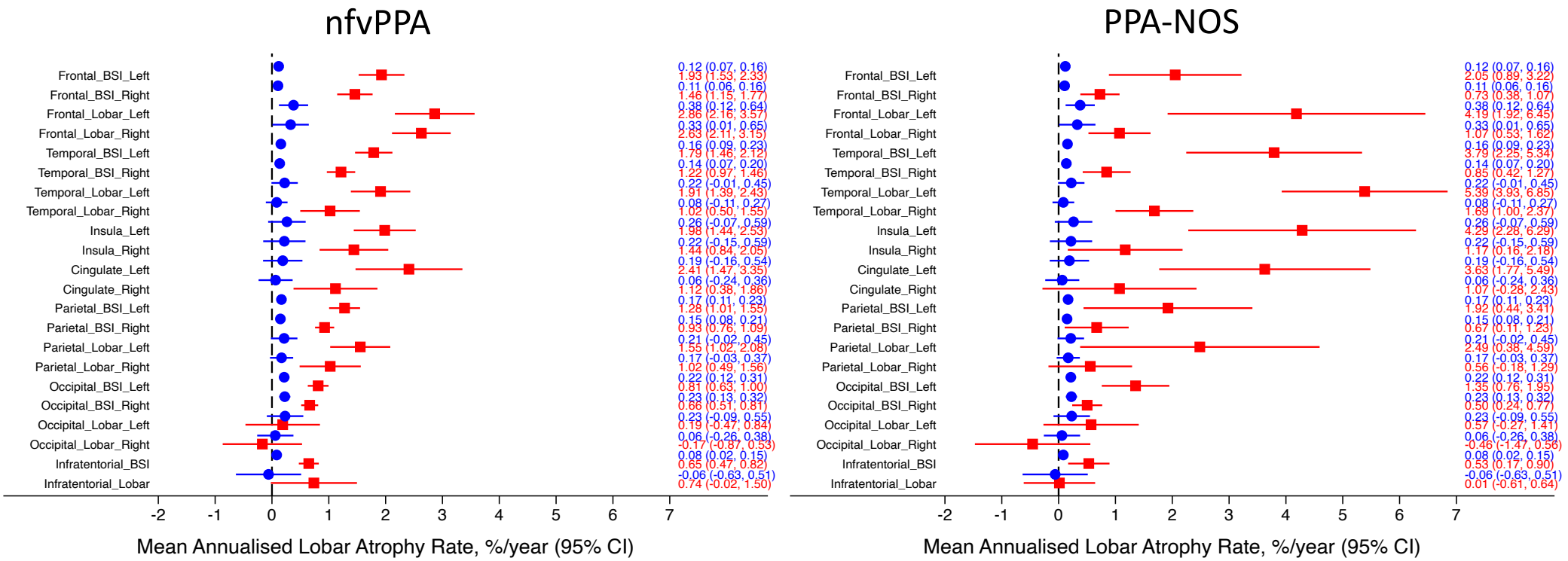

Figure 4-2 continued: Mean annual rates of lobar volume change with 95% confidence intervals for controls (in blue) and each FTD subgroup (red) for all direct and indirect longitudinal measures.

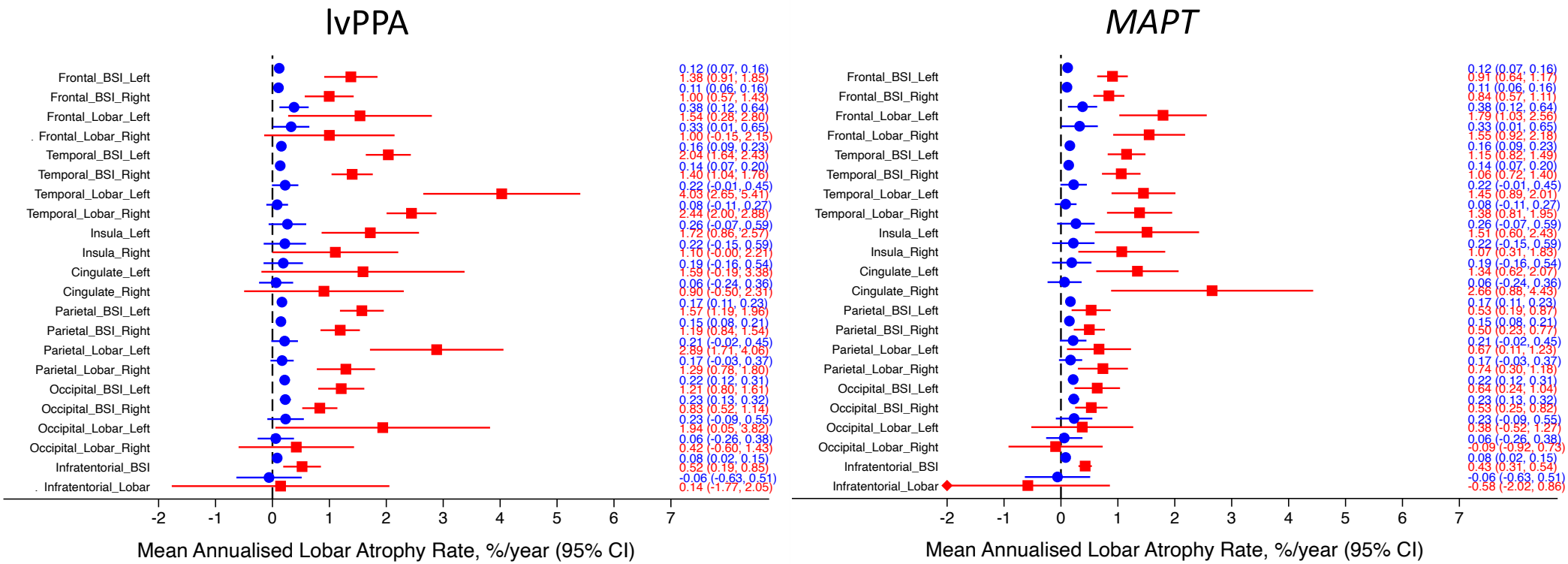

Figure 4-2 continued: Mean annual rates of lobar volume change with 95% confidence intervals for controls (in blue) and each FTD subgroup (red) for all direct and indirect longitudinal measures.

♦ represents a capped value where the 95% limit lies outside that chosen for the x-axis range. The actual value is listed on the right of the graph.

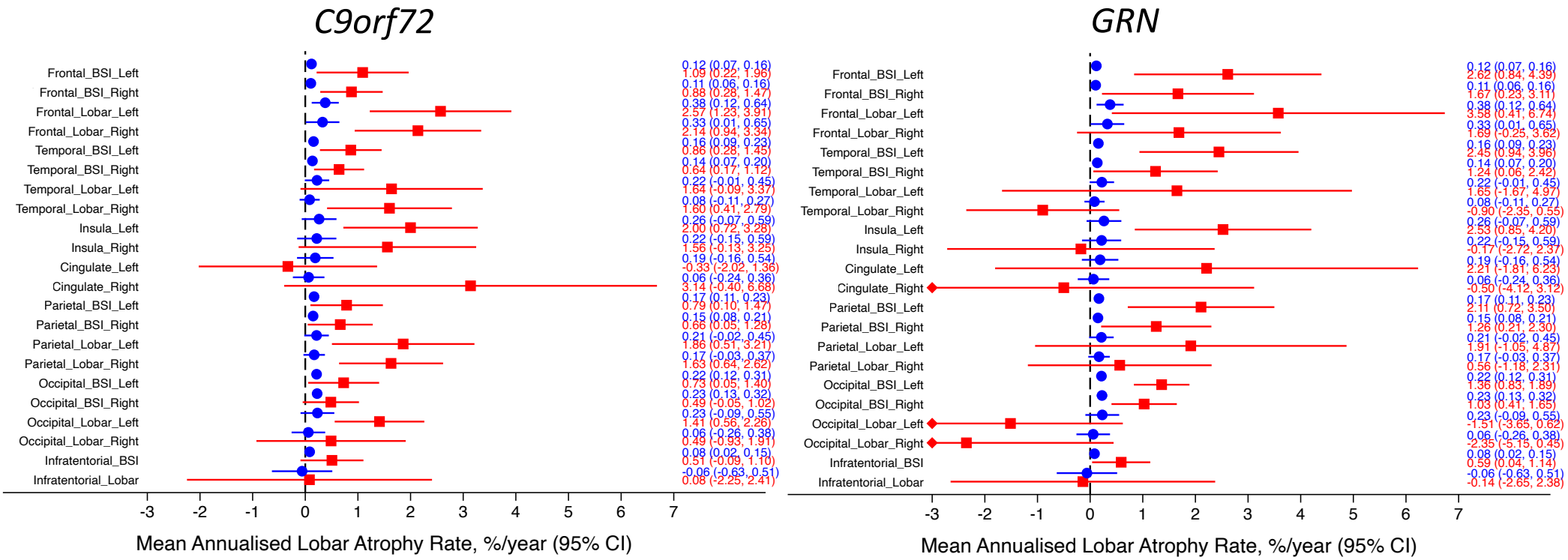

Figure 4-2 continued: Mean annual rates of lobar volume change with 95% confidence intervals for controls (in blue) and each FTD subgroup (red) for all direct and indirect longitudinal measures.

♦ represents a capped value where the 95% limit lies outside that chosen for the x-axis range. The actual value is listed on the right of the graph.

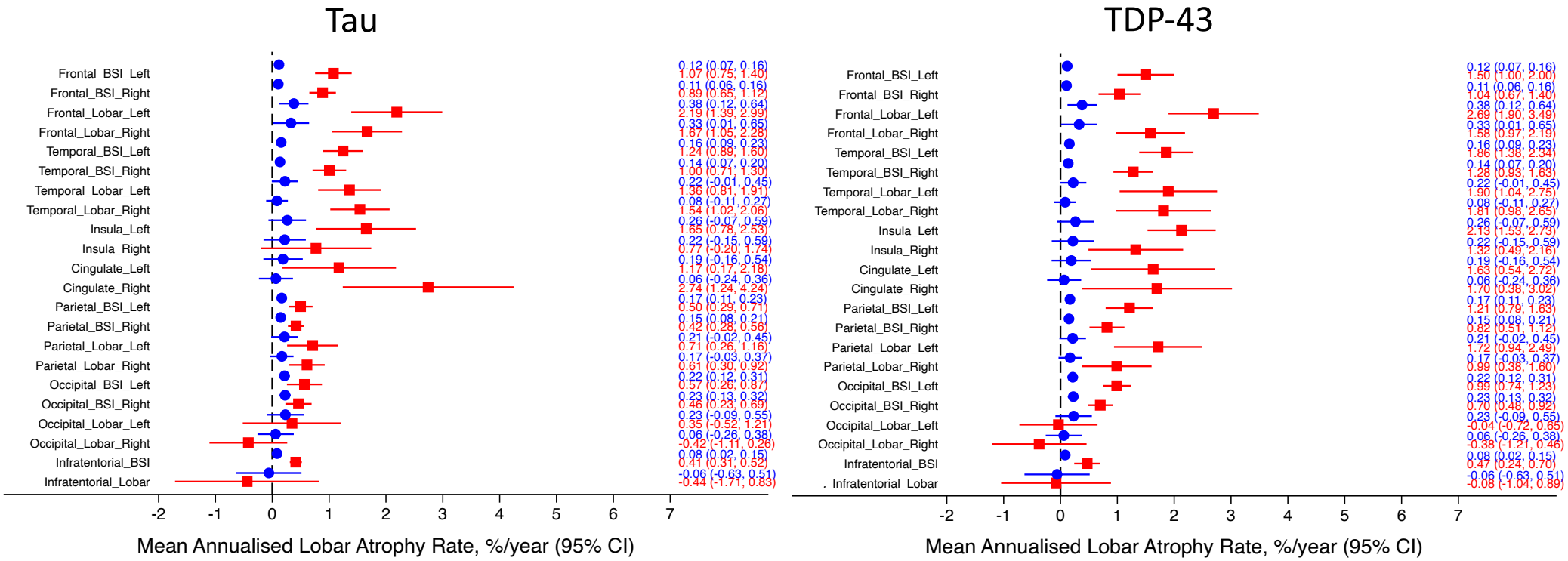

Figure 4-2 continued: Mean annual rates of lobar volume change with 95% confidence intervals for controls (in blue) and each FTD subgroup (red) for all direct and indirect longitudinal measures.

4.3.2 Regression results

Table 4-4 provides the regression analysis results with a 95% CI of the adjusted mean difference in atrophy rate for each measure and subgroup compared with controls. The regression coefficient (Coef.) indicates the point estimate of this adjusted mean difference when correcting for age, gender and scanner type. The PieBSI longitudinal measures were almost all significantly different from controls across the clinical FTD subgroups at the $p < 0.001$ level (in bold). This is except for the markedly left-dominant PPA-NOS group, whose right-sided PieBSI longitudinal measures failed to demonstrate the same level of distinction from control rates. The rates for the *MAPT* and Tau subgroups demonstrated a more anterior focal atrophy pattern with bilateral frontal and temporal lobe measures providing the best group separation from controls (all $p < 0.001$). *C9orf72* presented with a more distributed pattern of atrophy rates significantly higher than controls at the $p < 0.05$ level (in italics). All PieBSI annual measures of atrophy for the *GRN* and TDP-43 groups were significantly higher than controls at $p < 0.001$.

Given the higher variability of the indirect longitudinal measures derived from the GIF segmentation volumes, the regression analyses demonstrated significantly higher rates of change more frequently at the $p < 0.05$ level, with lower adjusted mean differences than the PieBSI measures. Reassuringly though, many of the characteristic measures of frontotemporal-insula-cingulate atrophy did provide clear group separation between FTD patients and controls at the more stringent $p < 0.001$ level (Table 4-4 and
Figure **4-2**).

Table 4-4 Regression analyses with mean adjusted difference in atrophy relative to controls (Coef.) and accompanying 95% confidence intervals for annual PieBSI and GIF volume-difference lobar change measures for each of the clinical FTD subgroups

Measure	Clinical subgroups									
	bvFTD (n = 64)		svPPA (n = 44)		nfvPPA (n = 45)		PPA-NOS (n = 7)		lvPPA (n = 21)	
PieBSI (n = 258)	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**
Frontal_left	1.0 [0.7–1.4]	**<0.001**	1.0 [0.7–1.4]	**<0.001**	1.8 [1.4–2.2]	**<0.001**	2.0 [1.2–2.7]	**<0.001**	1.3 [0.8–1.8]	**<0.001**
Frontal_right	1.0 [0.7–1.3]	**<0.001**	0.6 [0.3–0.9]	**<0.001**	1.3 [1.0–1.6]	**<0.001**	0.6 [0.0–1.3]	0.058	0.9 [0.5–1.3]	**<0.001**
Temporal_left	1.0 [0.7–1.3]	**<0.001**	2.1 [1.8–2.4]	**<0.001**	1.6 [1.3–1.9]	**<0.001**	3.6 [3.0–4.3]	**<0.001**	1.9 [1.4–2.3]	**<0.001**
Temporal_right	1.0 [0.7–1.2]	**<0.001**	1.4 [1.1–1.7]	**<0.001**	1.0 [0.7–1.3]	**<0.001**	0.7 [0.1–1.3]	*0.027*	1.2 [0.8–1.6]	**<0.001**
Parietal_left	0.5 [0.2–0.7]	**<0.001**	0.8 [0.5–1.0]	**<0.001**	1.1 [0.8–1.3]	**<0.001**	1.8 [1.3–2.3]	**<0.001**	1.4 [1.1–1.7]	**<0.001**
Parietal_right	0.5 [0.3–0.7]	**<0.001**	0.4 [0.2–0.6]	**<0.001**	0.7 [0.5–0.9]	**<0.001**	0.5 [0.1–0.9]	*0.020*	1.0 [0.7–1.3]	**<0.001**
Occipital_left	0.5 [0.3–0.7]	**<0.001**	0.7 [0.5–1.0]	**<0.001**	0.6 [0.3–0.8]	**<0.001**	1.1 [0.7–1.6]	**<0.001**	1.0 [0.7–1.3]	**<0.001**
Occipital_right	0.4 [0.2–0.6]	**<0.001**	0.4 [0.2–0.6]	**<0.001**	0.4 [0.2–0.6]	**<0.001**	0.3 [-0.1–0.7]	0.209	0.5 [0.3–0.8]	**<0.001**
Infratentorial	0.4 [0.3–0.6]	**<0.001**	0.2 [0.1–0.4]	*0.011*	0.5 [0.3–0.7]	**<0.001**	0.4 [0.1–0.8]	*0.025*	0.4 [0.1–0.6]	**<0.001**

Table 4-4 continued: Regression analyses with mean adjusted difference in atrophy relative to controls (Coef.) and accompanying 95% confidence intervals for annual PieBSI and GIF volume-difference lobar change measures for each of the clinical FTD subgroups

Measure	Clinical subgroups									
	bvFTD (n = 66)		svPPA (n = 45)		nfvPPA (n = 45)		PPA-NOS (n = 7)		lvPPA (n = 21)	
GIF GM volume difference (n = 262)	Coef. [95% CI]	*p*-value	Coef. [95% CI]	*p*-value	Coef. [95% CI]	*p*-value	Coef. [95% CI]	*p*-value	Coef. [95% CI]	*p*-value
Frontal_GM_left	2.2 [1.3–3.0]	**<0.001**	2.4 [1.5–3.4]	**<0.001**	3.1 [2.0–4.0]	**<0.001**	4.6 [2.7–6.5]	**<0.001**	3.6 [2.4–4.9]	**<0.001**
Frontal_GM_right	2.0 [1.1–2.9]	**<0.001**	1.1 [0.1–2.1]	*0.034*	2.6 [1.6–3.6]	**<0.001**	1.2 [-0.9–3.3]	0.257	3.2 [1.8–4.6]	**<0.001**
Temporal_GM_left	2.0 [1.2–2.8]	**<0.001**	2.6 [1.8–3.5]	**<0.001**	2.4 [1.5–3.3]	**<0.001**	5.1 [3.3–6.9]	**<0.001**	4.4 [3.3–5.6]	**<0.001**
Temporal_GM_right	1.9 [1.1–2.6]	**<0.001**	3.8 [3.0–4.6]	**<0.001**	1.2 [0.4–2.0]	*0.005*	1.8 [0.1–3.5]	*0.042*	3.8 [2.7–4.9]	**<0.001**
Insula_GM_left	2.7 [1.3–4.0]	**<0.001**	4.5 [3.0–6.0]	**<0.001**	3.7 [2.2–5.2]	**<0.001**	7.7 [4.5–10.8]	**<0.001**	4.4 [2.4–6.4]	**<0.001**
Insula_GM_right	2.0 [0.5–3.5]	*0.010*	3.8 [2.2–5.7]	**<0.001**	3.3 [1.7–5.0]	**<0.001**	1.1 [-2.3–4.6]	0.528	3.8 [1.6–6.0]	*0.001*
Cingulate_GM_left	1.2 [0.1–2.3]	*0.037*	2.3 [1.2–3.5]	**<0.001**	2.4 [1.2–3.6]	**<0.001**	3.3 [0.8–5.8]	*0.010*	2.2 [0.5–3.7]	*0.009*
Cingulate_GM_right	1.5 [0.5–2.6]	*0.005*	1.3 [0.2–2.5]	*0.024*	1.1 [-0.1–2.2]	0.065	0.8 [-1.7–3.2]	0.541	1.1 [-0.4–2.7]	0.161
Parietal_GM_left	1.3 [0.2–2.4]	*0.020*	1.4 [0.2–2.6]	*0.019*	1.9 [0.8–3.0]	*0.001*	3.7 [1.2–6.2]	*0.004*	2.0 [0.4–3.6]	*0.014*
Parietal_GM_right	1.0 [0.1–1.8]	*0.023*	0.3 [-0.7–1.2]	0.576	1.2 [0.3–2.2]	*0.008*	0.7 [-1.1–2.7]	0.442	1.6 [0.4–2.9]	*0.010*
Occipital_GM_left	1.0 [-0.3–2.3]	0.140	0.2 [-1.2–1.7]	0.770	0.6 [-0.9–2.1]	0.421	1.0 [-2.0–4.1]	0.499	-1.6 [-3.5–0.4]	0.112
Occipital_GM_right	0.7 [-0.3–1.7]	0.185	0.3 [-0.8–1.4]	0.578	-0.2 [-1.3–0.9]	0.686	-1.1 [-3.4–1.2]	0.343	-1.3 [-2.7–0.2]	0.091

Table 4-4 continued: Regression analyses with mean adjusted difference in atrophy relative to controls (Coef.) and accompanying 95% confidence intervals for annual PieBSI and GIF volume-difference lobar change measures for each of the clinical FTD subgroups

Measure	Clinical subgroups									
	bvFTD (n = 66)		svPPA (n = 45)		nfvPPA (n = 45)		PPA-NOS (n = 7)		lvPPA (n = 21)	
GIF lobar volume difference (n = 262)	Coef. [95% CI]	*p*-value	Coef. [95% CI]	*p*-value	Coef. [95% CI]	*p*-value	Coef. [95% CI]	*p*-value	Coef. [95% CI]	*p*-value
Frontal_left	1.4 [0.8–2.1]	**<0.001**	1.5 [0.8–2.2]	**<0.001**	2.6 [1.9–3.2]	**<0.001**	3.8 [2.4–5.3]	**<0.001**	1.4 [0.4–2.3]	*0.004*
Frontal_right	1.5 [0.9–2.2]	**<0.001**	0.5 [-0.2–1.2]	0.131	2.3 [1.6–3.0]	**<0.001**	0.7 [-0.7–2.1]	0.341	0.7 [-0.2–1.6]	0.142
Temporal_left	1.2 [0.5–1.8]	**<0.001**	2.5 [1.8–3.2]	**<0.001**	1.6 [0.9–2.3]	**<0.001**	5.0 [3.6–6.5]	**<0.001**	3.6 [2.7–4.5]	**<0.001**
Temporal_right	1.3 [0.7–1.8]	**<0.001**	2.7 [2.1–3.3]	**<0.001**	0.9 [0.3–1.5]	*0.002*	1.6 [0.4–2.8]	*0.008*	2.4 [1.6–3.2]	**<0.001**
Insula_left	1.1 [0.5–1.7]	*0.001*	1.8 [1.1–2.5]	**<0.001**	1.8 [1.1–2.5]	**<0.001**	4.0 [2.6–5.4]	**<0.001**	1.6 [0.7–2.5]	*0.001*
Insula_right	1.1 [0.4–1.8]	*0.004*	1.3 [0.6–2.1]	*0.001*	1.4 [0.6–2.2]	*0.001*	1.0 [-0.7–2.6]	0.249	1.1 [0.1–2.2]	*0.037*
Cingulate_left	1.3 [0.3–2.2]	*0.008*	2.1 [1.1–3.2]	**<0.001**	2.3 [1.3–3.3]	**<0.001**	3.5 [1.4–5.6]	**<0.001**	1.7 [0.3–3.0]	*0.016*
Cingulate_right	1.5 [0.6–2.4]	*0.001*	1.3 [0.4–2.2]	*0.007*	1.0 [0.1–2.0]	*0.031*	1.0 [-0.9–3.0]	0.300	1.0 [-0.3–2.1]	0.142
Parietal_left	0.7 [0.1–1.2]	*0.013*	1.2 [0.6–1.8]	**<0.001**	1.3 [0.7–1.9]	**<0.001**	2.2 [1.0–3.4]	**<0.001**	2.6 [1.8–3.4]	**<0.001**
Parietal_right	0.8 [0.3–1.3]	**<0.001**	0.2 [-0.3–0.7]	0.391	0.7 [0.2–1.3]	*0.005*	0.2 [-0.8–1.3]	0.651	0.9 [0.2–1.5]	*0.016*
Occipital_left	0.0 [-0.7–0.7]	0.916	-0.3 [-1.1–0.4]	0.411	-0.1 [-0.9–0.6]	0.743	0.2 [-1.3–1.8]	0.772	1.5 [0.5–2.5]	*0.003*
Occipital_right	0.2 [-0.5–0.8]	0.640	0.1 [-0.6–0.8]	0.756	-0.3 [-1.0–0.4]	0.447	-0.6 [-2.1–0.9]	0.438	0.2 [-0.7–1.2]	0.625
Infratentorial	0.1 [-0.9–1.1]	0.871	-1.3 [-2.5–-0.2]	*0.020*	0.8 [-0.3–1.9]	0.168	0.3 [-2.1–2.6]	0.834	0.5 [-1.0–2.0]	0.488

Table 4-5 Regression analyses with mean adjusted difference in atrophy relative to controls (Coef.) and accompanying 95% confidence intervals for annual PieBSI and GIF volume-difference lobar change measures for each of the genetic and pathology FTD subgroups

Measure	Genetic subgroups						Pathology subgroup			
	***MAPT* (n = 16)**		***C9orf72* (n = 10)**		***GRN* (n = 8)**		**Tau (n = 19)**		**TDP-43 (n = 34)**	
PieBSI (n = 258)	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**
Frontal_left	0.9 [0.5–1.3]	**<0.001**	1.1 [0.6–1.5]	**<0.001**	2.5 [2.0–2.9]	**<0.001**	1.1 [0.7–1.5]	**<0.001**	1.4 [1.1–1.7]	**<0.001**
Frontal_right	0.8 [0.5–1.2]	**<0.001**	0.8 [0.4–1.2]	**<0.001**	1.1 [1.1–2.0]	**<0.001**	0.9 [0.5–1.1]	**<0.001**	0.9 [0.7–1.2]	**<0.001**
Temporal_left	1.1 [0.7–1.4]	**<0.001**	0.7 [0.2–1.1]	*0.003*	2.3 [1.8–2.7]	**<0.001**	1.2 [0.8–1.6]	**<0.001**	1.7 [1.4–2.0]	**<0.001**
Temporal_right	1.0 [0.7–1.3]	**<0.001**	0.5 [0.1–0.8]	*0.017*	1.1 [0.7–1.5]	**<0.001**	0.9 [0.6–1.2]	**<0.001**	1.1 [0.9–1.4]	**<0.001**
Parietal_left	0.5 [0.1–0.8]	*0.010*	0.6 [0.2–1.1]	*0.004*	1.9 [1.5–2.4]	**<0.001**	0.4 [0.0–0.8]	*0.026*	1.1 [0.8–1.3]	**<0.001**
Parietal_right	0.4 [0.1–0.7]	*0.005*	0.5 [0.1–0.8]	*0.008*	1.1 [0.8–1.5]	**<0.001**	0.3 [0.0–0.6]	*0.022*	0.7 [0.5–0.9]	**<0.001**
Occipital_left	0.5 [0.2–0.8]	*0.001*	0.5 [0.1–0.9]	*0.007*	1.2 [0.8–1.5]	**<0.001**	0.4 [0.1–0.7]	*0.005*	0.8 [0.6–1.0]	**<0.001**
Occipital_right	0.4 [0.1–0.6]	*0.008*	0.2 [-0.1–0.5]	0.226	0.9 [0.5–1.2]	**<0.001**	0.3 [0.0–0.6]	*0.024*	0.5 [0.3–0.7]	**<0.001**
Infratentorial	0.4 [0.2–0.6]	**<0.001**	0.4 [0.21–0.7]	*0.003*	0.5 [0.2–0.8]	**<0.001**	0.4 [0.2–0.6]	*0.001*	0.4 [0.2–0.6]	**<0.001**

Table 4-5 continued: Regression analyses with mean adjusted difference in atrophy relative to controls (Coef.) and accompanying 95% confidence intervals for annual PieBSI and GIF volume-difference lobar change measures for each of the genetic and pathology FTD subgroups

Measure	Genetic subgroups						Pathology subgroup			
	***MAPT* (n = 16)**		***C9orf72* (n = 10)**		***GRN* (n = 8)**		**Tau (n = 19)**		**TDP-43 (n = 34)**	
GIF GM volume difference (n = 262)	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**
Frontal_GM_left	2.9 [1.8–3.9]	**<0.001**	3.2 [2.0–4.5]	**<0.001**	3.1 [1.8–4.4]	**<0.001**	3.0 [2.0–3.9]	**<0.001**	3.1 [2.3–3.8]	**<0.001**
Frontal_GM_right	2.3 [1.3–3.2]	**<0.001**	2.8 [1.6–3.9]	**<0.001**	0.7 [-0.6–1.9]	0.208	2.0 [1.1–2.9]	**<0.001**	1.7 [1.0–2.4]	**<0.001**
Temporal_GM_left	1.8 [0.8–2.8]	*0.001*	1.6 [0.4–2.8]	*0.012*	2.4 [1.1–3.7]	**<0.001**	1.4 [0.5–2.4]	*0.003*	1.8 [1.0–2.5]	**<0.001**
Temporal_GM_right	2.0 [1.1–2.9]	**<0.001**	1.7 [0.6–2.8]	*0.002*	-0.5 [-1.6–0.7]	0.414	2.2 [1.2–3.2]	**<0.001**	1.9 [1.2–2.7]	**<0.001**
Insula_GM_left	4.0 [2.2–5.8]	**<0.001**	4.1 [1.9–6.3]	**<0.001**	3.3 [1.0–5.6]	*0.006*	4.8 [3.1–6.5]	**<0.001**	3.5 [2.2–4.9]	**<0.001**
Insula_GM_right	3.8 [1.6–5.9]	*0.001*	2.7 [0.1–5.3]	*0.040*	-4.4 [-7.2–-1.7]	*0.002*	3.7 [1.6–5.9]	*0.001*	1.7 [0.0–3.3]	0.053
Cingulate_GM_left	1.2 [-0.3–2.6]	0.112	-0.8 [-2.6–0.8]	0.345	2.2 [0.3–4.0]	*0.023*	1.0 [-0.3–2.4]	0.144	1.5 [0.4–2.6]	*0.007*
Cingulate_GM_right	2.6 [1.0–4.2]	*0.002*	3.2 [1.2–5.2]	*0.002*	0.0 [-2.1–2.0]	0.970	2.8 [1.4–4.3]	**<0.001**	1.8 [0.6–2.9]	*0.003*
Parietal_GM_left	1.6 [0.6–2.6]	*0.002*	2.5 [1.3–3.7]	*0.002*	3.4 [2.2–4.7]	**<0.001**	1.1 [0.1–2.1]	*0.026*	2.6 [1.9–3.4]	**<0.001**
Parietal_GM_right	1.1 [0.0–2.3]	0.060	1.6 [0.2–2.9]	*0.030*	2.1 [0.6–3.6]	*0.006*	0.8 [-0.3–1.7]	0.146	1.3 [0.4–2.1]	*0.003*
Occipital_GM_left	0.6 [-1.1–2.3]	0.510	0.5 [-1.6–2.6]	0.641	3.4 [1.2–6.0]	*0.003*	0.3 [-1.3–1.9]	0.710	0.7 [-0.5–2.0]	0.260
Occipital_GM_right	0.0 [-1.6–1.7]	0.965	0.5 [-1.5–2.4]	0.654	1.2 [-0.9–3.4]	0.248	-0.4 [-2.0–1.1]	0.593	-0.1 [-1.3–1.1]	0.863

Table 4-5 continued: Regression analyses with mean adjusted difference in atrophy relative to controls (Coef.) and accompanying 95% confidence intervals for annual PieBSI and GIF volume-difference lobar change measures for each of the genetic and pathology FTD subgroups

Measure	Genetic subgroups						Pathology subgroup			
	***MAPT* (n = 16)**		***C9orf72* (n = 10)**		***GRN* (n = 8)**		**Tau (n = 19)**		**TDP-43 (n = 34)**	
GIF lobar volume difference (n = 262)	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**
Frontal_left	1.6 [0.7–2.4]	*0.001*	2.3 [1.2–3.4]	**<0.001**	3.1 [2.0–4.3]	**<0.001**	1.9 [1.0–2.8]	**<0.001**	2.3 [1.7–3.0]	**<0.001**
Frontal_right	1.4 [0.6–2.3]	*0.001*	2.0 [0.9–3.0]	**<0.001**	1.3 [0.2–2.4]	*0.021*	1.5 [0.7–2.3]	**<0.001**	1.2 [0.6–1.9]	**<0.001**
Temporal_left	1.3 [0.4–2.2]	*0.003*	1.2 [0.1–2.3]	*0.027*	1.5 [0.3–2.6]	*0.012*	1.2 [0.4–2.0]	*0.004*	1.6 [1.0–2.3]	**<0.001**
Temporal_right	1.4 [0.8–2.0]	**<0.001**	1.4 [0.7–2.2]	**<0.001**	-1.0 [-1.7–-0.2]	*0.016*	1.5 [0.7–2.2]	**<0.001**	1.7 [1.1–2.2]	**<0.001**
Insula_left	1.4 [0.5–2.3]	*0.002*	1.8 [0.8–2.9]	*0.001*	2.3 [1.2–3.4]	**<0.001**	1.6 [0.8–2.4]	**<0.001**	2.0 [1.3–2.6]	**<0.001**
Insula_right	0.9 [-0.1–2.0]	0.071	1.6 [0.3–2.8]	*0.014*	-0.4 [-1.7–0.9]	*0.548*	0.7 [-0.3–1.7]	0.187	1.2 [0.5–2.0]	*0.002*
Cingulate_left	1.3 [0.2–2.5]	*0.021*	-0.5 [-1.9–0.9]	0.487	1.9 [0.5–3.4]	*0.010*	1.1 [0.–2.2]	0.052	1.4 [0.5–2.3]	*0.002*
Cingulate_right	2.6 [1.2–4.0]	*0.001*	2.9 [1.1–4.6]	*0.001*	-0.6 [-2.4–1.3]	0.541	2.6 [1.3–3.9]	**<0.001**	1.5 [0.5–2.6]	*0.003*
Parietal_left	0.6 [-0.2–1.4]	0.159	1.6 [0.6–2.6]	*0.002*	1.6 [0.6–2.7]	*0.003*	0.6 [-0.2–1.3]	0.142	1.5 [0.9–2.1]	**<0.001**
Parietal_right	0.7 [0.1–1.3]	*0.027*	1.3 [0.6–2.0]	*0.001*	0.5 [-0.3–1.2]	0.236	0.5 [0.0–1.1]	0.063	0.8 [0.3–1.2]	*0.001*
Occipital_left	0.2 [-0.6–1.1]	0.573	1.0 [-0.1–2.0]	*0.071*	-1.7 [-2.8–-0.6]	*0.003*	0.2 [-0.7–1.0]	0.667	-0.3 [-1.0–0.3]	0.338
Occipital_right	0.0 [-1.0–0.9]	0.956	0.3 [-0.8–1.5]	0.572	-2.4 [-3.6–-1.1]	**<0.001**	-0.4 [-1.3–0.5]	0.371	-0.4 [-1.2–0.3]	0.225
Infratentorial	-0.3 [-1.9–1.3]	0.674	0.3 [-1.6–2.2]	0.773	-0.2 [-2.2–1.8]	0.857	-0.3 [-1.8–1.2]	0.705	0.0 [-1.2–1.1]	0.937

4.3.3 Effect size

Effect size and 95% BCa CI derived from these annual rates of change are reported in Table 4-6 and Table 4-7. Effect sizes of one or above are presented in bold, indicating a large effect (good group separation from controls), as based on Cohen's (1988) formulation. The more heterogeneous bvFTD patients demonstrated lower effect sizes than the PPA subgroups but were all 0.8 or higher for the PieBSI measures. The left temporal PieBSI measure provided the largest effect size for all PPA groups, ranging from 1.5 [1.2–1.8] in nfvPPA to 2.8 [2.2–3.6] in svPPA. More posterior atrophy detected using the parietal and left occipital PieBSI also provided a strong measure of change within the PPA subgroups. Here, effect sizes ranged from 0.9 [0.5–1.5] for right parietal to 1.8 [1.0–2.7] for left occipital both in the PPA-NOS subgroup. The GIF volume measures also provided effect size calculations that performed well in the more anterior regions and within the PPA subgroups clinically. *MAPT* and Tau patients presented with effect sizes generally between 1 and 2 for all frontal and temporal lobes. Effect sizes were lower and more variable for the *C9orf72*, *GRN* and TDP-43 groups. In addition, measures of the left insula and left cingulate change provided higher effect size calculations, particularly for the svPPA and PPA-NOS patients.

Table 4-6 Effect size calculations (95% BCa confidence intervals) for PieBSI and GIF volume-difference lobar measures for the clinical FTD subgroups*

Measure	Clinical subgroups									
	bvFTD (n = 64)		svPPA (n = 44)		nfvPPA (n = 45)		PPA-NOS (n = 7)		lvPPA (n = 21)	
PieBSI (n = 258)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)
Frontal_left	0.8	[0.7–1.0]	**1.2**	[0.8–1.7]	**1.4**	[1.1–1.7]	**1.5**	[0.8–2.5]	**1.2**	[0.8–1.8]
Frontal_right	0.9	[0.7–1.0]	**1.0**	[0.7–1.4]	**1.3**	[1.1–1.6]	**1.7**	[0.7–3.0]	0.9	[0.5–1.4]
Temporal_left	**1.0**	[0.8–1.3]	**2.8**	[2.2–3.6]	**1.5**	[1.2–1.8]	**2.2**	[1.1–3.8]	**2.2**	[1.5–2.8]
Temporal_right	0.9	[0.7–1.1]	**2.0**	[1.6–2.5]	**1.3**	[1.0–1.7]	**1.6**	[1.0–2.5]	**1.6**	[1.1–2.1]
Parietal_left	0.8	[0.6–1.0]	**1.7**	[1.4–2.1]	**1.2**	[1.0–1.6]	**1.1**	[0.7–1.6]	**1.7**	[1.1–2.2]
Parietal_right	0.7	[0.5–0.9]	**1.1**	[0.8–1.4]	**1.4**	[1.1–1.8]	0.9	[0.5–1.5]	**1.4**	[0.9–1.8]
Occipital_left	0.8	[0.5–1.0]	**1.1**	[0.8–1.5]	**1.0**	[0.6–1.3]	**1.8**	[1.0–2.7]	**1.1**	[0.7–1.5]
Occipital_right	0.7	[0.4–0.9]	0.7	[0.4–1.0]	0.9	[0.5–1.3]	**1.0**	[-0.1–1.8]	0.9	[0.4–1.4]
Infratentorial	0.8	[0.6–1.1]	0.5	[0.1–0.8]	**1.0**	[0.6–1.3]	**1.1**	[0.3–2.0]	0.6	[0.2–0.9]

* Effect sizes of 1.0 or above are in **bold** to indicate a high degree of group separation from controls

Table 4-6 continued: Effect size calculations (95% BCa confidence intervals) for PieBSI and GIF volume-difference lobar measures for the clinical FTD subgroups

Measure	Clinical subgroups									
	bvFTD (n = 64)		svPPA (n = 44)		nfvPPA (n = 45)		PPA-NOS (n = 7)		lvPPA (n = 21)	
GIF GM volume difference (n = 262)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)
Frontal_GM_left	0.9	[0.5–1.3]	**1.3**	[0.9–1.9]	**1.0**	[0.6–1.5]	**1.6**	[0.1–3.7]	0.7	[0.3–1.0]
Frontal_GM_right	0.8	[0.5–1.2]	0.5	[0.0–0.9]	0.8	[0.4–1.2]	**1.0**	[0.0–2.0]	0.5	[0.0–0.7]
Temporal_GM_left	0.7	[0.4–0.9]	**1.0**	[0.7–1.3]	**1.2**	[0.8–1.6]	**2.1**	[1.5–3.5]	**1.3**	[0.9–2.6]
Temporal_GM_right	0.7	[0.5–1.0]	**1.7**	[1.2–2.2]	0.5	[0.1–0.9]	**1.4**	[0.5–4.3]	**1.2**	[0.8–1.6]
Insula_GM_left	0.5	[0.2–0.8]	**1.1**	[0.7–1.5]	0.8	[0.4–1.2]	**3.0**	[1.5–6.1]	0.8	[0.2–1.3]
Insula_GM_right	0.3	[-0.1–0.7]	**1.1**	[0.7–1.5]	0.7	[0.4–1.0]	0.4	[-0.7–1.3]	0.6	[0.0–1.0]
Cingulate_GM_left	0.3	[0.0–0.6]	0.8	[0.4–1.2]	0.7	[0.3–1.0]	**1.7**	[0.6–2.6]	0.4	[-0.2–0.7]
Cingulate_GM_right	0.4	[0.1–0.7]	0.6	[0.3–1.0]	0.4	[0.0–0.8]	0.4	[-0.6–1.3]	0.2	[-0.2–0.6]
Parietal_GM_left	0.4	[0.1–0.7]	0.4	[-0.1–1.4]	0.7	[0.3–1.1]	**1.1**	[0.3–1.6]	0.2	[-0.3–1.4]
Parietal_GM_right	0.3	[0.1–0.6]	0.1	[-0.3–0.5]	0.6	[0.2–1.0]	0.5	[-0.4–1.3]	0.6	[0.0–1.2]
Occipital_GM_left	0.3	[0.0–0.6]	0.1	[-0.3–0.5]	0.2	[-0.2–0.5]	0.6	[-0.4–1.5]	-0.2	[-0.6–0.4]
Occipital_GM_right	0.2	[-0.1–0.5]	0.0	[-0.4–0.4]	-0.1	[-0.5–0.3]	-0.4	[-1.4–0.7]	-0.2	[-0.7–0.3]

* Effect sizes of 1.0 or above are in **bold** to indicate a high degree of group separation from controls

Table 4-6 continued: Effect size calculations (95% BCa confidence intervals) for PieBSI and GIF volume-difference lobar measures for the clinical FTD subgroups

Measure	Clinical subgroups									
	bvFTD (n = 64)		svPPA (n = 44)		nfvPPA (n = 45)		PPA-NOS (n = 7)		lvPPA (n = 21)	
GIF lobar volume difference (n = 262)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)
Frontal_left	0.7	[0.4–1.0]	**1.1**	[0.7–1.6]	**1.1**	[0.6–1.5]	**1.6**	[0.7–2.6]	0.4	[-0.2–1.2]
Frontal_right	0.7	[0.4–1.0]	0.4	[0.0–0.7]	**1.3**	[1.0–1.7]	**1.3**	[0.1–2.2]	0.3	[-0.3–1.0]
Temporal_left	0.7	[0.4–0.9]	**1.1**	[0.8–1.5]	**1.0**	[0.5–1.4]	**3.3**	[2.3–4.9]	**1.3**	[0.8–2.2]
Temporal_right	0.7	[0.5–1.0]	**1.4**	[0.8–2.0]	0.5	[0.1–1.0]	**2.2**	[1.1–3.2]	**2.4**	[1.6–3.3]
Insula_left	0.4	[0.1–0.7]	**1.1**	[0.7–1.4]	0.9	[0.6–1.3]	**1.9**	[1.3–2.6]	0.8	[0.2–1.2]
Insula_right	0.3	[0.0–0.6]	0.5	[0.0–1.0]	0.6	[0.2–1.0]	0.9	[-0.5–1.5]	0.4	[-0.2–0.8]
Cingulate_left	0.4	[0.1–0.6]	0.8	[0.3–1.3]	0.7	[0.4–0.9]	**1.7**	[1.1–2.6]	0.4	[-0.2–0.7]
Cingulate_right	0.4	[0.1–0.7]	0.8	[0.4–1.1]	0.4	[0.0–0.8]	0.7	[-0.3–1.4]	0.3	[-0.2–0.6]
Parietal_left	0.5	[0.2–0.8]	0.9	[0.5–1.3]	0.8	[0.3–1.2]	**1.0**	[0.4–1.6]	**1.0**	[0.6–1.4]
Parietal_right	0.6	[0.3–0.9]	0.2	[-0.2–0.5]	0.5	[0.1–0.9]	0.5	[-0.7–1.2]	**1.0**	[0.5–1.5]
Occipital_left	0.1	[-0.2–0.4]	-0.2	[-0.6–0.2]	0.0	[-0.4–0.3]	0.4	[-0.6–1.4]	0.4	[-0.1–0.6]
Occipital_right	0.1	[-0.2–0.4]	0.0	[-0.3–0.4]	-0.1	[-0.4–0.3]	-0.5	[-1.4–0.6]	0.2	[-0.3–0.7]
Infratentorial	0.1	[-0.2–0.4]	-0.4	[-0.6–0.0]	0.3	[-0.1–0.8]	0.1	[-1.1–2.6]	0.0	[-0.4–0.5]

* Effect sizes of 1.0 or above are in **bold** to indicate a high degree of group separation from controls

Table 4-7 Effect size calculations (95% BCa confidence intervals) for PieBSI and GIF volume-difference lobar measures for the genetic and pathology FTD subgroups

Measure	Genetic subgroups						Pathology subgroup			
	MAPT (n = 16)		*C9orf72* (n = 10)		*GRN* (n = 8)		Tau (n = 19)		TDP43 (n = 34)	
PIEBSI (n=258)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)
Frontal_left	**1.6**	[1.0 – 2.1]	0.9	[0.5 – 1.3]	**1.2**	[0.4 – 1.9]	**1.4**	[1.0 – 1.9]	**1.0**	[0.7 – 1.3]
Frontal_right	**1.4**	[0.9 – 1.9]	**1.0**	[0.6 – 1.5]	0.9	[0.4 – 1.6]	**1.6**	[1.1 – 2.1]	0.9	[0.7 – 1.1]
Temporal_left	**1.6**	[1.1 – 2.1]	0.9	[0.5 – 1.4]	**1.3**	[0.5 – 2.2]	**1.5**	[1.1 – 1.9]	**1.3**	[0.9 – 1.7]
Temporal_right	**1.5**	[1.0 – 1.9]	0.8	[0.3 – 1.4]	0.8	[0.3 – 1.1]	**1.4**	[1.0 – 1.8]	**1.2**	[0.8 – 1.6]
Parietal_left	0.6	[0.0 – 0.9]	0.7	[0.0 – 1.2]	**1.2**	[0.7 – 1.8]	0.8	[0.3 – 1.3]	0.9	[0.6 – 1.2]
Parietal_right	0.7	[0.2 – 1.0]	0.6	[-0.1 – 1.2]	0.9	[0.5 – 1.2]	0.9	[0.4 – 1.4]	0.8	[0.5 – 1.0]
Occipital_left	0.6	[0.1 – 0.9]	0.6	[0.0 – 1.1]	**1.8**	[1.2 – 2.7]	0.6	[0.1 – 0.9]	**1.1**	[0.8 – 1.5]
Occipital_right	0.6	[0.0 – 1.0]	0.4	[-0.4 – 0.9]	**1.1**	[-0.1 – 1.6]	0.5	[0.0 – 0.9]	0.8	[0.4 – 1.1]
Infratentorial	**1.6**	[0.9 – 2.1]	0.5	[-0.2 – 1.1]	0.8	[-0.2 – 1.9]	**1.5**	[1.0 – 2.1]	0.6	[0.3 – 0.9]

* Effect sizes of 1.0 or above are in **bold** to indicate a high degree of group separation from controls

Table 4-7 continued: Effect size calculations (95% BCa confidence intervals) for PieBSI and GIF volume-difference lobar measures for the genetic and pathology FTD subgroups

Measure	Genetic subgroups						Pathology subgroups			
	MAPT (n = 16)		*C9orf72* (n = 10)		*GRN* (n = 8)		Tau (n = 19)		TDP-43 (n = 34)	
GIF GM volume difference (n = 262)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)
Frontal_GM_left	**1.8**	[1.1–2.5]	**1.5**	[1.0–2.3]	0.8	[-0.3–1.9]	**1.8**	[1.3–2.3]	**1.2**	[0.6–1.7]
Frontal_GM_right	**1.4**	[0.8–2.1]	0.9	[0.1–1.5]	0.4	[-0.6–1.2]	**1.3**	[0.7–1.9]	0.8	[0.3–1.2]
Temporal_GM_left	**1.0**	[0.3–1.8]	0.7	[0.1–1.2]	0.5	[-0.3–1.6]	0.9	[0.2–1.4]	0.6	[0.2–1.0]
Temporal_GM_right	**1.8**	[0.6–3.1]	0.7	[0.0–1.2]	-0.3	[-1.1–0.7]	**1.6**	[0.7–2.5]	0.7	[0.3–1.0]
Insula_GM_left	**1.4**	[0.8–2.1]	**1.3**	[0.6–2.0]	0.5	[-0.5–2.1]	**1.3**	[0.8–1.9]	0.9	[0.3–1.4]
Insula_GM_right	**1.4**	[0.8–2.1]	0.9	[0.1–1.9]	-0.4	[-0.9–0.4]	**1.3**	[0.8–1.9]	0.2	[-0.3–0.8]
Cingulate_GM_left	0.6	[-0.1–1.1]	-0.3	[-1.0–0.5]	0.4	[-0.7–2.8]	0.4	[-0.2–1.0]	0.4	[0.0–0.8]
Cingulate_GM_right	0.8	[0.4–1.6]	0.6	[-0.2–0.9]	0.0	[-1.0–1.0]	0.9	[0.5–1.3]	0.5	[0.0–0.8]
Parietal_GM_left	0.8	[0.4–1.3]	0.9	[0.4–1.3]	**1.0**	[0.3–1.6]	0.6	[-0.1–1.2]	**1.0**	[0.7–1.3]
Parietal_GM_right	0.6	[0.0–1.1]	0.8	[0.2–1.2]	0.4	[-0.4–1.2]	0.4	[-0.2–0.9]	0.4	[0.1–0.8]
Occipital_GM_left	0.3	[-0.4–1.0]	0.3	[-0.5–1.1]	0.5	[-0.3–1.3]	0.1	[-0.5–0.7]	0.2	[-0.3–0.5]
Occipital_GM_right	0.0	[-0.7–0.7]	0.3	[-0.6–1.4]	0.2	[-0.7–1.0]	-0.3	[-0.8–0.3]	-0.1	[-0.5–0.3]

* Effect sizes of 1.0 or above are in **bold** to indicate a high degree of group separation from controls

Table 4-7 continued: Effect size calculations (95% BCa confidence intervals) for PieBSI and GIF volume-difference lobar measures for the genetic and patholgoy FTD subgroups

Measure	Genetic subgroups						Pathology subgroups			
	MAPT (n = 16)		*C9orf72* (n = 10)		*GRN* (n = 8)		Tau (n = 19)		TDP-43 (n = 34)	
GIF lobar volume difference (n = 262)	**Effect Size**	**95% CI (BCa)**	**Effect Size**	**95% CI (BCa)**	**Effect Size**	**95% CI (BCa)**	**Effect Size**	**95% CI (BCa)**	**Effect Size**	**95% CI (BCa)**
Frontal_left	**1.0**	[0.4–1.5]	**1.2**	[0.5–2.0]	0.8	[-0.1–2.0]	**1.1**	[0.6–1.6]	**1.0**	[0.6–1.4]
Frontal_right	**1.0**	[0.4–1.7]	**1.1**	[0.3–1.7]	0.6	[-0.5–1.8]	**1.1**	[0.5–1.6]	0.7	[0.3–1.2]
Temporal_left	**1.2**	[0.3–2.0]	0.6	[-0.2–1.1]	0.4	[-0.5–1.2]	**1.0**	[0.3–1.7]	0.7	[0.3–1.0]
Temporal_right	**1.2**	[0.7–1.7]	0.9	[0.4–1.4]	-0.6	[-1.5–0.2]	**1.3**	[0.9–1.8]	0.7	[0.3–1.1]
Insula_left	0.7	[0.1–1.4]	**1.0**	[0.2–2.3]	**1.1**	[0.0–2.2]	0.8	[0.2–1.3]	**1.1**	[0.6–1.5]
Insula_right	0.6	[0.0–1.3]	0.6	[-0.5–1.1]	-0.1	[-0.9–1.2]	0.3	[-0.4–0.9]	0.5	[-0.1–1.0]
Cingulate_left	0.8	[0.3–1.4]	-0.2	[-1.1–0.6]	0.4	[-0.7–2.9]	0.5	[-0.2–1.2]	0.5	[0.0–0.9]
Cingulate_right	0.8	[0.5–2.6]	0.6	[0.0–0.9]	-0.1	[-1.0–0.9]	0.9	[0.6–2.1]	0.4	[0.0–0.8]
Parietal_left	0.4	[-0.2–0.9]	0.9	[0.1–2.0]	0.5	[-0.6–1.8]	0.5	[0.0–1.1]	0.7	[0.2–1.1]
Parietal_right	0.7	[0.1–1.3]	**1.1**	[0.1–2.6]	0.2	[-0.8–1.0]	0.7	[0.1–1.3]	0.5	[0.1–0.9]
Occipital_left	0.1	[-0.5–0.7]	**1.0**	[0.2–1.6]	-0.7	[-1.4–0.1]	0.1	[-0.5–0.5]	-0.1	[-0.5–0.3]
Occipital_right	-0.1	[-0.6–0.6]	0.2	[-0.7–0.9]	-0.7	[-1.5–-0.1]	-0.3	[-0.8–0.2]	-0.2	[-0.5–0.2]
Infratentorial	-0.2	[-0.7–0.4]	0.0	[-0.8–0.9]	0.0	[-0.9–1.1]	-0.1	[-0.6–0.4]	0.0	[-0.4–0.4]

* Effect sizes of 1.0 or above are in **bold** to indicate a high degree of group separation from controls

4.3.4 Sample size

Table 4-8 reports the sample size calculations along with 95% BCa CI to detect a 30% treatment effect with 80% power and 5% statistical significance. These were adjusted for the longitudinal rates of change in the healthy controls for each measure. Sample size estimates under 100 participants required per treatment arm are in blue and bold. Those between 100 and 200 are green and underlined. Measures that provided estimates with a potentially infinite upper 95% BCa CI limit are shown in red. For these measures, multiple independent random repetitions of the 2000 bootstrap samples failed to include patients with significantly different atrophy rates from controls. Therefore, an accurate distribution could not be determined for that measure and subgroup. Estimates with an unfeasible upper limit of over 10,000 individuals required per treatment arm are also shown in red.

4.3.4.1 Clinical subgroups

The PieBSI performed well across the clinical subgroups with many measures producing sample sizes well below 100. In the bvFTD group, the left temporal lobe PieBSI performed best with an estimate of 161 [109–249]. The ratio paired comparison analysis demonstrated this was a significantly lower estimate than provided by any of the other eight PieBSI longitudinal measures at the $p < 0.05$ significance level. The frontal, temporal and parietal PieBSI measures performed well across all PPA subgroups. For svPPA, the left temporal also produced the lowest estimate at 23 [14–37] individuals per treatment arm. The right temporal and left parietal also produced comparatively low estimates of 43 [29–72] and 60 [41–94], respectively. However, the sample size-ratio paired comparisons demonstrated that the left temporal PieBSI measure produced a significantly lower estimate than all PieBSI measures. The pattern in the nfvPPA group was more distributed with the left frontal, left temporal and right parietal longitudinal measures all producing the lowest sample size estimates below 100. The ratio comparisons revealed these were all significantly better than the left and right occipital and the infratentorial measures of change, but not

the contralateral hemisphere within the same lobe or the other frontal, temporal and parietal measures. Despite a strong left-hemisphere dominance in terms of the mean rate of change for the PPA-NOS subgroup, the significantly increased variability that accompanied these elevated rates resulted in comparable bilateral sample size estimates for several lobes. The left and right measures for the frontal and temporal lobes, as well as the left occipital measure, all produced low estimates of less than 100. Again, the left temporal measure produced the lowest estimate at 37 [12–142] per treatment arm. Whilst this was not significantly different from the other four measures under 100, it was the only one of them to demonstrate a significantly lower estimate than both parietal measures and the infratentorial region ($p < 0.05$). The PieBSI longitudinal measure of the left temporal lobe again provided the lowest sample size estimate for the lvPPA subgroup at 37 [22–78]. This was followed by the left parietal lobe (63 [37–132]), right temporal (69 [38–146]) and right parietal lobes at 93 [54–199]. All four of these measures outperformed those for the right frontal lobe, both occipital lobes and the infratentorial measure at the $p < 0.05$ level. Importantly, the left temporal measure outperformed all of the other measures except the left parietal lobe ($p = 0.081$ [-0.03–0.57]).

As with the rates of change and effect sizes used to derive the sample size estimate, results for the GIF GM and full lobar volume differences were considerably more variable, had wider 95% BCa CI and higher sample size estimates overall. As a result, the majority of the head-to-head paired ratio comparison analyses failed to show significant differences between the measures that provided any additional data over and above the PieBSI comparisons discussed above and are thus not included here.

None of the 25 GIF volume measures provided estimates below 200 in the bvFTD group, although the lower 95% BCa CI limit of several of the frontal and temporal measures did fall below 200. The lowest estimate was derived from the left frontal GM measure with 227 required per treatment arm. In the svPPA group, the left frontal (GM) and right temporal measures (GM and

lobar) all produced low estimates of less than 100, with narrow 95% BCa CIs. The lowest was the right temporal GM measure with 59 participants required to detect the putative treatment effect. The left temporal (GM and lobar), left insula (GM and lobar) and right insula (GM only) also showed potential with estimates ranging from 136 for the left frontal lobar measure to 172 for the left temporal GM measure. The nfvPPA showed a good signal in the left frontal (GM and lobar), right frontal (GM only), left temporal (GM and lobar) and left insula (GM only) measures. The best measure was the right frontal lobar change with an estimate of 97 [60–187]. As with the PieBSI results, frontal and temporal measures performed well in the PPA-NOS group. The lowest estimate was again provided by the left temporal lobar measure at 16 [7–33] individuals required per treatment arm to detect a 30% reduction in atrophy. In addition, the GIF results demonstrate the usefulness of measures of the left insula and left cingulate atrophy in this patient population, with estimates ranging from 20 [5–79] in the left insula GM to 64 [26–422] with the cingulate GM measure, although the high upper limit for the latter is worth noting. Few GIF volume-change measures in the lvPPA group provided feasible sample sizes, except for the temporal lobe (bilateral GM and lobar) and parietal (lobar only) measures, with the right temporal lobar measure providing the best estimate at 29 [16–64] participants per treatment arm.

4.3.4.2 Genetic subgroups

In the *MAPT* subgroup, the sample size calculations show a strong frontotemporal signal with all four PieBSI measures of these lobes providing sample sizes of less than 100. Surprisingly, the infratentorial region also provided a comparably low estimate at 72 [38–204]. The ratio comparisons demonstrated that these five measures provided significantly lower sample sizes than all parietal and occipital measures at $p < 0.05$ but did not significantly differ from each other in any paired comparison combination. The right frontal PieBSI provided the lowest sample size for the *C9orf72* subgroup at 175 [80–432]. However, because of the more distributed pattern of atrophy and the higher variability demonstrated by wide and overlapping

BCa CI in this subgroup, the paired comparisons demonstrated that the right frontal-lobe sample size did not significantly differ from any of the other measures at $p < 0.05$. The *GRN* subgroup presented with a left-dominant profile for the sample size estimates, the smallest of which was 53 [24–119] participants per treatment arm when using the left occipital measure. Paired comparison analysis showed that this was not significantly smaller than the estimates produced using the left frontal, left temporal or left parietal lobe measures, which all produced estimates below 130. However, it was significantly lower than the estimates produced from the right frontal, right temporal, right occipital and the infratentorial PieBSI measures. It is important to note that whilst the right occipital PieBSI measure provided a reasonable point estimate of 150 participants required per treatment arm, the upper 95% BCa CI of 11,056 individuals demonstrates the importance of including confidence intervals when providing sample size estimates for a correct interpretation, particularly if based on small cohorts.

Similar to the PieBSI, bilateral frontal and temporal measures of change produced the lowest estimates in the *MAPT* subgroup, particularly for the GM measures. The smallest sample size was for the left frontal GM measure, producing an estimate of 53 [28–132]. In addition, measures of insula GM loss demonstrate promise with estimates of 85 [38–290] for the left and 88 [39–292] for the right. In the *C9orf72* subgroup, bilateral frontal and left insula GM measures provided feasible sample sizes under 200, with the right frontal GM measure providing the lowest estimate at 73 [33–167]. In addition, the right parietal and left occipital lobar measures appear to provide reasonable estimates at 156 and 178, respectively. However, again, the BCa CI demonstrates that the confidence interval should be carefully considered when making a biomarker choice with the 95% upper limits at 16,260 and 3,266 participants, respectively, for these measures. Unfeasible upper limits were also produced for the right frontal GM measure and left insula lobar measure despite feasible point estimates. In contrast to the PieBSI, the GIF volume-difference measures for the *GRN* performed poorly, with the majority of measures indicating an infinite upper 95% BCa CI as a

possibility. In the GM measures, the left parietal region showed some potential with an estimate of 187 [70–1919] and the left insula lobar measure producing an estimate of 137. However, the confidence interval for this latter measure was [37–122,877], indicating it would be an unreliable marker of change.

4.3.4.3 Pathology subgroups

The Tau subgroup presented with the same frontotemporal and infratentorial pattern of sample size estimates as the *MAPT* group, with all four parietal and occipital measures providing significantly higher estimates and no differences amongst the five best performing measures. In the TDP-43 subgroup, the left temporal lobe measure performed best with an estimate of 105, followed by the right temporal, left occipital and left frontal PieBSI measures. The sample size estimates derived from these four measures did not significantly differ from each other but were significantly smaller than the right frontal, both parietal, right occipital and infratentorial estimates (all $p < 0.05$).

For the sample sizes derived from the GIF segmentations, bilateral frontal, temporal and insula measures performed well in the Tau pathology group, particularly the GM regions. The left frontal GM measure provided the lowest estimate at 56. For the TDP-43 subgroup, left frontal (GM and lobar), left insula (lobar only) and left parietal (GM only) provided estimates below 200. Again, the left frontal GM measure produced the smallest sample size estimate with 113 participants required per treatment arm.

Table 4-8 Sample size estimates with 95% BCa CI to detect a 30% reduction in atrophy rate with 80% power and 5% significance for lobar measures split by FTD subgroups. Results in **blue** and bold are for measures that required <100 participants per treatment arm, green and underlined demonstrate sample size estimates between 100 and 200 individuals per treatment arm and those highlighted in red have an infinite or unfeasible upper 95% confidence interval produced using 2000 bootstrap replications.

Measure	Clinical subgroups									
	bvFTD (n = 64)		svPPA (n = 44)		nfvPPA (n = 45)		PPA-NOS (n = 7)		lvPPA (n = 21)	
PIEBSI (n = 258)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (Bca)
Frontal_left	245	[174–386]	120	[61–243]	93	[62–141]	74	[27–306]	116	[53–295]
Frontal_right	227	[162–339]	193	[92–396]	100	[69–148]	63	[20–314]	196	[85–625]
Temporal_left	161	[109–249]	23	[14–37]	76	[54–114]	37	[12–142]	37	[22–78]
Temporal_right	212	[146–335]	43	[29–72]	100	[61–178]	72	[29–184]	69	[38–146]
Parietal_left	283	[179–540]	60	[41–94]	113	[72–184]	146	[70–359]	63	[37–132]
Parietal_right	327	[204–640]	143	[92–268]	88	[55–155]	238	[79–737]	93	[54–199]
Occipital_left	307	[179–721]	135	[81–279]	176	[100–451]	56	[25–166]	140	[74–411]
Occipital_right	360	[203–987]	370	[188–1266]	221	[106–833]	185	[56–9131]	218	[90–1069]
Infratentorial	270	[162–575]	730	[267–9319]	183	[100–552]	135	[44–1424]	484	[200–4280]

Table 4-8 continued: Sample size estimates with 95% BCa CI to detect a 30% reduction in atrophy rate with 80% power and 5% significance for lobar measures split by FTD subgroup

Measure	Clinical subgroups									
	bvFTD (n = 64)		svPPA (n = 44)		nfvPPA (n = 45)		PPA-NOS (n = 7)		lvPPA (n = 21)	
GIF GM volume difference (n = 262)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (Bca)
Frontal_GM_left	227	[113–633]	97	[58–219]	178	[78–560]	69	[12–37962]	404	[164–2494]
Frontal_GM_right	246	[131–620]	620	[256–8394]	275	[115–1303]	190	[43–88360]	790	[393–∞]
Temporal_GM_left	393	[218–1139]	172	[104–421]	113	[65–269]	38	[14–74]	104	[25–236]
Temporal_GM_right	319	[176–668]	59	[37–137]	693	[217–8741]	92	[9–776]	118	[65–278]
Insula_GM_left	614	[254–2893]	143	[86–383]	259	[127–923]	20	[5–79]	269	[104–3608]
Insula_GM_right	2246	[416–1.1e+06]	150	[83–400]	338	[175–1147]	862	[102–∞]	521	[190–∞]
Cingulate_GM_left	2151	[621–9.6e+05]	259	[135–1775]	373	[191–1456]	64	[26–422]	1220	[391–∞]
Cingulate_GM_right	1119	[416–21363]	463	[194–2170]	1346	[306–∞]	913	[106–∞]	3399	[475–∞]
Parietal_GM_left	947	[366–11141]	1024	[133–74133]	344	[138–1841]	143	[66–1682]	2837	[95–∞]
Parietal_GM_right	1458	[462–50530]	33139	[3793–∞]	502	[170–6349]	659	[98–∞]	557	[129–2.2e+06]
Occipital_GM_left	1980	[610–2.1e+05]	46880	[10461–∞]	4187	[589–∞]	538	[80–∞]	7322	[1062–∞]
Occipital_GM_right	4717	[918–4.8e+05]	83757	[12745–∞]	44340	[1968–∞]	1027	[409–∞]	3492	[1819–∞]

Table 4-8 continued: Sample size estimates with 95% BCa CI to detect a 30% reduction in atrophy rate with 80% power and 5% significance for lobar measures split by subgroup

Measure	Clinical subgroups									
	bvFTD (n = 64)		svPPA (n = 44)		nfvPPA (n = 45)		PPA-NOS (n = 7)		lvPPA (n = 21)	
GIF lobar volume difference (n = 262)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (Bca)
Frontal_left	362	[196–1097]	136	[71–374]	154	[83–415]	72	[26–337]	1000	[129–∞]
Frontal_right	370	[196–1067]	1339	[374–∞]	97	[60–187]	110	[36–35325]	2447	[171–∞]
Temporal_left	412	[213–1184]	140	[82–283]	186	[90–615]	16	[7–33]	111	[35–286]
Temporal_right	318	[176–824]	85	[45–243]	602	[165–10207]	37	[17–152]	29	[16–64]
Insula_left	1028	[310–18705]	157	[87–386]	194	[106–547]	51	[25–96]	291	[112–2805]
Insula_right	1614	[460–∞]	664	[164–7.7e+08]	471	[192–2965]	229	[77–∞]	1308	[256–∞]
Cingulate_left	1256	[438–18348]	243	[107–1431]	346	[200–874]	60	[26–151]	1366	[359–∞]
Cingulate_right	897	[349–12099]	307	[139–968]	949	[243–76130]	369	[87–∞]	2343	[414–∞]
Parietal_left	751	[290–9969]	224	[121–1814]	301	[121–1814]	175	[121–1814]	162	[121–1814]
Parietal_right	480	[218–2160]	7658	[728– ∞]	766	[227–23015]	734	[163–∞]	175	[75–817]
Occipital_left	35226	[1169–∞]	4646	[1886–∞]	473268	[1886–∞]	1236	[1886–∞]	1026	[1886–∞]
Occipital_right	35157	[1288–∞]	93187	[2742–∞]	18135	[2742–∞]	790	[2742–∞]	6679	[2742–∞]
Infratentorial	62527	[1344–∞]	1345	[308–∞]	1743	[308–∞]	14484	[308–∞]	74577	[308–∞]

Table 4-8 continued: Sample size estimates with 95% BCa CI to detect a 30% reduction in atrophy rate with 80% power and 5% significance for lobar measures split by subgroup

Measure	Genetic subgroups						Pathology subgroup			
	MAPT (n = 16)		*C9orf72* (n = 10)		*GRN* (n = 8)		Tau (n = 19)		TDP-43 (n = 34)	
PIEBSI (n = 258)	**Sample size**	**95% CI (BCa)**	**Sample size**	**95% CI (BCa)**	**Sample size**	**95% CI (BCa)**	**Sample size**	**95% CI (BCa)**	**Sample size**	**95% CI (Bca)**
Frontal_left	**72**	**[40–172]**	239	[107–674]	127	[47–894]	**85**	**[50–170]**	174	[103–319]
Frontal_right	**84**	**[47–225]**	175	[80–432]	213	[67–873]	**67**	**[38–159]**	209	[137–390]
Temporal_left	**68**	**[40–142]**	202	[92–546]	109	[37–545]	**79**	**[47–142]**	105	[62–198]
Temporal_right	**83**	**[49–192]**	261	[100–1073]	285	[141–1966]	**87**	**[53–177]**	125	[71–267]
Parietal_left	548	[218–14829]	366	[130 - 3087]	128	[56–367]	287	[113–2130]	215	[131–422]
Parietal_right	376	[174–2376]	423	[142–6234]	223	[124–708]	199	[87–1011]	284	[172–800]
Occipital_left	552	[202–10587]	509	[178–3105]	**53**	**[24–119]**	547	[204–6803]	134	[80–278]
Occipital_right	523	[189–17953]	1249	[205–1.9e+05]	150	[64–11056]	676	[214–25388]	282	[141–930]
Infratentorial	**72**	**[38–213]**	589	[175–88309]	297	[48–73842]	**77**	**[41–166]**	468	[203–2039]

Table 4-8 continued: Sample size estimates with 95% BCa CI to detect a 30% reduction in atrophy rate with 80% power and 5% significance for lobar measures split by subgroup

Measure	Genetic subgroups						Pathology subgroup			
	MAPT (n = 16)		*C9orf72* (n = 10)		*GRN* (n = 8)		Tau (n = 19)		TDP-43 (n = 34)	
GIF GM volume difference (n = 262)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (Bca)
Frontal_GM_left	53	[28–132]	73	[33–167]	270	[48–87302]	56	[33–105]	113	[59–438]
Frontal_GM_right	84	[38–299]	197	[73–11229]	1382	[125–∞]	107	[46–331]	267	[115–1611]
Temporal_GM_left	163	[56–1769]	352	[123–69400]	649	[65–∞]	228	[85–3512]	430	[184–3086]
Temporal_GM_right	56	[18–467]	398	[126–∞]	2294	[358–∞]	66	[28–378]	387	[166–1699]
Insula_GM_left	85	[38–290]	101	[43–518]	711	[39–∞]	98	[50–250]	241	[93–1774]
Insula_GM_right	88	[39–292]	228	[48–10944]	1113	[1023–∞]	105	[48–309]	3107	[258–∞]
Cingulate_GM_left	548	[134–∞]	1900	[713–∞]	1159	[23–∞]	1032	[169–∞]	1089	[244–∞]
Cingulate_GM_right	275	[65–875]	460	[198–∞]	909703	[168–∞]	223	[98–628]	816	[261–∞]
Parietal_GM_left	254	[101–1319]	224	[99–1004]	187	[70–1919]	516	[130–∞]	177	[110–345]
Parietal_GM_right	532	[141–8.9e+05]	284	[114–5876]	1053	[129–∞]	1170	[227–∞]	872	[306–28597]
Occipital_GM_left	1783	[178–∞]	1560	[154–∞]	715	[99–∞]	14219	[332–∞]	7360	[703–∞]
Occipital_GM_right	13551312	[330–∞]	2378	[88–∞]	6697	[162–∞]	1960	[1972–∞]	19459	[1923–∞]

Table 4-8 continued: Sample size estimates with 95% BCa CI to detect a 30% reduction in atrophy rate with 80% power and 5% significance for lobar measures split by subgroup

Measure	Genetic subgroups						Pathology subgroup			
	MAPT (n = 16)		*C9orf72* (n = 10)		*GRN* (n = 8)		Tau (n = 19)		TDP-43 (n = 34)	
GIF lobar volume difference	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (BCa)	Sample size	95% CI (Bca)
Frontal_left	181	[76–1129]	128	[42–662]	244	[43–71466]	147	[70–518]	163	[87–568]
Frontal_right	165	[62–1243]	150	[62–1567]	506	[51–∞]	157	[70–671]	326	[128–2116]
Temporal_left	129	[44–1852]	509	[157–∞]	1354	[117–∞]	177	[61–1653]	364	[161–1995]
Temporal_right	120	[60–336]	208	[93–1019]	545	[2930–∞]	96	[55–231]	325	[136–1553]
Insula_left	329	[87–17617]	184	[34–4362]	137	[37–1.2e+05]	298	[102–4235]	143	[73–429]
Insula_right	495	[104–∞]	538	[138–∞]	10430	[116–∞]	2343	[212–∞]	786	[189–∞]
Cingulate_left	242	[88–1753]	3623	[562–∞]	985	[21–∞]	780	[123–∞]	797	[212–6.2e+05]
Cingluate_right	288	[26–817]	452	[214–8.5e05]	10250	[199–∞]	236	[41–460]	907	[264–∞]
Parietal_left	941	[121–1814]	231	[45–57309]	757	[121–1814]	616	[121–1814]	369	[121–1814]
Parietal_right	366	[97–16224]	156	[26–16260]	4914	[167–∞]	373	[97–17483]	767	[236–59948]
Occipital_left	23684	[1886–∞]	178	[71 - 3266]	374	[1886–∞]	41530	[1886–∞]	9184	[1886–∞]
Occipital_right	18021	[2742–∞]	3690	[2742–∞]	336	[2742–∞]	1525	[2742–∞]	5106	[2742–∞]
Infratentorial	4702	[308–∞]	93842	[308–∞]	264092	[308–∞]	8292	[308–∞]	3768529	[308–∞]

4.4 Discussion

The current study investigated the utility of fully automated measures of lobar atrophy across the FTD spectrum. Rates of GM and full GM/WM lobar regional change were successfully measured using both direct and indirect longitudinal methodologies. As with the previous whole-brain chapter, the direct PieBSI technique provided a more robust measure of change across all nine lobar measures and ten FTD subgroups. The benefit of the indirect GIF volume measure was the ability to investigate the insula and cingulate regional change not currently possible with the PieBSI. Both regions have previously been shown to be heavily involved across the FTD spectrum, particularly in bvFTD and later in the disease course for svPPA and nfvPPA (Schroeter et al., 2007). Regional atrophy was detected across all lobar measures, with differing profiles and variability between measures and subgroups. Many of these rates of change produced large effect sizes and small feasible sample sizes under 100 participants required to detect a 30% treatment effect with 80% power and 5% statistical significance in a putative intervention trial. Temporal lobe measures proved to be particularly useful.

Overall, the pattern of lobar change across the clinical, genetic and pathology groups was consistent with previous literature and generally supported the regional measures hypothesised that would be useful in the Introduction. The bvFTD presented with the most heterogeneous rates in the clinical subgroups, with frontal and temporal regions providing the best measures. However, the current study failed to detect expected insula atrophy with the indirect GIF GM or lobar measures. As relatively smaller regions, they were possibly disproportionately affected by segmentation error, which mainly involved misclassification of the GM boundary. So, despite an overall mean rate of 3.0% and 2.5% change for the left and right insula GM measures respectively, the high standard deviations resulting from these segmentation errors masked the potential signal from this region. Application of the double-window BSI (Hobbs et al., 2009) to masks derived from these GIF insula segmentations may provide a more robust measure in the future to detect this potential signal.

SvPPA had the highest and most consistent rates of change bilaterally in the temporal lobes as predicted. Despite a characteristically left-dominant profile across the majority of lobes, the high rates in the right temporal lobe found here are in line with previous investigations (Bocchetta, Iglesias, Russell et al., 2019; Rohrer et al., 2008) and the established understanding that the contralateral hemisphere (irrespective of initial dominant side) increasingly becomes involved as the disease progresses. Aside from Rohrer and colleagues (2008), additional rates of left and right temporal lobe atrophy in svPPA patients have yet to be published. The majority of longitudinal studies investigate patterns not rates of regional temporal lobe involvement. The few that have reported rates either combined the left and right into a single temporal lobe metric (Binney et al., 2017) or include the svPPA with other clinical phenotypes and present only an FTD rate (Pankov et al., 2016) making in-depth comparisons difficult. The results of these two studies are summarised in Table 4-1. It is reassuring to note that the sample sizes reported using the PieBSI in this study are as low, if not lower than previously published rates at 23 and 43 for the left and right temporal lobe, respectively. As the disease progresses, atrophy spreads from the initial focal temporal profile, moving both anteriorly to include orbitofrontal, insular and anterior cingulate cortices, as well as posteriorly to include temporoparietal regions (Schroeter et al., 2007). This pattern was also demonstrated in the current findings, with measures of the frontal, insula and parietal atrophy, particular left-sided also showing promise as biomarkers for svPPA. The mean disease duration for the current svPPA cohort at baseline was 4.7 (1.9) years, with individuals ranging from 1.5 to 10.2 years. Whilst staging in FTD remains an open question, it is likely this svPPA cohort covers a wide range of disease severity and thus more dispersed regions, typically involved later in the disease were detected. Anterior cingulate atrophy is also commonly reported as the disease progresses (Convery et al., 2019; Montembeault, Brambati, Gorno-Tempini, & Migliaccio, 2018). Whilst both GIF GM and lobar measures of the cingulate were significantly different to controls ($p < 0.001$ for the left: $p < 0.015$ for

the right), these did not translate into as feasible sample sizes as the frontal, temporal, insula and parietal measures. This is likely due to substantially less atrophy in the posterior region diluting the signal of the anterior cingulate and again the issues discussed with increased noise and susceptibility to segmentation errors inherent in indirect longitudinal measures of change.

In line with the current nfvPPA literature, results for the left temporal, left frontal and left parietal measures all provided small sample sizes $\leq$ 100 individuals per treatment arm in a putative trial. The PieBSI performed the best across these regions. The profile of disease progression demonstrates an increasing contralateral spread to analogous regions in the right hemisphere (Gorno-Tempini et al., 2004; Rogalski et al., 2011), which was also demonstrated in the current study. In addition, insula involvement is also reported (Grossman, 2012). Whilst the regression analysis reveals significantly increased insula rates of change compared with control rates ($p < 0.001$ for left GM, right GM and left lobar measures; $p < 0.003$ for right lobar measure), the adjusted mean differences were amongst the lowest, resulting in poorer group separation and higher sample sizes. The lowest sample size for the insula was for the left lobar measure with 194 [106–547]. This has the potential for improvement if included in a direct method of longitudinal change. The currently reported sample sizes are comparable to the frontal and temporal rates reported by Binney and colleagues (2017) and substantially lower than both frontal measures reported by Staffaroni and colleagues (2019), summarised in Table 4-1. Rates of parietal and insula atrophy with accompanying sample size calculations have yet to be published in nfvPPA, but the current study suggests they have potential as imaging biomarkers in this population.

To date, there is no published literature investigating the global or lobar rates of atrophy in PPA-NOS patients. The current study demonstrates several findings that may prove useful in developing a neuroanatomical picture of these patients. As discussed in the previous chapter, whole-brain

longitudinal investigations produced consistently high rates of annual volume loss using multiple automated measures of change (see Table 3-4). The lobar rates of change in the current study suggest a left-dominant presentation, particularly in the temporal lobe. The disease duration of the cohort was 2.4 years at baseline, ranging from 0.2 to 4.2 years since diagnosis. Taking into account the issue of variable and often prolonged time from symptom onset to diagnosis seen in early-onset and rare dementias (Draper et al., 2016; van Vliet et al., 2013), along with variable trajectories of decline (Whitwell et al., 2015), I believe it is reasonable to say, based on the recorded disease duration, that the cohort represents a spectrum from relatively early to more moderately advanced stages of the disease. Figure 4-3 presents the individual PPA-NOS patient's PieBSI lobar rates of change. They are numbered sequentially based on disease duration at baseline. At the individual level, the left-dominant pattern in the temporal, as well as frontal and parietal lobe, can be clearly seen. However, given there are only seven individuals in this cohort and the variability of their rates when combined, the sample size calculations at the group level did not consistently demonstrate this pattern across the lobes, although the left insular GM and left temporal PieBSI produced the lowest estimates of 20 [5–79] and 37 [12–142], respectively. The high rates of volume loss in this patient population resulted in both the direct and indirect longitudinal measures providing low estimates for bilateral frontal, bilateral temporal, left insula (GM and lobar), left cingulate (GM and lobar), left parietal and left occipital regions. It is important to note that two of the PPA-NOS patients also tested positive for the *GRN* mutation (PPA-NOS_1 and PPA-NOS_5), consistent with this more dispersed asymmetric presentation. These initial findings will need validation in a larger cohort of PPA-NOS patients. In addition, PPA-NOS patients do not present with a characteristic profile of language deficits because they are predominantly defined based on failing to fulfil the diagnostic criteria of the other PPA syndromes. It will, therefore, be of value to investigate the relationship of the pattern of regional volume loss and primary language deficit to add to this emerging understanding of the neuroanatomical profile in PPA-NOS.

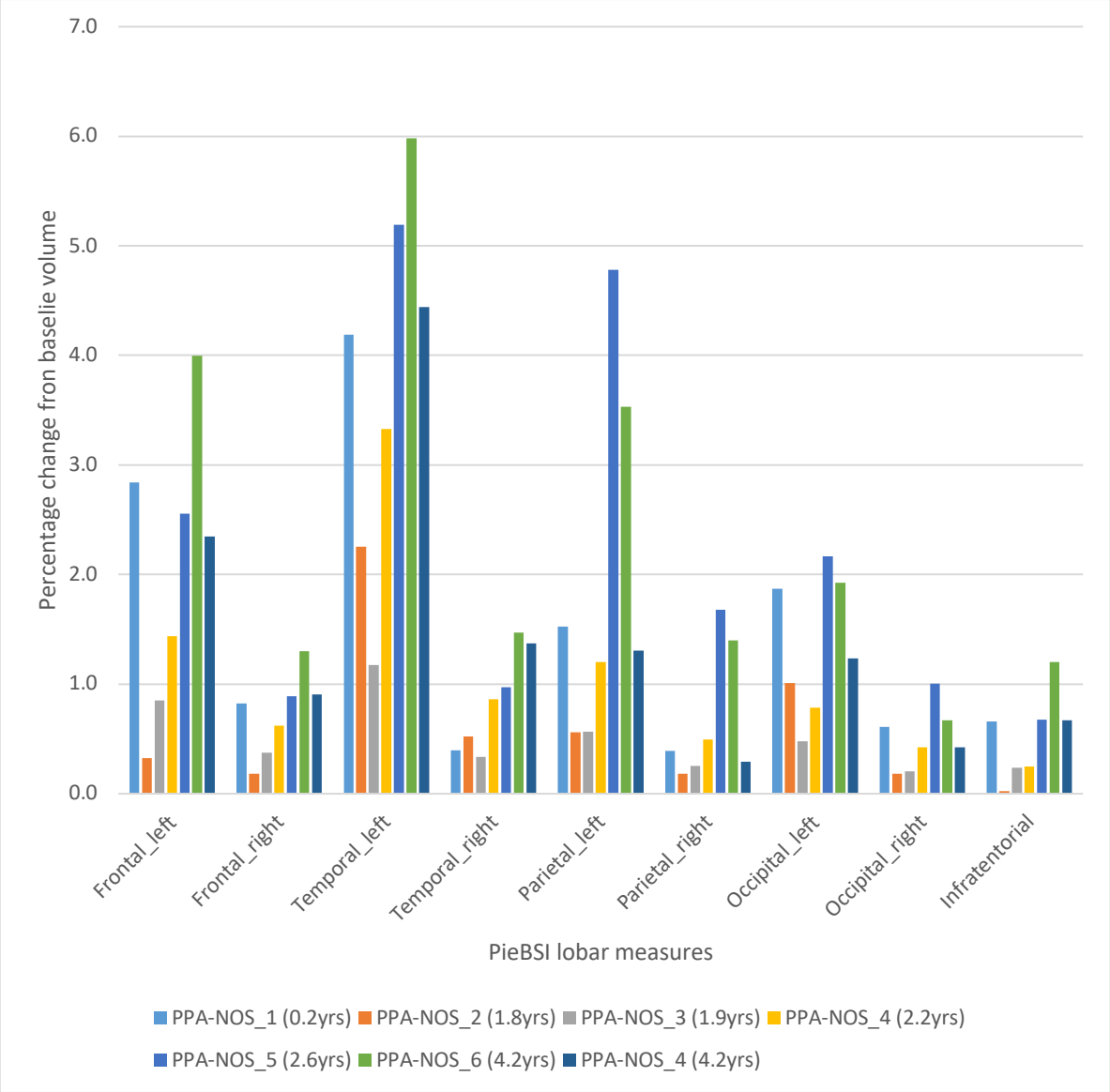

Figure 4-3 Lobar PieBSI results, expressed as an annual percentage change from baseline volume for the individual PPA-NOS patients (disease duration in brackets for each patient). PPA-NOS_1 and PPA-NOS_5 tested positive for a GRN mutation.

Unlike the other PPA subgroups, the lvPPA group presented with a symmetric pattern of atrophy in the temporoparietal regions, consistent with the fact that the underlying pathology is most likely AD (Mesulam et al., 2008). The smallest sample size was derived using the right temporal GIF lobar measure at 29 [16–64]. This was closely followed by the PieBSI temporal and parietal bilateral estimates ranging from 37 [22–78] to 93 [54–199] for the left temporal and right parietal, respectively. There are no previously published sample size calculations in an lvPPA population for

comparison, but it is promising that fully automated measures of these four lobar regions in the current cohort all produce low sample size estimates to detect a 30% treatment effect.

In the genetic subgroups, the best potential candidate measures varied between groups in a manner broadly predicted base on previous literature. The *MAPT* subgroup showed a strong focal symmetric frontotemporal signal with all four measures of these lobes providing sample sizes of less than 100 with the PieBSI. The indirect measures of the insula GM bilaterally also produced small sample size estimates below 100, despite the increased noise inherent in these indirect longitudinal measures. This suggests measures of the insula would be promising candidates, particularly if refined using a direct measure of the change in this region. A surprising result came from the infratentorial PieBSI, which provided a comparably low estimate at 72 [38–204]. Reduced volume within the cerebellum has previously been reported as a feature of *C9orf72* (Cash et al., 2018; Mahoney et al., 2012), however, there is some indication that the vermis may also be reduced in *MAPT* patients (Bocchetta et al., 2016). Upon closer inspection of these scan pairs, the cerebellum was the most susceptible region to show acquisition-related geometric distortion upon visual review of the registrations. Assuming proper positioning in the scanner, the cerebellum is the structure furthest away from the magnet isocentre and thus more prone to effects of gradient non-linearity during acquisition. During the affine registration process, the cerebellum is further away from the centre of rotation defined by the transformation parameters compared with other more medial structures. Therefore, it is proportionally more affected by the rotation and shearing stages of the registration if the patient was positioned substantially differently for the acquisition of the two native space images. The results from the GIF lobar volume change that do not include any transformation suggest very little change in this region. Finally, the data suggest that GM changes may prove a more valuable measure than the full lobe in this subgroup. Previous work has shown WM damage in *MAPT* patients is substantially less than in *C9orf72* (Jiskoot et al., 2018); however,

the relative value of GM versus lobar regional change will need to be further investigated. To date, only one study has published lobar sample size calculations in the three main genetic populations (Whitwell et al., 2015) reporting an estimate of 205 for a combined frontal measure and 152 for temporal measures in their *MAPT* cohort (Table 4-1). Without confidence intervals, it is difficult to make an accurate comparison, but the current results suggest the PieBSI improved performance as a longitudinal lobar measure by providing a substantially lower estimate for these regional measures across the genetic FTD subgroups.

Consistent with previous predominantly cross-sectional studies, the *C9orf72* patients presented with a more distributed pattern of atrophy. Previous limited longitudinal work suggests they may have the slowest rate of global change compared with the other two mutation types (Gordon et al., 2016), which was also demonstrated in the previous chapter for all nine whole-brain BSI longitudinal methodologies (Table 3-4). Broadly speaking, the lowest sample size estimates were derived from frontal (right PieBSI and bilateral GM and lobar GIF), left insula (GM and lobar GIF), right parietal (GIF lobar) and left occipital (GIF lobar) measures, all reported as under 200 participants required per treatment arm. However, the 95% BCa CI upper limit for many of these (e.g. 11,229 for right frontal GM or 16,260 for right parietal GIF lobar measure) demonstrates caution is necessary when interpreting the point estimate on its own and these measures would likely be poor candidates for detecting treatment effects. The lowest estimate with a reasonable confidence interval was derived from the right frontal GM measure with 73 [33–167]. Although not analogous to Whitwell and colleagues (2015) combined frontal lobar measure that produced an estimate of 122 participants, it does indicate regional frontal measures may be of value. However, given their characteristically distributed profile of atrophy, results from the current cohort suggest that global measures may provide better markers in *C9orf72* patients with the GIF whole-brain KBSI resulting in a sample size estimate of 156 [55–1573]. Further work will be of value to improve and validate regional measures in *C9orf72* patients.

Despite historically demonstrating the fastest rate of global atrophy (which was also consistent with the results in the previous chapter), the regional measures in *GRN* failed to provide as reliable a marker of change based on sample size calculations. The measures demonstrated a clear left-dominant profile, with the best candidate measure derived from the left occipital PieBSI at 53 [24–119]. The other measures providing estimates between 100 and 200 individuals per treatment arm were left frontal PieBSI, left temporal PieBSI, left insula GIF lobar, left parietal (PieBSI and GM) and right occipital PieBSI with estimates of 127, 109, 137, 128, 187 and 150, respectively. Whilst all these point estimates appear reasonable and comparable when reviewed on their own, incorporating the BCa CI data shows a different story. With this additional valuable information, it can be seen that the upper limit of the left insula reached 122,877 and the right occipital PieBSI was 11,056, demonstrating these measures were performing substantially worse at differentiating patient and control populations based on atrophy rates. Overall, the GIF volume-difference measures performed very poorly in the *GRN* group with the majority having an infinite upper limit to detect a treatment effect. Upon further visual inspection, the significant atrophy in these patients evident at baseline meant segmentation error had a disproportionately high impact on the measures. Reassuringly, the application of the direct longitudinal PieBSI measure corrected much of this error to provide a more robust measure of the change in these patients.

It is important to note that these *GRN* patients had varying clinical presentations, contributing to the variability of regional measurements in this subgroup. Four patients were bvFTD, two presented with a nfvPPA phenotype and two were classified as PPA-NOS at the time of scanning. This lack of a one-to-one relationship between clinical, genetic and pathological presentations again highlights the complexity and key challenges in FTD research and why continued investigations in larger comprehensive cohorts are important to better understand and model these

connections. One of the contributors to the poor performance of these lobar measures at the group level was due to the differing neuroanatomical profiles of the eight *GRN* patients on the individual level as presented in Figure 4-4. Therefore, it may be important to incorporate clinical presentation or weight biomarkers choice based on symptomatic and therefore likely underlying neuroanatomical profiles in future investigations.

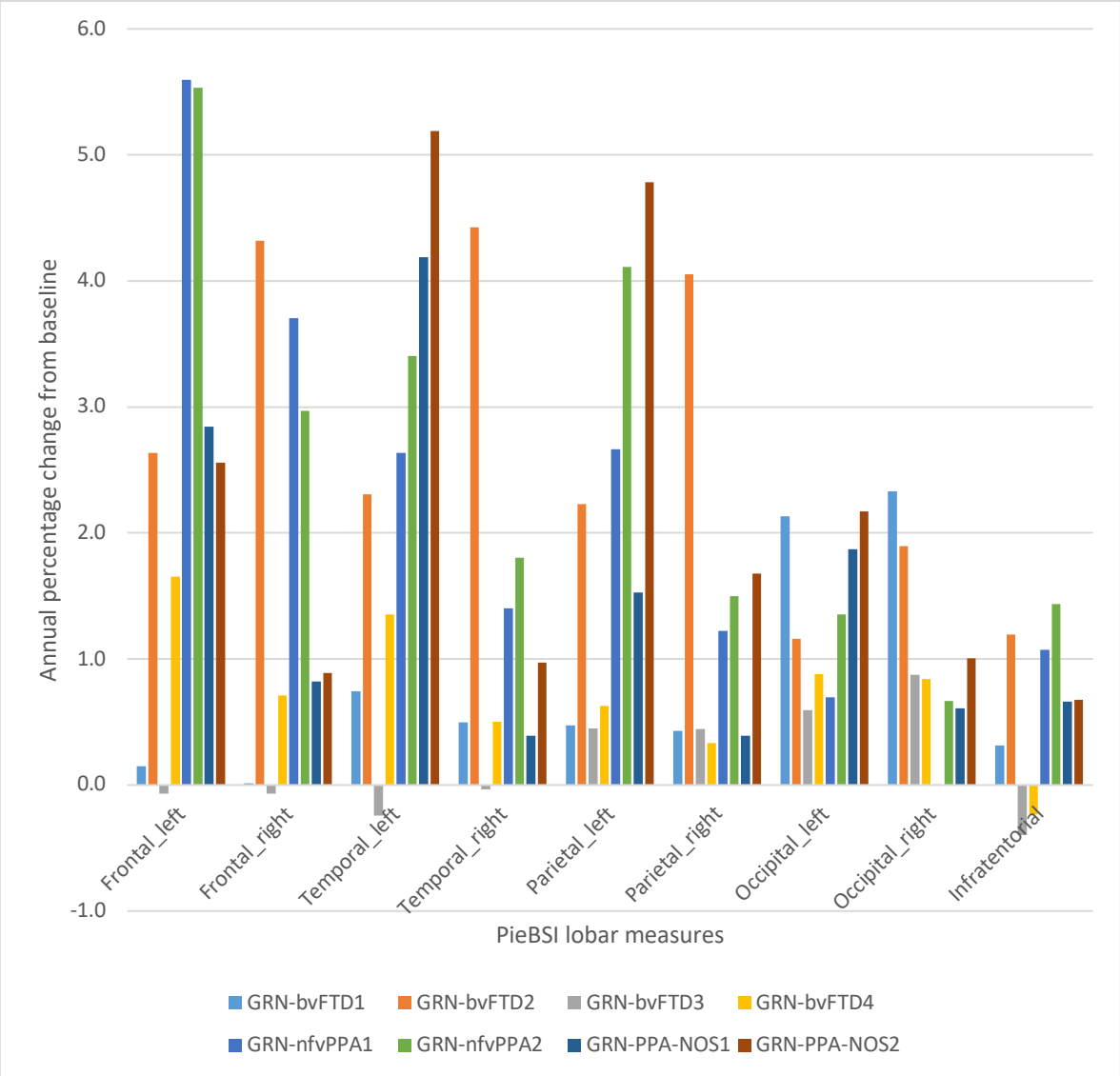

Figure 4-4 Annual PieBSI lobar change for each of the *GRN* patients, demonstrating the different atrophy patterns across the subgroup. Clinical diagnosis included for each patient.

The current study has several caveats to consider when interpreting the results. First, the PieBSI is a recently developed technique that is still under improvement and has no previously published validation or data in neurodegenerative or healthy cohorts. Therefore, it is not clear whether the

technique may be over- or underestimating regional change or producing errors in terms of the lobar delineation. Despite these potential uncertainties, visual inspection of the delineations failed to find any lobar parcellation that did not accurately reflect each of the nine lobar sectors. In addition, all nine regional BSI results summed to an overall whole-brain BSI result that closely matched the GIF KBSI results reported in Chapter 3, providing support for the current findings. An additional caveat relates to the genetic cohorts being relatively small. Whilst they provided robust global measures, the regional measures were less consistent, and the bootstrapping provided relatively wide BCa CI for several of the measures.

An important issue highlighted by the current results relates to reporting and interpreting of sample sizes. Issues with the reporting of sample sizes have been previously raised in the literature, demonstrating that many articles in high-impact medical journals failed to provide adequate or accurate reporting (Charles et al., 2009). The current dataset has multiple examples of sample size estimates within the clinical, genetic and pathology groups where the point estimate appears feasible but have an unobtainable upper 95% BCa CI limit. For example, in PPA-NOS the right frontal GM measure provided an estimate of 190, and the left parietal GIF lobar measure provided an estimate of 175. If the interpretation of biomarker feasibility was based solely on these estimates, it would be reasonable to conclude that they may be comparable. However, including information about the likely confidence interval of these estimates shows how misleading reliance on the estimate alone can be. The BCa CI for the right frontal GM is [43–88,360] compared with [121–1814] for the left parietal GIF lobar measure, demonstrating substantial difference in the utility of these measures despite initial point estimate similarity.

Only one of the publications summarised in Table 4-1 (Staffaroni et al., 2019), calculated confidence intervals for their sample size estimates, provided separately in the supplementary data. These showed that several of the calculations had wide corresponding confidence intervals, with the

largest for the right temporal measure in the nfvPPA group ranging from 344 to 58,677 to detect a 25% treatment effect using the right temporal lobe measure. This lack of calculating or reporting confidence intervals makes interpretation and comparison of previously published results problematic, particularly if calculations were based on smaller cohorts as is common in rare conditions such as FTD. I would argue that the current data highlight the importance of including confidence intervals for sample size calculations as standard given that they provide a much more realistic result to inform the design of trials, biomarker choice and to accurately plan recruitment. This will be essential to avoid the costly and ethical issues related to underpowered trials.

The current set of results demonstrates the potential of regional longitudinal measures in FTD. However, additional work is required to confirm and extend the current lobar findings. In particular, investigating additional direct measures of change may prove beneficial to improve the signal-to-noise ratio of some measures and provide a better understanding of the potential of different regional measures across the different subgroups. Application of the double-window BSI (Hobbs et al., 2009) to the summed lobar GIF segmentations for these regions will be valuable in investigating the potential improvement in robustness of these regional measures and a corresponding reduction in sample sizes. In addition, applying the GIF lobar masks to the longitudinal SPM deformation maps derived in the previous chapter may also provide improved lobar measurements less prone to segmentation error and provide more robust measures of lobar change within FTD.

The current chapter provides support for the use of fully automated regional measures as potential biomarkers across the FTD spectrum. Many of the frontal and temporal PieBSI results produced estimates well below 100 participants required per arm to detect a meaningful treatment effect across the PPA subgroups, the *MAPT* and the Tau subgroups. Regional measures may likely prove best as a supplement to, instead of replacement for, the

more established global measures as is the case in AD trials, which generally include whole-brain, ventricular and hippocampal measures as standard. The utility of some key subcortical regions will be investigated in the following chapter to ascertain if the FTD subgroups may have an analogous additional subcortical biomarker to the hippocampus in AD and the caudate for HD.

5 Automated subcortical measures of change in FTD

In the previous two data chapters, several fully automated global whole-brain and regional lobar measures were investigated, with many showing promise as biomarkers for use in upcoming and putative therapeutic interventions across the FTD spectrum. An important finding from these chapters was that optimal biomarker choice (both region and method of longitudinal measurement) varied based on the FTD subgroup being investigated. It is possible that combining a global measure alongside a regional measure in place of a single outcome may prove useful to detect treatment effects. What remains unclear in FTD is whether measures of subcortical structures may also prove valuable in a similar way as that demonstrated for the hippocampus in AD trials and caudate measures in HD. In addition, it has yet to be investigated whether different subcortical structures prove more or less useful depending on the FTD subgroup of interest. Therefore, the focus of the current project was to investigate fully automated measures of four key subcortical structures across the FTD spectrum including the amygdala, the hippocampus, the caudate and the thalamus. Both indirect volume-difference and direct BSI measures of change were investigated for each hemisphere separately.

5.1 Introduction

Although the majority of structural neuroimaging research has focused on cortical volume-loss measures, more recent imaging and neuropathological research support the importance of subcortical involvement in FTD. Reduced volumes in subcortical structures have been shown across much of the FTD spectrum when compared with control subjects both symptomatically (Broe et al., 2003; Garibotto et al., 2011; Kril & Halliday, 2004; Rohrer et al., 2011; Schroeter et al., 2007; Seeley et al., 2008) and presymptomatically in mutation carriers (Rohrer et al., 2015). However, systematic longitudinal investigations of rates of change in subcortical structures are currently lacking in FTD and assessing their utility as possible outcome measures for therapeutic trials is necessary and increasingly

pressing as growing numbers of promising candidate interventions move in clinical trials in humans.

5.1.1 The amygdala in FTD

The amygdala is an important part of the limbic system involved in emotion, threat detection, reward evaluation, motivation and several other cognitive functions such as attention and perception that are often affected in FTD. Within the clinical subgroups, amygdalar atrophy is reportedly more pronounced in svPPA patients than bvFTD and nfvPPA (Garibotto et al., 2011; Whitwell et al., 2005). Reported volume reduction in svPPA compared with controls range from 51–65% on the left and 33–54% on the right, demonstrating the asymmetry characteristic of this group (Chan et al., 2001; Lehmann et al., 2010; Rohrer et al., 2008; Whitwell et al., 2005). Although less pronounced, investigations have reported that amygdalar volumes in bvFTD are approximately 22% smaller on the left and 19% smaller on the right than controls (Boccardi et al., 2002). Longitudinal investigations have been limited and demonstrate inconsistent results. To date, only two studies have reported rates of change involving the amygdala in bvFTD (Frings et al., 2014; Landin-Romero & Piguet, 2017). Frings and colleagues (2014) report a mean (sd) annual rate of change of 3.6 (3.4)%; however, the atlas used for the results combined both hippocampal and amygdalar regions for this calculation, making it difficult to interpret the contribution of the amygdala alone (or either hemisphere). Landin-Romero and colleagues (2017) also report that rates of amygdalar atrophy in bvFTD were significantly higher than controls with an annual percentage change of 2.1% (a result calculated from the supplementary data and additional input following email correspondence with the author). No other longitudinal investigations of amygdalar atrophy in any of the FTD subgroups have been reported. Involvement of the amygdala is not a commonly recognised feature of nfvPPA or lvPPA (Agosta et al., 2012) and has yet to be investigated in PPA-NOS. Based on these previous investigations, I hypothesise longitudinal rates of amygdalar atrophy will provide the strongest biomarkers of change in svPPA bilaterally, with low sample size

estimates of less than 100 using the direct BSI method. Bilateral measures in bvFTD will also provide promise, with low sample sizes of less than 200 given the greater heterogeneity of these patients. Given the strong signal provided by the left temporal lobe in PPA-NOS reported in the previous chapter, I further hypothesis that left amygdalar atrophy will demonstrate a robust signal of change and low resulting sample size estimates.

In the genetic subgroups, the marked medial temporal lobe atrophy characteristic of *MAPT* mutation carriers includes clear amygdala atrophy (Rohrer et al., 2010; Whitwell et al., 2012) and has been detected 10–15 years before symptom onset in a presymptomatic cohort (Rohrer et al., 2015). A recent investigation of the amygdala subnuclei also found that the *MAPT* group had the smallest subnuclei across the amygdala in both hemispheres (37–43% smaller than controls), compared with the *C9orf72* and *GRN* mutation patients who demonstrated similar volume differences to controls in the most affected hemisphere (14–29% smaller) (Bocchetta et al., 2019). Investigations in a pathology-confirmed cohort also reported increased amygdalar atrophy in *MAPT* individuals compared with Pick's Disease and FTD-ubiquitin-positive patients, although this was only shown for the right amygdala (Whitwell et al., 2005). Studies have been almost exclusively cross-sectional; therefore, comprehensive longitudinal investigations into the rate of amygdalar atrophy across the full FTD spectrum will be necessary to confirm and expand on these findings.

5.1.2 The hippocampus in FTD

Hippocampal atrophy has traditionally been associated with AD given its primary role in episodic memory but is increasingly recognised as a key feature in FTD. This is particularly true in bvFTD, with bilateral hippocampal atrophy reported as equivalent to that measured for AD patients (Hornberger et al., 2012). Moreover, early and striking asymmetric hippocampal atrophy is considered one of the hallmark features in svPPA patients (Landin-Romero et al., 2016). Longitudinal rates of hippocampal atrophy in bvFTD have been reported at 3.6% for a combined (left and right) measure (Landin-

Romero et al., 2017; supplementary data). To date, only one publication has provided rates of hippocampal atrophy (left at 7.2%; right at 7.5%) with accompanying sample size estimates in a svPPA population (Rohrer et al., 2008). Measures of the left hippocampal BSI produced an estimate of 104 participants required per treatment arm, whilst the right hippocampal BSI measure resulted in 81 individuals required to detect a 30% treatment effect (90% power, 5% significance), adjusting for control rates of change. Several other cross-sectional publications have also reported significant percentage differences from controls in hippocampal volume for bvFTD and svPPA in the range of 13–37% (Barnes et al., 2006; Chan et al., 2001; Lehmann et al., 2010; Schroeter et al., 2007) informing the hypothesis that hippocampal atrophy will provide a good longitudinal marker in these two FTD subgroups and result in low sample sizes bilaterally. Investigations with nfvPPA are limited but have generally failed to demonstrate any significant volume reduction in comparison to controls (van de Pol et al., 2006), and so may prove less useful in this subgroup. Inconsistent results have been reported for lvPPA. Several studies have failed to report hippocampal involvement (Agosta et al., 2012; Teichmann et al., 2013). In contrast, due to the most likely underlying pathology being AD (Gorno-Tempini et al., 2004), reduced bilateral hippocampal volumes have also been reported (Beber et al., 2014) and will need further investigation in the current longitudinal cohort. As with the amygdala, there are currently no investigations into hippocampal involvement in PPA-NOS; however, the high rates of temporal lobe atrophy reported in the previous chapter and visual inspection of these patient images suggest that left-sided hippocampal measures will prove valuable longitudinal markers of change resulting in low sample size estimates.

Within the genetic FTD subgroups, results also vary. Hippocampal volumes are most reduced in *MAPT* compared with controls, followed by *GRN* and *C9orf72* carriers. Bocchetta, Iglesias and colleagues (2018) also demonstrated different patterns of subfield involvement across these genetic mutations. However, other studies failed to find differences in annual rates of hippocampal atrophy between *MAPT* and *GRN* (Whitwell et al.,

2011) or reported reduced hippocampal volume in *MAPT* compared with *C9orf72* (Mahoney et al., 2012). In a presymptomatic genetic cohort, reduced cross-sectional hippocampal volumes were observed bilaterally, 15 years before expected symptom onset (Rohrer et al., 2015). This was not detected for the presymptomatic *GRN* or *C9orf72* carriers. Further longitudinal investigation is needed to gain a clearer picture of hippocampal involvement across the sporadic and genetic FTD subgroups.

5.1.3 The caudate in FTD

The caudate is a subcortical nucleus that forms part of the dorsal striatum. This region is a key component involved in motor functioning, particularly sensorimotor processes. Although historically associated with motor networks and thus Parkinson's disease and HD, it has widely distributed connections to cortical areas involved in motivation, cognition, emotional processing, inhibitory control and goal-directed behaviour (Graff-Radford et al., 2017; Haber, 2016), all of which are affected in FTD. Several cross-sectional investigations have reported significantly smaller caudate nuclei when comparing FTD patients and controls. In bvFTD patients, results generally show a bilateral pattern of volume reduction ranging from 16–21% on the left to 11–27% on the right compared with controls (Garibotto et al., 2011; Halabi et al., 2013; Macfarlane et al., 2015). Longitudinally, a high rate of caudate atrophy has been reported, with an annual loss of 10% in bvFTD (Landin-Romero et al., 2017; calculated from supplementary data). Reported caudate volumes for svPPA demonstrated an asymmetric left-sided reduction with 19% on the left versus 14% on the right (Garibotto et al., 2011). Results for nfvPPA patients have been more inconsistent in terms of laterality. Caudate atrophy was originally reported as bilateral (Gorno-Tempini et al., 2004) but more recently reported with reductions of 42% on the left and 32% on the right (Garibotto et al., 2011). Furthermore, Looi and colleagues (2008) specifically tested for asymmetry and found a significantly more left-sided reduction ($p = 0.003$) with 27% on the left and 15% on the right. Caudate atrophy has not been reported as a feature of lvPPA and has not been previously investigated in PPA-NOS.

In the genetic subgroups, caudate atrophy has been most associated with the *GRN* mutation. Significantly reduced caudate volume was reported in *GRN* but not in *C9orf72* or *MAPT* patients when compared with control volumes in a cross-sectional analysis of a large genetic cohort (Cash et al., 2018). There is also evidence of volume reduction in the striatum using a combined caudate and putamen measure five years before symptom onset (Rohrer et al., 2015). An additional cross-sectional study demonstrated a bilateral reduction of 55% in symptomatic *GRN* patients compared with healthy controls versus a lower 26% reduction in the sporadic FTD group (Premi et al., 2014). This is in line with previous neuropathological investigations where 67% of *GRN* cases had marked caudate atrophy at post-mortem, with degenerative changes of neuronal loss and gliosis in 100% of cases evident during the microscopic examination (Mackenzie et al., 2006). As there is currently only one longitudinal investigation of caudate atrophy with the FTD spectrum (including only bvFTD), it will be important to ascertain the potential utility of this measure as a future biomarker within the clinical, genetic and pathology subgroups. Based on the available cross-sectional investigations, I hypothesis that longitudinal caudate atrophy measures will provide valuable biomarkers on the left in svPPA and PPA-NOS and bilaterally in bvFTD, nfvPPA and *GRN* patients.

5.1.4 The thalamus in FTD

The thalamus consists of a cluster of nuclei, often characterised as the hub or relay station of the brain due to its extensive connectivity to all lobes of the cerebral cortex as well as midbrain and other subcortical structures. As such, it is involved in the communication or regulation of a wide range of neurological functions (Herrero, Barcia, & Navarro, 2002). Reduced thalamic volumes have been widely reported across the spectrum of clinical, genetic and pathology-confirmed FTD patients (Bocchetta et al., 2018, 2019, 2020; Cardenas et al., 2007; Cash et al., 2018; Garibotto et al., 2011; Landin-Romero et al., 2017; Premi et al., 2014; Rohrer et al., 2010, 2015). The bvFTD subgroup has a bilateral presentation of atrophy with previously

reported reductions of 10–17% for the left thalamus and 8–19% for the right compared with control volumes (Bocchetta et al., 2018; Borroni et al., 2007; Garibotto et al., 2011). An additional cross-sectional study using a direct deformation-based morphometry method in SPM found an average of 26% volume reduction in relation to controls (Cardenas et al., 2007). Rates of change have been reported as significantly higher in bvFTD than controls in a longitudinal study, which reported a combined thalamic annual atrophy rate at 3.7% of baseline volume (Landin-Romero et al., 2017; calculated from supplementary data). In line with the characteristically asymmetric presentation in svPPA, a recent study showed a 9% reduction on the left and 0.1% reduction on the right compared with controls (Bocchetta et al., 2018). However, a previous study has shown a more bilateral reduction, with 11% on the left and 10% on the right compared with healthy controls (Garibotto et al., 2011). It is important to note that the disease duration in the former was 4.6 (2.3) and the latter was 2.2 (2.3) years, so the inconsistency may be due to capturing different stages of the disease and the potential of inclusion of a right temporal variant as asymmetry was not investigated at the individual level (Chan et al., 2009). NfvPPA patients show a high and asymmetric presentation, with 10–21% reduction in the left thalamus and 6–16% reduction on the right compared with controls (Bocchetta et al., 2018; Garibotto et al., 2011). Thalamic atrophy is not a recognised feature of lvPPA. Thalamic atrophy has been investigated in a cohort of PPA-NOS patients and shown to be significantly higher than controls, with an asymmetric volume reduction of 12% and 2% (left and right). They also show bilaterally smaller thalami compared with svPPA (Bocchetta et al., 2018).

In the genetic groups, thalamic atrophy has been most associated with the *C9orf72* expansion. Cross-sectional analyses in *C9orf72* symptomatic patients demonstrate significantly reduced thalamic volumes bilaterally compared with controls (Bocchetta et al., 2018; Cash et al., 2018; Lee et al., 2014), which has also been observed in presymptomatic carriers (Cash et al., 2018; Lee et al., 2017; Rohrer et al., 2015). Longitudinal analysis of *GRN*

patients has demonstrated asymmetric thalamic atrophy but only once images were flipped so that the most affected hemispheres were on the same side (Rohrer et al., 2010). Cross-sectional analyses have failed to find any asymmetry, which may be due to averaging across a cohort with both left- and right-dominant patterns of atrophy. Premi and colleagues (2014) found a large reduction bilaterally in *GRN* patients, with 28% on the left and 24% on the right compared with healthy controls. This was significantly higher than the sporadic FTD cohort which presented with a 10% volume reduction in both hemispheres (Premi et al., 2014). *MAPT* patients also show significantly reduced volumes compared with controls (Bocchetta et al., 2018). Interestingly, recent work has demonstrated differential involvement of the thalamic nuclei across the FTD spectrum. The mediodorsal nucleus was universally affected in all clinical, genetic and pathology FTD subgroups, with 16–33% reduction compared with controls. The laterodorsal nucleus was particularly affected in the three genetic groups (28–38% smaller than controls) and the pulvinar only affected in the *C9orf72* group (16% smaller) (Bocchetta et al., 2020). Together, the literature suggests thalamic atrophy will be a useful marker across the FTD spectrum for future interventions and thus hypothesise that increased annual rates will be detected bilaterally for bvFTD, *MAPT* and *C9orf72* patients and on the left for the PPA and *GRN* subgroups, producing low feasible sample sizes less than 200 using the direct BSI measure.

5.2 Methods

5.2.1 Participants

Participants consisted of the full retrospective neuroimaging cohort (n = 262) described in detail in Section 2.2 and Table 2-1.

5.2.2 Image analysis

All images were aligned to MNI-125 template space to ensure neuroanatomical regions were in similar stereotactic space irrespective of positional differences during initial scan acquisition. This transformation reorients and resamples the image to produce isotropic voxels of 1 mm $\times$ 1 mm $\times$ 1 mm irrespective of the variety of acquisition protocols included in the cohort (see Table 2-3). Because subcortical structures are relatively small in relation to the whole-brain, correct alignment is important to ensure correspondence of structures and for optimal matching of the subcortical template libraries to the target structure. Three template libraries were used for the Similarity and Truth Estimation for Propagated Segmentations (STEPS) analysis: the hippocampus, the caudate and the thalamus. The amygdala regions were derived using the GIF parcellation method. As described in Section 2.3.3.8, the STEPS algorithm utilises these large template libraries of manually segmented regions and uses those that are best matched to the unsegmented target image to propagate a new region. This is initially done using a global similarity metric to narrow the template choices to the most optimal at the image level, and then employs a local similarity metric based on the fast LNCC method at the voxel level for increased precision (Cardoso et al., 2013). STEPS was performed for each baseline and repeat image on the left and right hemispheres separately. This was repeated for each of the hippocampus, caudate and thalamus template libraries independently.

5.2.2.1 Template libraries

The hippocampal template library consists of 128 T1-weighted MR images from individuals with several diagnoses (including AD, MCI, bvFTD and

PPA) and healthy controls. All scans were in MNI template space and had both hippocampi manually segmented by expert raters. These were then flipped in the coronal plane such that the right side appeared on the left. Including these mirror images resulted in a template library of 512 hippocampal segmentations (256 left and 256 right).

The caudate library consists of 101 MNI images from 40 healthy controls and 61 early HD patients. As with the hippocampal library, all images and manual segmentations were flipped along the coronal plane to create mirror-reverse images, doubling the template library size to 404 caudates (202 per hemisphere) as potential best matches to the target image.

The thalamus library consists of 325 MNI images drawn from the DIAN (Dominantly Inherited Alzheimer Network) study, including controls and genetic AD participants at various symptomatic and presymptomatic stages. All images were manually segmented, and a final combined flipped and non-flipped set of potential matches consisted of 650 candidate thalamic labels for each hemisphere.

5.2.2.2 Subcortical segmentation and registration

I did not have access to a template library of manually segmented amygdalar regions for this project. Instead, I utilised the left and right amygdalar labels from the GIF segmentations used for Chapters 3 and 4. For consistency, I used the transformation parameters produced as part of the MNI-125 image realignment and applied them to the full GIF segmentation mask before extracting the 1048 realigned MNI amygdalar masks. Figure 5-1 provides a representation of each automated segmentation with both left and right regions overlaid on a subject's MNI image.

STEPS was run on all 262 subjects (524 images) from the FTD cohort, using each of the hippocampus, caudate and thalamus template libraries separately. This produced 1048 binary segmentations (a left and right

region) for each of the three structures. After some initial scripting adjustments due to space requirements and correction of a large number of output files, 3134 of the possible 3144 STEPS segmentations completed the full pipeline to produce a final left and right propagated region for the hippocampus, the caudate and the thalamus. Ten of the thalamus STEPS segmentations failed to complete during this project. Volumes from the 4182 final subcortical segmentations were then outputted and used to calculate the indirect annual volume-difference measure as described in Section 2.3.4

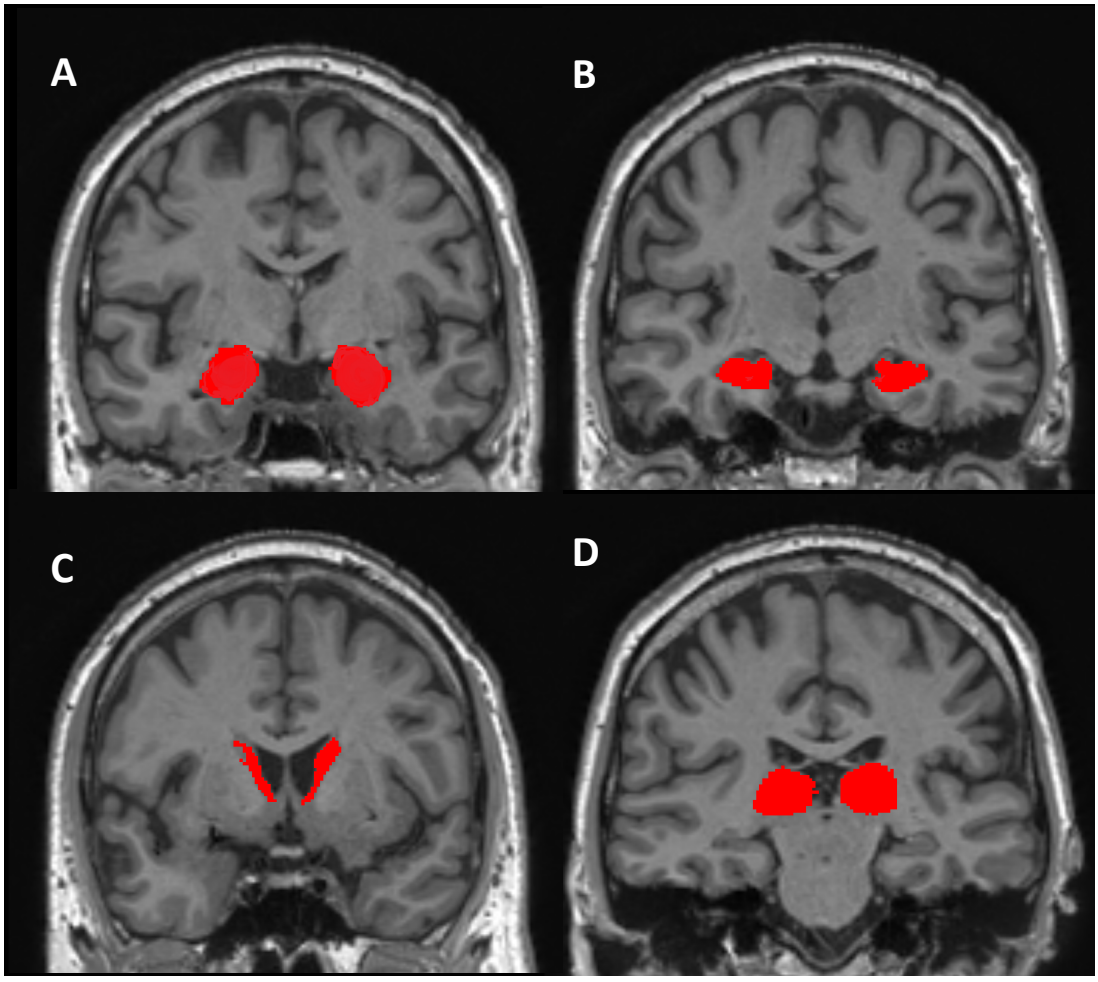

Figure 5-1 Representation of the automated region segmentations for the left and right A) amygdala GIF, B) hippocampal STEPS, C) caudate STEPS and D) thalamus STEPS overlaid on the MNI image.

The direct longitudinal subcortical BSI was calculated using a double intensity window approach (Hobbs et al., 2009) to capture the boundary shift across both the CSF–subcortical GM border and the subcortical GM–WM border, see Figure 5-2 below. Because the MR images in the cohort were

acquired using multiple scanners and protocols, the intensity windows for the CSF–GM and GM–WM boundaries were automatically estimated separately for each image pair. The global KBSI method was applied as described in Section 2.3.4.1, with an additional local 6 dof (three translations, three rotations) subcortical–subcortical registration applied separately for the left and right regions. The KBSI was calculated using the XOR derived from the subcortical regions dilated by two voxels to account for potential segmentation error. The final output from this included: 509 amygdalar BSI ('ABSI') – 15 failed to complete the full registration pipeline; 511 hippocampal BSI ('HBSI') – 13 failed to complete; 515 caudate BSI ('CBSI') – nine failed to complete, and 508 thalamus BSI ('TBSI') – six failed to complete and 10 had a STEPS region unavailable.

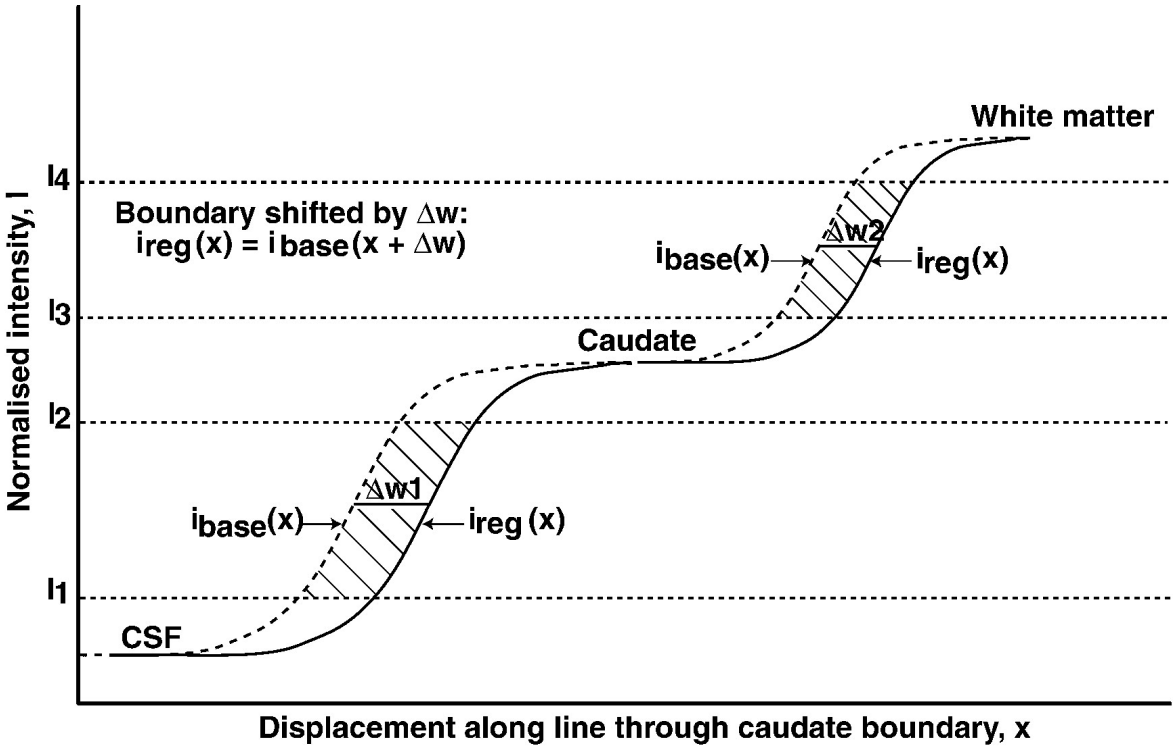

Figure 5-2 One-dimensional representation of the intensity change associated with a boundary shift between a baseline scan and a locally registered repeat scan for the caudate as an example. The dotted line represents the intensity profile (boundary or edge) for the baseline scan; the solid line represents the repeat scan; the shift in edge, Δw represents atrophy. I_1 to I_2 is the intensity window for estimating change at the CSF–GM boundary and I_3 to I_4 is the intensity window for estimating change at the GM–WM boundary. Reproduced from Hobbs et al. (2009).

5.2.2.3 Quality control

To ensure the integrity of the data, several visual checks were performed. Before the STEPS segmentations were run, all 1048 MNI-transformed amygdalar masks were loaded with accompanying MNI images to check that both the MR image and amygdalar mask realignment had been successful. All 1048 regions and all 524 images were correctly realigned into MNI space ensuring all STEPS segmentations were propagated onto correctly reoriented MNI images. Following completion of the 3134 subcortical STEPS segmentations and subsequent running of the 2091 subcortical BSI paired local registrations, the following range checking was performed.

All segmentations and registrations with an annual negative BSI or GIF/STEPS volume change (i.e. growth) of 10% or above were reviewed. Upon visual review, all of these cases (n = 30 out of 4134) showed a clear failure of part of the analysis pipeline (i.e. a segmentation or registration error) or non-biological artefacts causing longitudinal inconsistencies (i.e. motion or flow differences) that were biologically implausible and thus excluded. The remaining datasets for each of the eight BSI and eight GIF/STEPS volume-difference measures were then modelled separately and all values that fell outside a 1.5 interquartile range (IQR) from the upper and lower quartile were flagged for visual assessment (red data points in Figure 5-3).

This was also done for image pairs where the absolute difference between the GIF/STEPS volume measure and the equivalent BSI measure for that individual fell outside the allowed IQR values, indicating the indirect and direct measures from the same image pair were providing substantially different results (Figure 5-4). A random selection of 25 image pairs (~10%) for each of the 16 longitudinal measures was also visually reviewed, selected using an online random number generator (www.random.org/). Based on these criteria, 681 image pairs were flagged for review. This list of image pairs was used to program an automated loading script that

allowed review and QC response to be entered blinded to the scan pair and diagnosis. The local registered image pairs with region overlays were used for the review, with additional checks of the unregistered regions if I flagged the segmentation quality was questionable. Final numbers included in the analysis following QC can be seen in Table 5-1.

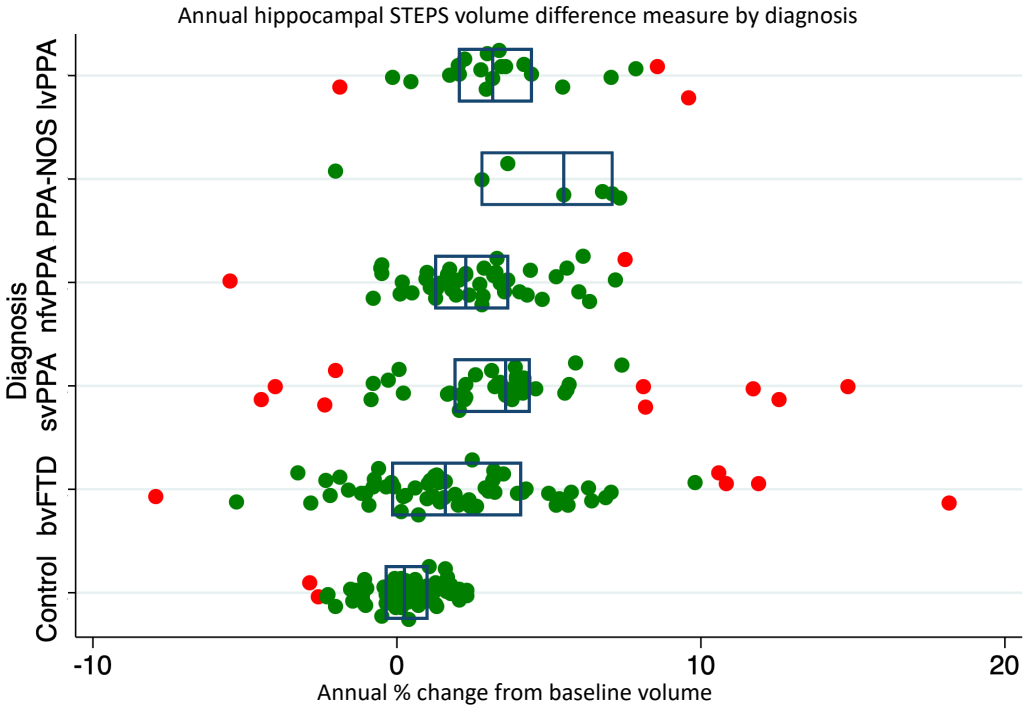

Figure 5-3 Graphical representation of the outliers flagged for review using the annual hippocampal STEPS indirect volume-loss measure for the left hemisphere (Hippocampal_STEPS_Left) as an example. All pairs of images in red fell outside the range and were visually assessed.

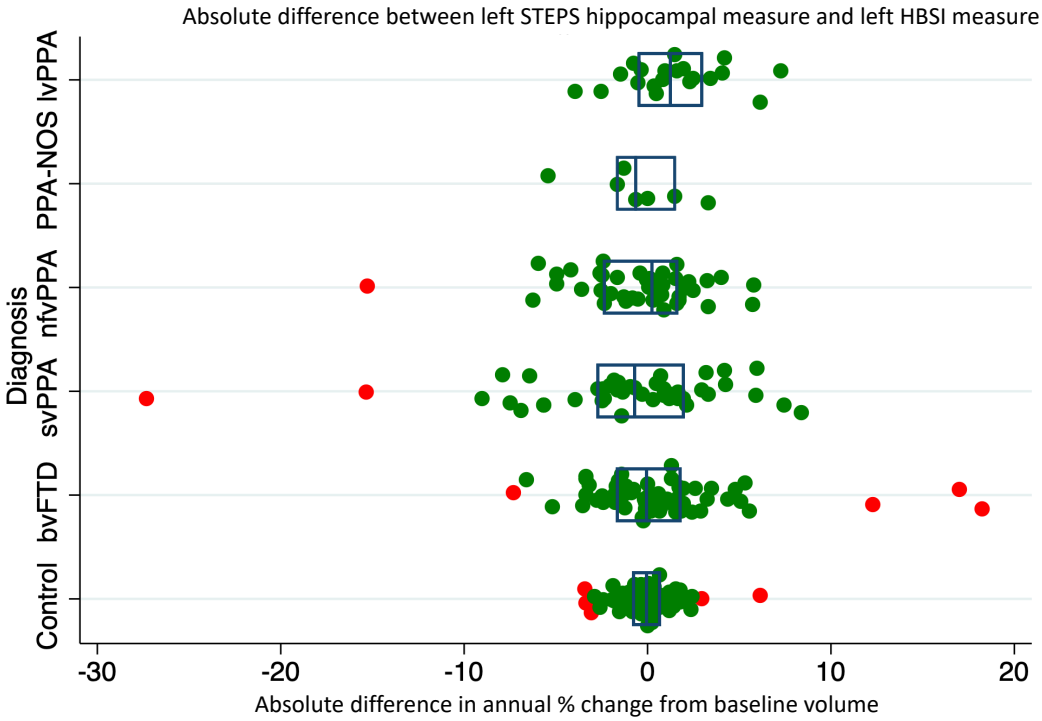

Figure 5-4 Graphical representation for outliers flagged for review where the indirect and direct measures of change for a scan pair were substantially different.

5.3 Results

5.3.1 Rates of change and regression analysis

Table 5-1 presents the mean and sd for the subcortical annual rates of atrophy, expressed as a percentage change from baseline volume. The number of image pairs that passed QC is listed in the observations (Obs) column for each subcortical measure and each subgroup. The original total number of all individuals in each subgroup is listed in the header. Figure 5-5 shows these mean subcortical atrophy rates with 95% confidence intervals for the 16 longitudinal measures presenting the controls (in blue) compared with each FTD subgroup (in red) separately. Note the extended x-axis required for PPA-NOS, *C9orf72* and *GRN* graphs due to substantially wider 95% CI for the left caudate measures in these subgroups. Table 5-2 presents the results of the linear regression comparing each FTD subgroup with controls across all subcortical longitudinal measures. The adjusted mean difference (regression coefficient (Coef.)) demonstrating the additional atrophy in the FTD subgroup compared to controls when correcting for age, gender and scanner is listed with 95% CIs and associated *p*-values.

Control rates were consistently low across both direct and indirect longitudinal measures for the amygdala, hippocampus, caudate and thalamus in both left and right hemispheres. Remaining close to a 0% volume change for all subcortical regions, the measures ranged from a slight increase of –0.1 (1.4)% volume loss for the right hippocampal STEPS measure to 0.6 (1.4)% volume loss for the right CBSI.

Table 5-1 Mean (sd) of annual rates of subcortical change expressed as the annual percentage change from baseline volume for each FTD subgroup

Measure	Controls (n = 78)		Clinical subgroups									
			bvFTD (n = 66)		svPPA (n = 45)		nfvPPA (n = 45)		PPA-NOS (n = 7)		lvPPA (n = 21)	
Subcortical BSI	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**
ABSI_Left	76	0.0 (2.0)	62	3.2 (4.1)	41	5.8 (4.6)	44	5.1 (5.1)	7	4.8 (1.9)	21	3.7 (4.8)
ABSI_Right	76	0.0 (1.5)	60	2.9 (4.7)	41	6.8 (4.6)	44	3.4 (4.3)	7	1.6 (2.5)	20	2.2 (3.8)
HBSI_Left	77	0.3 (1.3)	59	2.4 (3.6)	43	4.1 (3.0)	45	2.9 (3.0)	7	5.0 (1.3)	20	2.4 (2.3)
HBSI_Right	77	0.1 (1.2)	62	1.7 (3.2)	43	2.8 (3.4)	45	2.0 (2.2)	7	1.5 (1.5)	20	0.8 (2.7)
CBSI_Left	77	0.4 (1.6)	65	6.5 (8.5)	44	8.0 (4.0)	44	9.0 (8.4)	7	11.8 (5.1)	20	7.0 (3.6)
CBSI_Right	77	0.6 (1.4)	65	5.9 (7.1)	45	5.3 (4.1)	44	7.5 (6.2)	7	4.3 (2.4)	20	4.1 (2.2)
TBSI_Left	77	0.2 (0.7)	64	1.3 (1.5)	45	1.7 (0.9)	45	2.0 (1.8)	7	3.3 (1.8)	19	1.5 (0.8)
TBSI_Right	77	0.1 (0.6)	64	0.9 (1.0)	43	1.2 (0.7)	38	1.0 (0.8)	6	0.8 (1.1)	19	0.9 (0.9)
GIF and STEPS volume difference												
Amygdala_GIF_Left	76	0.2 (1.9)	62	1.5 (3.1)	41	1.4 (3.7)	44	3.6 (3.6)	7	3.6 (2.3)	20	4.1 (4.0)
Amygdala_GIF_Right	76	0.2 (1.9)	60	2.6 (4.4)	45	5.0 (4.1)	43	2.3 (3.7)	7	1.2 (2.2)	21	2.1 (3.9)
Hippocampal_STEPS_Left	78	0.2 (1.2)	63	2.5 (3.9)	39	3.5 (2.8)	44	2.7 (2.1)	7	4.5 (3.3)	21	3.6 (2.8)
Hippocampal_STEPS_Right	78	-0.1 (1.4)	59	2.1 (3.3)	44	3.7 (2.4)	45	2.4 (3.3)	7	3.4 (2.9)	21	2.9 (1.8)
Caudate_STEPS_Left	77	0.0 (2.4)	60	5.3 (7.9)	45	5.9 (6.9)	42	5.8 (6.2)	7	5.2 (2.4)	19	3.2 (3.4)
Caudate_STEPS_Right	75	0.3 (2.7)	62	2.7 (4.8)	44	3.0 (3.9)	42	2.9 (4.5)	7	1.2 (2.4)	19	1.5 (3.2)
Thalamus_STEPS_Left	78	0.4 (1.5)	65	2.0 (2.8)	44	2.2 (1.6)	44	3.2 (2.7)	7	3.2 (2.5)	21	2.5 (2.8)
Thalamus_STEPS_Right	78	0.1 (2.2)	64	1.3 (2.7)	42	0.1 (2.1)	38	1.2 (2.5)	6	-0.1 (0.9)	20	0.6 (2.6)

Table 5-1 continued: Mean (sd) of annual rates of subcortical change expressed as the annual percentage change from baseline volume for each FTD subgroup

Measure	Genetic subgroups						Pathology subgroups			
	MAPT (n = 16)		*C9orf72* (n = 10)		*GRN* (n = 8)		Tau (n = 19)		TDP-43 (n = 34)	
Subcortical BSI	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**
ABSI_Left	14	3.8 (3.8)	9	2.1 (3.8)	7	3.0 (1.8)	17	3.4 (2.8)	30	3.8 (4.8)
ABSI_Right	15	3.8 (3.3)	9	1.5 (2.8)	8	-0.1 (4.0)	18	3.9 (3.3)	31	4.5 (6.0)
HBSI_Left	15	2.4 (3.5)	10	0.7 (2.8)	6	3.9 (2.9)	18	2.7 (3.4)	31	2.7 (2.7)
HBSI_Right	15	2.4 (2.6)	10	0.8 (1.7)	8	-0.1 (2.6)	18	2.3 (2.4)	33	2.1 (2.9)
CBSI_Left	16	6.7 (4.3)	10	8.1 (11.2)	8	12.3 (9.8)	19	7.5 (5.5)	33	8.6 (8.7)
CBSI_Right	16	5.6 (5.4)	10	6.5 (5.2)	8	5.4 (6.7)	19	5.9 (5.1)	33	5.7 (4.9)
TBSI_Left	16	1.5 (0.9)	9	1.1 (1.7)	7	3.1 (3.2)	19	1.6 (0.9)	31	1.5 (2.0)
TBSI_Right	15	1.0 (0.9)	9	0.3 (0.5)	6	0.8 (1.7)	19	1.2 (0.9)	29	0.8 (1.0)
GIF and STEPS volume difference										
Amygdala_GIF_Left	16	1.2 (2.5)	9	0.2 (2.1)	8	1.4 (3.9)	19	1.4 (2.3)	31	0.8 (2.6)
Amygdala_GIF_Right	15	1.6 (3.5)	10	2.9 (5.6)	7	-0.7 (3.1)	18	2.7 (4.1)	32	3.0 (4.3)
Hippocampal_STEPS_Left	16	0.7 (2.6)	10	2.1 (2.8)	8	4.3 (4.4)	19	1.4 (3.0)	33	2.6 (3.2)
Hippocampal_STEPS_Right	16	0.3 (2.1)	10	1.6 (1.7)	7	0.8 (2.9)	19	0.4 (2.0)	32	3.0 (3.1)
Caudate_STEPS_Left	15	4.8 (5.2)	10	7.0 (14.2)	7	7.2 (7.4)	18	4.5 (4.9)	31	6.1 (9.7)
Caudate_STEPS_Right	16	3.8 (5.8)	10	2.2 (4.9)	7	4.3 (4.9)	19	3.8 (5.5)	32	2.8 (4.0)
Thalamus_STEPS_Left	16	1.6 (1.7)	10	2.3 (2.6)	7	3.4 (3.4)	19	2.0 (1.7)	32	2.4 (2.3)
Thalamus_STEPS_Right	15	0.8 (1.5)	10	1.7 (2.9)	6	-0.1 (3.0)	19	0.7 (1.5)	30	0.8 (2.5)

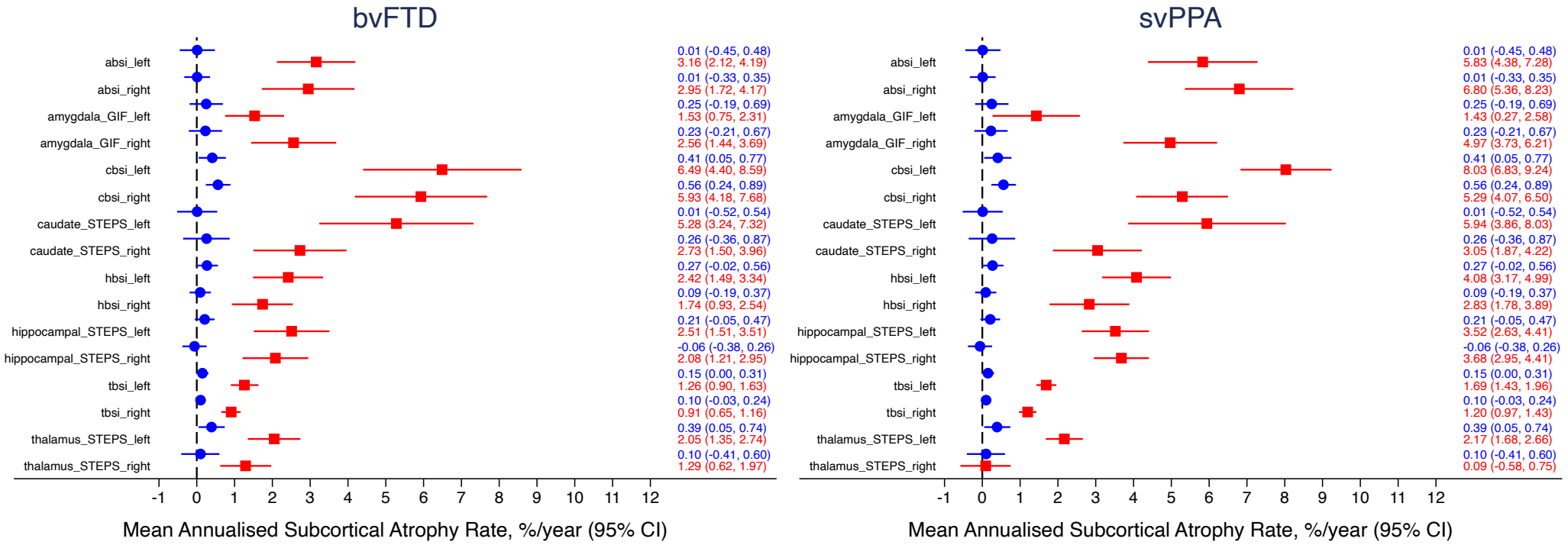

Figure 5-5 Mean and 95% confidence interval for annual subcortical atrophy rates for direct BSI and indirect volume change using the GIF and STEPS segmentation volumes for controls (blue) and each FTD subgroup (red).

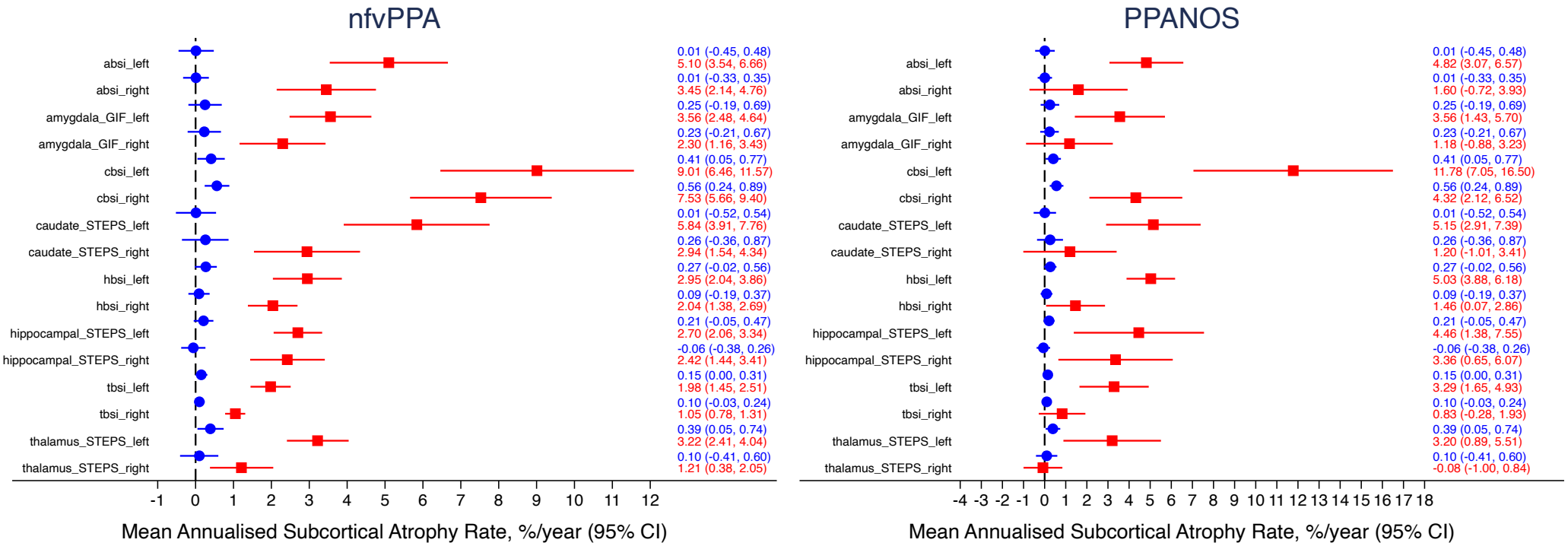

Figure 5-5 continued: Mean and 95% confidence interval for annual subcortical atrophy rates for direct BSI and indirect volume change using the GIF and STEPS segmentation volumes for controls (blue) and each FTD subgroup (red).

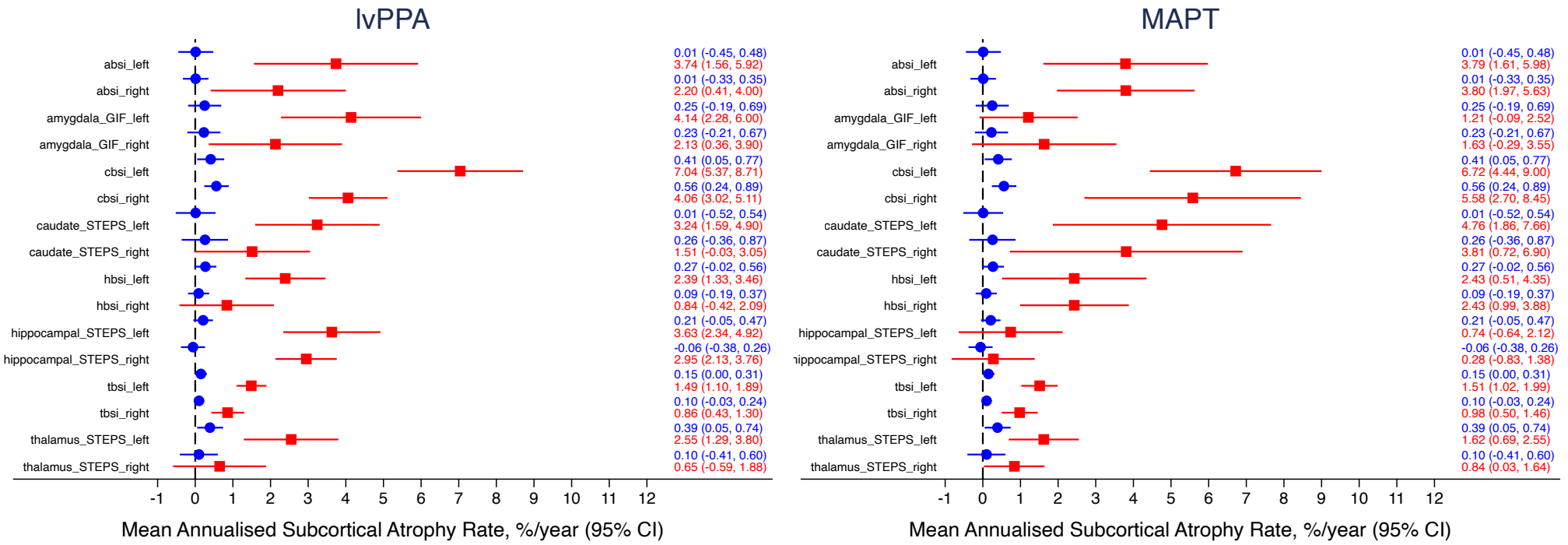

Figure 5-5 continued: Mean and 95% confidence interval for annual subcortical atrophy rates for direct BSI and indirect volume change using the GIF and STEPS segmentation volumes for controls (blue) and each FTD subgroup (red).

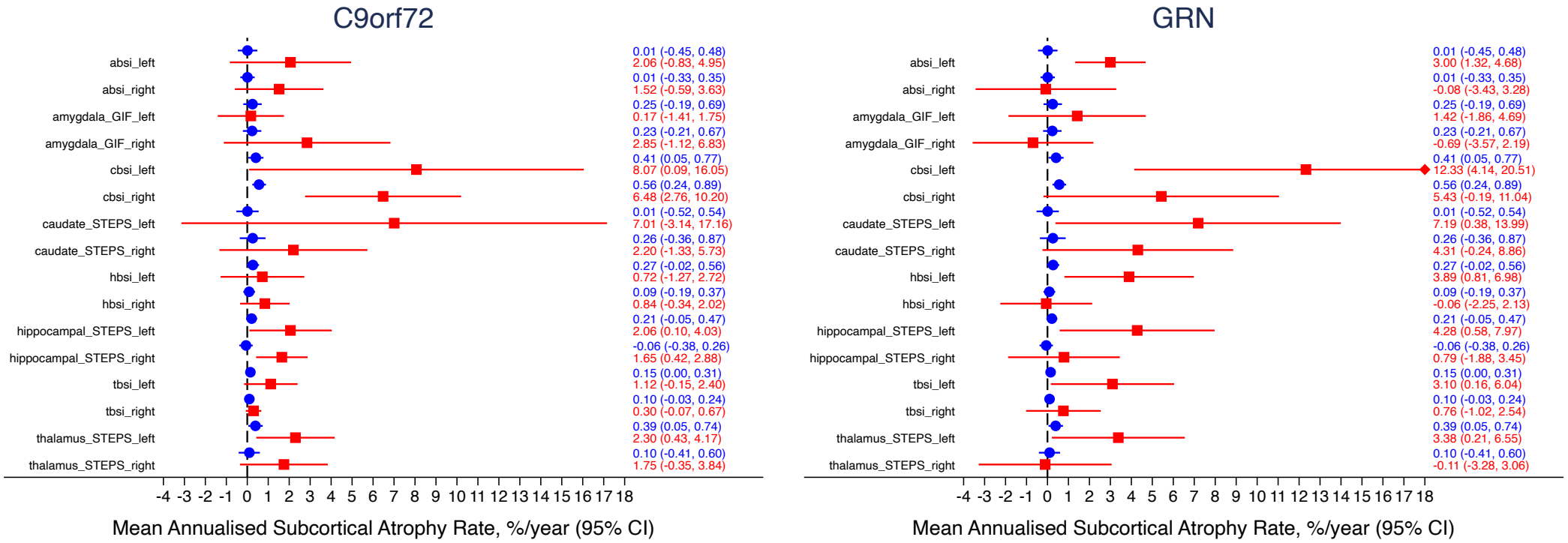

Figure 5-5 continued: Mean and 95% confidence interval for annual subcortical atrophy rates for direct BSI and indirect volume change using the GIF and STEPS segmentation volumes for controls (blue) and each FTD subgroup (red).

♦ represents a capped value where the 95% limit lies outside that chosen for the x-axis range. The actual value is listed on the right of the graph.

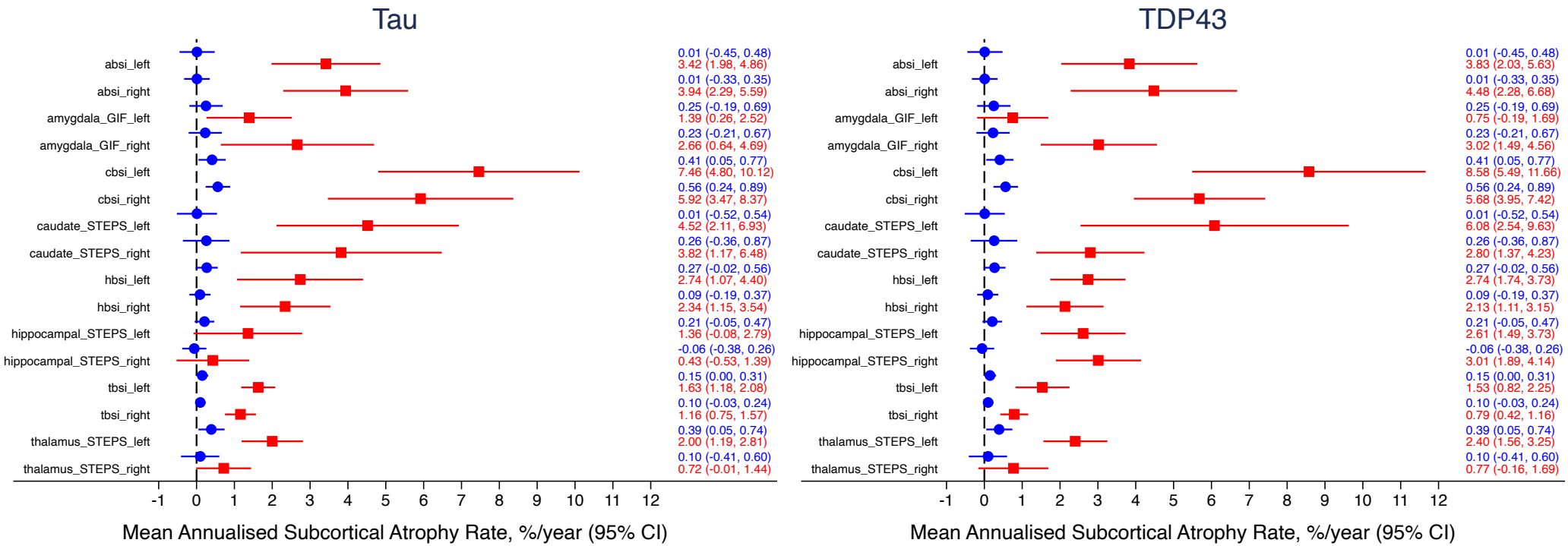

Figure 5-5 continued: Mean and 95% confidence interval for annual subcortical atrophy rates for direct BSI and indirect volume change using the GIF and STEPS segmentation volumes for controls (blue) and each FTD subgroup (red).

5.3.1.1 Clinical subgroups

Results for the bvFTD subgroup demonstrated significantly higher rates of change than controls for all subcortical measures. Results from the regression analysis comparing FTD subcortical atrophy rates with controls are shown in Table 5-2. The 95% CI and adjusted mean difference (Coef.) of the annual patient rates of change after correcting for age, gender and scanner type are also included. The highest raw and adjusted mean difference in rates of change for bvFTD were produced by the left CBSI at 6.5%, closely followed by the right CBSI at 5.9%. The smallest annual change resulted from the right TBSI at 0.9% per year, but importantly it was also the measure that has the highest mean-to-sd ratio from the raw atrophy rates. Given its documented heterogeneity, it is unsurprising that all the subcortical measures for bvFTD had a mean-to-sd ratio of <1.0. A surprising finding when comparing the direct measures of caudate atrophy with the indirect volume-difference measure results is the asymmetry produced by the latter compared to the relatively symmetric result using the CBSI noted above. This can be more clearly seen in Figure 5-5. Looking in detail at the individual data points that contributed to these distributions, the source of this apparent asymmetry becomes clearer. Six of the left indirect STEPS volume-difference measures failed QC compared with only one of the corresponding left CBSI measures. The majority of the left indirect QC fails had large negative annual volume loss due to the repeat segmentation over-estimating the caudate boundary. The removal of these datapoints substantially increased the mean for this side whereby artificially overemphasising the relative left-right difference. The application of the BSI on these segmentation pairs provided a more robust measure that was not significantly impacted by the segmentation errors that had resulted in the QC failures. This also highlights the importance of using identical datasets when comparing across measures to ensure the same individuals are contributing data to produce a more representative conclusion on relative performance.

SvPPA patients also demonstrated the highest rate of change in the caudate as measured by the left CBSI. Both left and right amygdala showed substantial atrophy as measured by the ABSI and hippocampal rates of change also demonstrated good group separation from controls on the left with the HBSI and bilaterally using the hippocampal STEPS volume difference. Again, the thalamus measures produced the lowest annual rates of change, but importantly also had the lowest standard deviation with the TBSI reporting 1.7 (0.9)% and 1.2 (0.7)% volume loss for the left and right hemispheres, respectively. The linear regression analysis confirmed that all eight subcortical BSI rates of change in the svPPA were significantly higher than controls ($p < 0.001$). The indirect GIF/STEPS longitudinal measures were also significantly higher than controls at $p < 0.001$, except for the left GIF amygdala ($p = 0.048$) and right STEPS thalamus measure ($p = 0.978$). Reassuringly, application of the BSI to these segmentations substantially improved the mean-to-sd ratio for all regions, except for the right hippocampal measure. It also substantially increased the mean adjusted difference in atrophy rate as derived from the linear regression model (Coef.) indicating detection of larger increases in atrophy rate compared to controls. Reassuringly, in general, these adjusted mean difference results from the regressions analysis closely reflected the raw mean rates of change measured in the FTD subgroups (Table 5-1) confirming the minimal atrophy in the reference control group and relatively small impact of any gender, age or scanner type differences found between subgroups (Table 2-1).

NfvPPA patients also demonstrated significantly higher rates of subcortical atrophy than controls across all 16 measures. Linear regression analysis results all reached a significant difference in rates at $p < 0.001$, except for the right amygdala GIF measure ($p = 0.002$) and right thalamus STEPS measure ($p = 0.020$). The highest rates of change in the nfvPPA patients were observed in the left caudate using both the BSI and STEPS volume-difference measures. The right CBSI as well as the left amygdala, left hippocampus and left thalamus measures also demonstrated high annual rates of volume loss and good group separation from controls.

The PPA-NOS subgroup again presented with clear left-sided asymmetry in their pattern of atrophy across all subcortical regions, with the left caudate measures producing the highest atrophy rates. The adjusted mean difference derived from the regression analysis demonstrated that PPA-NOS patients had an additional 11.6% caudate atrophy compared with controls using the left CBSI and 7.0% for STEPS volume measure after correcting for age, gender and scanner (both $p < 0.001$). Regression analyses also revealed that BSI measures for the left amygdala, left hippocampus and left thalamus demonstrated significantly higher rates than controls ($p < 0.001$). These three structures also produced significantly higher annual rates of volume loss than controls at $p < 0.05$ for the indirect GIF/STEPS volume-difference measures. These results suggest that measures of volume loss in several subcortical regions in the left hemisphere show promise for tracking disease progression and treatment effects within the PPA-NOS clinical population.

As with the other PPA subgroups, lvPPA patients presented with the highest rates of atrophy in the caudate, seen bilaterally when measured by the CBSI. Significantly higher rates of atrophy were also seen bilaterally in the hippocampus using the STEPS volume-difference measure (both $p < 0.001$) and on the left for the HBSI ($p = 0.002$). Both indirect and direct measures for the left amygdala and left thalamus provided atrophy rates significantly higher than controls ($p < 0.001$) and also for the right thalamus when using the TBSI ($p < 0.001$), suggesting that a direct measure of left subcortical loss would prove valuable in this population.

5.3.1.2 Genetic subgroups

In the *MAPT* subgroups, the eight subcortical BSI measures demonstrated significantly higher rates of atrophy than controls (all $p < 0.001$). The caudate produced the highest rate of atrophy for both the BSI and the STEPS volume-difference measure, ranging from 3.8 to 6.7% annual volume loss. The slightly lower mean-to-sd ratio in the indirect STEPS measures

produced a regression result of $p = 0.002$ and $p = 0.001$ for the left and right caudate, respectively. The thalamus measures had the lowest annual rate of change but again had the highest mean-to-sd ratio for the left thalamus, suggesting it may be a useful marker in this patient population. The results for the *C9orf72* subgroup were more variable with several measures overlapping with controls as can be seen in Figure 5-5. The higher standard deviations in this small cohort meant that only the CBSI (bilaterally) and the left caudate STEPS measure provided the best group separation from controls at the $p < 0.001$ level. The ABSI demonstrates higher bilateral amygdala atrophy rates (both $p < 0.05$) and greater left thalamic atrophy than controls as measured using the TBSI ($p = 0.015$). In addition, the hippocampal STEPS volume difference showed bilateral increased rates of atrophy ($p = 0.006$; left and $p = 0.004$; right); however, this was not demonstrated using the HBSI measures ($p = 0.495$; left and $p = 0.174$; right). Rates of change and regression results in the *GRN* subgroup demonstrated a generalised left-dominant pattern of atrophy across all four subcortical regions. The left caudate measures provided the highest rates of annual volume change, with 12.3 (9.8)% for the CBSI and 7.2 (7.4)% for the caudate STEPS volume-difference measures. In addition, the left amygdalar atrophy rate as measured using the ABSI had the highest mean-to-sd ratio, suggesting it may provide the most robust measure of change in this subgroup.

5.3.1.3 Pathology subgroups

Similar to the *MAPT* subgroup, all eight subcortical BSI measures were significantly higher in the Tau subgroup than controls ($p < 0.001$). The amygdalar GIF volume-difference measure also demonstrated higher rates of annual volume loss bilaterally, along with the left hippocampal STEPS and left thalamic STEPS measures (all $p < 0.05$). The caudate provided the highest rates of annual volume loss bilaterally using the CBSI measure. The left TBSI measure had the highest mean-to-sd ratio, followed by the ABSI bilaterally, suggesting that BSI measures of the caudate, left thalamus and bilateral amygdala may provide promising measures of change in the Tau

subgroup. Annual rates of change in the TDP-43 group demonstrated high rates across all subcortical regions (although with overall lower mean-to-sd ratios for many of the measures than the Tau group). The linear regression analyses revealed that these annual rates of subcortical volume loss were significantly higher than controls for all eight BSI measures at $p < 0.001$. In addition, rates of change in the right amygdala, bilateral hippocampus, bilateral caudate and left thalamus were all significantly higher than controls using the indirect GIF and STEPS volume-difference measures (all $p \leq 0.001$).

Table 5-2 Linear regression analysis results with 95% CI comparing annual rates of subcortical atrophy for each FTD subgroups with controls. *

Measure	Clinical subgroups									
	bvFTD		svPPA		nfvPPA		PPA-NOS		lvPPA	
Subcortical BSI	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**
ABSI_Left	3.1 [1.7–4.5]	**<0.001**	5.8 [4.3–7.4]	**<0.001**	5.1 [3.6–6.6]	**<0.001**	4.7 [1.7–7.8]	**<0.001**	3.6 [1.8–5.6]	**<0.001**
ABSI_Right	2.7 [1.3–4.0]	**<0.001**	6.7 [5.2–8.1]	**<0.001**	3.5 [2.1–4.9]	**<0.001**	1.7 [-1.3–4.6]	0.263	2.4 [0.4–4.3]	*0.017*
HBSI_Left	2.0 [1.0–2.9]	**<0.001**	3.7 [2.7–4.7]	**<0.001**	2.6 [1.7–3.6]	**<0.001**	4.9 [2.8–7.0]	**<0.001**	2.3 [0.9–3.6]	*0.001*
HBSI_Right	1.7 [0.8–2.5]	**<0.001**	2.7 [1.7–3.6]	**<0.001**	2.0 [1.1–3.0]	**<0.001**	1.6 [-0.4–3.5]	0.111	1.1 [-0.2–2.4]	0.087
CBSI_Left	6.5 [4.5–8.5]	**<0.001**	7.7 [5.5–9.9]	**<0.001**	9.0 [6.8–11.3]	**<0.001**	11.8 [7.2–16.4]	**<0.001**	7.7 [4.7–10.8]	**<0.001**
CBSI_Right	5.3 [3.7–6.9]	**<0.001**	4.6 [2.8–6.3]	**<0.001**	7.2 [5.4–9.0]	**<0.001**	4.1 [0.4–7.8]	*0.029*	4.3 [1.9–6.8]	**<0.001**
TBSI_Left	1.2 [0.8–1.6]	**<0.001**	1.5 [1.1–2.0]	**<0.001**	1.9 [1.4–2.3]	**<0.001**	3.2 [2.3–4.2]	**<0.001**	1.5 [0.9–2.2]	**<0.001**
TBSI_Right	0.8 [0.6–1.1]	**<0.001**	1.1 [0.8–1.4]	**<0.001**	1.0 [0.7–1.3]	**<0.001**	0.9 [0.2–1.5]	*0.011*	1.0 [0.5–1.4]	**<0.001**
GIF and STEPS volume difference										
Amygdala_GIF_Left	1.1 [0.0–2.2]	*0.043*	1.2 [0.0–2.4]	*0.046*	3.2 [2.1–4.4]	**<0.001**	3.0 [0.7–5.4]	*0.012*	3.5 [1.9–5.0]	**<0.001**
Amygdala_GIF_Right	2.1 [0.8–3.4]	*0.001*	4.7 [3.4–6.0]	**<0.001**	2.0 [0.76–3.3]	*0.004*	0.8 [-2.0–3.5]	0.573	1.6 [-0.2–3.3]	*0.084*
Hippocampal_STEPS_Left	2.2 [1.2–3.2]	**<0.001**	3.3 [2.2–4.3]	**<0.001**	2.4 [1.4–3.4]	**<0.001**	4.2 [2.0–6.3]	**<0.001**	3.2 [1.9–4.6]	**<0.001**
Hippocampal_STEPS_Right	2.0 [1.1–2.9]	**<0.001**	3.7 [2.7–4.6]	**<0.001**	2.4 [1.4–3.6]	**<0.001**	3.4 [1.4–5.4]	*0.001*	3.0 [1.8–4.3]	**<0.001**
Caudate_STEPS_Left	5.7 [3.7–7.7]	**<0.001**	5.9 [3.8–8.0]	**<0.001**	6.0 [3.8–8.2]	**<0.001**	5.6 [1.1–10.1]	*0.015*	3.8 [0.8–6.8]	*0.012*
Caudate_STEPS_Right	2.4 [1.1–3.8]	*0.001*	2.8 [1.3–4.2]	**<0.001**	2.8 [1.3–4.3]	**<0.001**	1.1 [-1.9–4.1]	0.483	1.4 [-0.6–3.4]	0.164
Thalamus_STEPS_Left	1.6 [0.8–2.3]	**<0.001**	1.8 [0.9–2.6]	**<0.001**	2.7 [1.9–3.6]	**<0.001**	2.7 [0.9–4.4]	*0.003*	1.9 [0.8–3.1]	*0.001*
Thalamus_STEPS_Right	1.1 [0.3–2.0]	*0.008*	-0.1 [-1.0–0.9]	0.898	1.0 [0.0–1.9]	*0.046*	-0.3 [-2.3–1.8]	0.794	0.3 [-0.9–1.6]	0.602

** All analyses adjusted for age, gender and scanner type:* ***p < 0.001 bolded****, p < 0.05 italicised*

Table 5-2 continued: Linear regression analysis results with 95% CI comparing annual rates of subcortical atrophy for each FTD subgroups with controls. *

Measure	Genetic subgroups						Pathology subgroups			
	MAPT		***C9orf72***		***GRN***		**Tau**		**TDP-43**	
Subcortical BSI	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**
ABSI_Left	3.7 [2.2–5.2]	**<0.001**	2.0 [0.1–3.8]	*0.036*	3.2 [1.2–5.1]	*0.002*	3.4 [1.7–5.2]	**<0.001**	3.9 [2.5–5.3]	**<0.001**
ABSI_Right	3.5 [2.3–4.8]	**<0.001**	1.0 [-0.1–2.6]	0.199	0.0 [-1.6–1.6]	0.998	3.6 [1.7–5.4]	**<0.001**	4.3 [2.8–5.7]	**<0.001**
HBSI_Left	2.1 [0.9–3.2]	*0.001*	0.1 [-1.2–1.4]	0.874	3.7 [2.0–5.3]	**<0.001**	2.3 [1.2–3.4]	**<0.001**	2.3 [1.=4–3.1]	**<0.001**
HBSI_Right	2.3 [1.3–3.3]	**<0.001**	0.7 [-0.4–1.8]	0.233	-0.2 [-1.4–1.0]	0.792	2.1 [1.1–3.2]	**<0.001**	1.9 [1.1–2.7]	**<0.001**
CBSI_Left	6.6 [4.1–9.2]	**<0.001**	8.7 [5.5–11.8]	**<0.001**	11.6 [8.3–14.9]	**<0.001**	7.4 [4.8–10.0]	**<0.001**	8.3 [6.3–10.3]	**<0.001**
CBSI_Right	5.1 [3.2–7.0]	**<0.001**	6.3 [4.0–8.5]	**<0.001**	4.7 [2.2–7.1]	**<0.001**	5.2 [3.5–6.9]	**<0.001**	5.0 [3.7–6.3]	**<0.001**
TBSI_Left	1.4 [0.7–2.0]	**<0.001**	1.1 [0.2–1.9]	*0.014*	2.9 [2.0–3.7]	**<0.001**	1.5 [0.9–2.1]	**<0.001**	1.4 [0.9–1.9]	**<0.001**
TBSI_Right	0.9 [0.4–1.3]	**<0.001**	0.2 [-0.3–0.7]	0.456	0.6 [0.0–1.2]	*0.042*	1.1 [0.7–1.5]	**<0.001**	0.7 [0.4–1.0]	**<0.001**
GIF and STEPS volume difference										
Amygdala_GIF_Left	0.9 [-0.3–2.2]	0.141	-0.5 [-2.1–1.1]	0.523	1.3 [-0.3–2.9]	0.115	1.1 [-0.1–2.2]	0.065	0.4 [-0.5–1.3]	0.426
Amygdala_GIF_Right	1.1 [-0.5–2.7]	0.164	2.0 [0.1–3.8]	*0.042*	-0.7 [-2.8–1.5]	0.529	2.1 [0.5–3.8]	*0.010*	2.6 [1.3–3.8]	**<0.001**
Hippocampal_STEPS_Left	0.7 [-0.4–1.8]	0.196	1.7 [0.3–3.0]	*0.016*	4.1 [2.7–5.6]	**<0.001**	1.2 [0.1–2.4]	*0.039*	2.4 [1.4–3.3]	**<0.001**
Hippocampal_STEPS_Right	0.4 [-0.5–1.4]	0.372	1.5 [0.3–2.7]	*0.014*	0.9 [-0.5–2.3]	0.189	0.6 [-0.5–1.7]	0.298	3.0 [2.1–3.9]	**<0.001**
Caudate_STEPS_Left	5.2 [2.2–8.3]	*0.001*	8.0 [4.4–11.6]	**<0.001**	7.0 [3.0–11.1]	*0.001*	5.0 [2.2–7.9]	*0.001*	6.3 [4.1–8.5]	**<0.001**
Caudate_STEPS_Right	3.2 [1.1–5.3]	*0.003*	1.5 [-0.1–4.0]	0.238	4.4 [1.5–7.2]	*0.003*	3.2 [1.3–5.1]	*0.001*	2.5 [1.0–4.0]	*0.001*
Thalamus_STEPS_Left	1.4 [0.4–2.4]	*0.008*	1.7 [0.5–3.0]	*0.008*	3.0 [1.5–4.4]	**<0.001**	1.8 [0.9–2.8]	**<0.001**	2.0 [1.2–2.7]	**<0.001**
Thalamus_STEPS_Right	1.2 [-0.1–2.5]	0.068	1.4 [-0.1–2.9]	0.066	-0.1 [-2.0–1.7]	0.829	1.0 [-0.1–2.1]	0.077	0.6 [-0.3–1.5]	0.185

* All analyses adjusted for age, gender and scanner type: ***p* < 0.001 bolded**, *p < 0.05 italicised*

5.3.2 Effect size and sample size results

Effect size results with 95% BCa CIs demonstrating the magnitude of group difference between controls and each FTD subgroup are reported in Table 5-3. Effect sizes over one are presented in bold because these are within the range considered as representing a large effect. For these bolded results, at least 84% of the control group atrophy rates were below the average rate in the corresponding FTD patient subgroup for that measure, suggesting good group separation.

Table 5-4 reports the sample size calculations with 95% BCa CI to detect a 30% treatment effect with 80% power and 5% statistical significance. These were adjusted for the longitudinal rates of change in the healthy controls for each measure. Sample size estimates under 100 participants required per treatment arm are in blue and bold. Those requiring between 100 and 200 individuals are shown in green and underlined. Measures that provided estimates with a potentially infinite upper 95% BCa CI limit are shown in red. For these measures, multiple repetitions of the 2000 random bootstrap samples failed to detect significant differences in subcortical atrophy rates between patients and controls for that measure and subgroup. Therefore, an accurate calculation of the distribution could not be determined; subsequently, an upper limit for the number of individuals needed to detect a treatment effect with these patients and measures could not be estimated.

5.3.2.1 Clinical subgroups

In line with the whole-brain (Chapter 3) and lobar (Chapter 4) results previously reported, the more heterogeneous bvFTD patients demonstrated overall lower effect sizes than the PPA subgroups. This resulted in none of the 16 measures providing a sample size estimate below 200 individuals per treatment arm. The lowest sample size estimate was provided by the right TBSI at 286, closely followed by the left ABSI, left TBSI and bilateral CBSI measures. Interestingly, despite lower effect sizes, the accompanying 95% BCa CI were relatively narrow and demonstrated a good degree of group

separation from controls as demonstrated by the regression results. Therefore, none of the subcortical measures resulted in an infinite upper 95% CI for the sample size estimate.

In the svPPA subgroup, all the subcortical BSI measures apart from the right HBSI had effect sizes over 1.0 with accompanying low sample size estimates. This was also the case for the indirect longitudinal measures of the right amygdala, bilateral hippocampus and left thalamus. The left CBSI measure produced the lowest sample size estimate of 47 [30–80] patients per treatment arm. The right ABSI, bilateral TBSI and right hippocampal STEPS volume-difference measure also all produced sample size estimates below 100, suggesting that several subcortical measures may prove useful as biomarkers of change in svPPA. The majority of other measures produced estimates between 100 and 200 with only the indirect measures of the left amygdala and right thalamus producing a potentially infinite upper CI estimate. The application of the BSI to these two subcortical measures substantially improved their potential as biomarkers of change.

The majority of effect size calculations in the nfvPPA subcortical measures grouped around the 1.0 with relatively narrow 95% CI and only the right STEPS thalamus measure produced an effect size below 0.5. Whilst this meant that none of the resulting sample sizes produced an estimate of fewer than 100 participants, most of the measures provided feasible estimates in the 100–200 range, with no measure producing a potential infinite 95% upper estimate. The lowest sample size was derived from left hippocampal STEPS volume-difference measure at 125 [78–246]. The other indirect longitudinal measures also showed an asymmetric pattern with only the left amygdala, left caudate and left thalamus providing estimates below 200; however, the 95% CI did overlap to some extent with the right. Following application of the BSI, a more bilateral pattern of biomarker performance emerged with the lowest estimates derived from the left and right CBSI and TBSI. Again, the 95% CI showed substantial overlap suggesting that the

amygdala, hippocampus, caudate and thalamus may provide valuable markers of change in this population.

The PPA-NOS subgroup presented with a strong left-sided asymmetric pattern of atrophy, with all four subcortical regions providing the highest effect size and lowest sample size estimate for that hemisphere. For the BSI measures, all left regions provided low sample size estimates of less than 100, which was also evident for the right CBSI. The lowest sample size was derived from the left HBSI measure which produced an estimate of 12 [6–33] participants required per treatment arm. This was closely followed by the left ABSI (27 [13–54), left CBSI (35 [15–80]) and left TBSI (56 [27–157]). The indirect measures for the left caudate and left amygdala also demonstrated low sample sizes of 39 and 85, respectively, with narrow 95% BCa CIs. These results suggest that all four measures of left subcortical atrophy may prove useful as biomarkers in this newly classified subgroup of FTD patients.

The lvPPA subgroup demonstrated a consistently left-sided asymmetric pattern of volume loss in the amygdala and thalamus. However, the direct and indirect measures gave slightly different results for the caudate and hippocampus. For the CBSI, effect sizes were high and sample sizes low (<100) bilaterally, which was only evident on the left when using the indirect STEPS volume difference. For the hippocampus, the indirect STEPS volume-difference measure provided a strong measure of change bilaterally with the right producing a sample size estimate of 62 and the left requiring 120 participants per treatment arm to detect an effect. Using the HBSI, only the left showed promise with a sample size estimate of 200.

5.3.2.2 Genetic subgroups

Overall effect sizes were lower and sample sizes higher in the genetic subgroups than in the larger clinical subgroups. The indirect measures of change produced high sample sizes with an upper 95% BCa CI limit that

could not be accurately estimated in the majority of measures, with only a few exceptions (Table 5-4).

In the *MAPT* subgroup, several of the BSI measures showed promise as biomarkers of change. The left caudate and left thalamus measures provided the lowest sample size estimate at 80 participants per treatment arm for both BSI outcomes. The ABSI produced relatively low sample size estimates of between 100 and 200 participants; however, the upper limit for the left ABSI was 179,022 participants, demonstrating that caution is required in interpreting the value of this measure at capturing potential disease-modifying effects in a treatment trial.

As can be seen in Figure 5-5, the majority of subcortical measures in the *C9orf72* cohort overlap with the control rates, and thus 13 of the 16 measures produced sample size estimates with the potential of an infinite upper limit of patients required to detect the treatment effect. Only the right CBSI and right hippocampal STEPS volume-difference measures provided estimates under 200 patients per treatment arm; however, the upper limit for the latter was 7,671, suggesting that right hippocampal measures may not be suitable in this population.

Finally, the subcortical measures in the *GRN* subgroup did not perform well at separating the patient group from controls and hence failed to provide feasible sample size estimates for all 16 subcortical longitudinal measures. Whilst the left ABSI, left HBSI, left CBSI and left caudate volume loss, based on the STEPS segmentations, provided estimates below 200, the upper potential 95% CI limit for these measures suggest that considerable caution in interpreting their utility as longitudinal biomarkers for this population is required. Even the lowest estimate, which was derived from the left ABSI at 65, had a 95% BCa CI of 14–4609, suggesting that this is likely to be an unreliable measure of atrophy and even less likely to detect a subtle treatment-related change to that atrophy rate.

5.3.2.3 Pathology subgroups

Given the substantial overlap of patients, the Tau subgroup results closely reflected those of the *MAPT* subgroup. Except for the left STEPS caudate measure, the indirect GIF/STEPS longitudinal volume-difference measures failed to provide feasible sample size estimates, especially when considering the upper 95% CI limit. The more robust BSI measures provided sample size estimates below 200 for seven of the eight measures, except for the left HBSI. The right TBSI provided the lowest estimate at 69 [40–176] patients required per treatment arm. In the TDP-43 group, the CBSI bilaterally and right hippocampal STEPS volume-difference measures performed the best, providing sample size estimates below 200.

Table 5-3 Effect size calculations with 95% BCa confidence intervals for each of the subcortical longitudinal measures by FTD subgroup*

Measure	Clinical subgroups									
	bvFTD		svPPA		nfvPPA		PPA-NOS		lvPPA	
Subcortical BSI	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)
ABSI_Left	0.8	[0.5–1.0]	**1.3**	**[0.9–1.6]**	**1.0**	**[0.8–1.2]**	**2.5**	**[1.8–3.7]**	0.8	[0.3–1.1]
ABSI_Right	0.6	[0.4–0.9]	**1.5**	**[1.1–1.9]**	0.8	[0.5–1.0]	0.6	[-0.6–1.5]	0.6	[0.0–0.9]
HBSI_Left	0.6	[0.3–0.9]	**1.3**	**[0.9–1.6]**	0.9	[0.6–1.1]	**3.8**	**[2.3–5.4]**	0.9	[0.5–1.4]
HBSI_Right	0.5	[0.3–0.7]	0.8	[0.5–1.1]	0.9	[0.5–1.2]	0.9	[0.2–1.5]	0.3	[-0.2–0.6]
CBSI_Left	0.7	[0.6–0.9]	**1.9**	**[1.5–2.4]**	**1.0**	**[0.8–1.3]**	**2.2**	**[1.5–3.4]**	**1.9**	**[1.3–2.4]**
CBSI_Right	0.8	[0.5–1.0]	**1.2**	**[0.8–1.5]**	**1.1**	**[0.9–1.4]**	**1.6**	**[1.0–2.3]**	**1.6**	**[1.0–2.1]**
TBSI_Left	0.8	[0.4–1.0]	**1.7**	**[1.4–2.1]**	**1.0**	**[0.8–1.3]**	**1.8**	**[1.1–2.6]**	**1.6**	**[1.0–2.2]**
TBSI_Right	0.8	[0.5–1.1]	**1.5**	**[1.1–1.9]**	**1.2**	**[0.6–1.7]**	0.7	[-0.7–1.7]	0.8	[0.1–1.4]
GIF and STEPS volume difference										
Amygdala_GIF_Left	0.4	[0.1–0.7]	0.3	[0.0–0.6]	0.9	[0.6–1.3]	**1.4**	**[1.0–2.0]**	**1.0**	**[0.5–1.4]**
Amygdala_GIF_Right	0.5	[0.3–0.8]	**1.2**	**[0.7–1.6]**	0.6	[0.2–0.9]	0.4	[-0.6–1.4]	0.5	[0.0–0.9]
Hippocampal_STEPS_Left	0.6	[0.3–0.8]	**1.2**	**[0.8–1.6]**	**1.2**	**[0.8–1.5]**	**1.3**	**[0.1–3.6]**	**1.2**	**[0.7–1.7]**
Hippocampal_STEPS_Right	0.6	[0.4–0.8]	**1.5**	**[0.8–2.1]**	0.8	[0.5–1.0]	**1.2**	**[0.1–2.3]**	**1.7**	**[1.2–2.3]**
Caudate_STEPS_Left	0.7	[0.4–0.8]	0.9	[0.6–1.1]	0.9	[0.6–1.2]	**2.1**	**[1.4–2.6]**	0.9	[0.4–1.5]
Caudate_STEPS_Right	0.5	[0.3–0.8]	0.7	[0.4–1.1]	0.6	[0.3–0.9]	0.4	[-0.7–1.9]	0.4	[-0.2–0.9]
Thalamus_STEPS_Left	0.6	[0.3–0.9]	**1.1**	**[0.7–1.5]**	**1.1**	**[0.6–1.4]**	**1.1**	**[0.4–1.9]**	0.8	[0.4–1.2]
Thalamus_STEPS_Right	0.4	[0.1–0.7]	0.0	[-0.4–0.4]	0.4	[0.1–0.8]	-0.2	[-1.3–1.4]	0.2	[-0.3–0.7]

* Effect sizes of 1.0 or above are in **bold** to indicate a high degree of group separation from controls

Table 5-3 continued: Effect size calculations with 95% BCa confidence intervals for each of the subcortical longitudinal measures by FTD subgroup

Measure	Genetic subgroups						Pathology subgroups			
	MAPT		***C9orf72***		***GRN***		**Tau**		**TDP-43**	
Subcortical BSI	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)
ABSI_Left	**1.0**	**[0.0–1.5]**	0.5	[-0.3–1.0]	**1.6**	**[0.2–3.6]**	**1.2**	**[0.5–2.0]**	0.8	[0.5–1.0]
ABSI_Right	**1.1**	**[0.6–1.9]**	0.6	[-0.2–1.3]	0.0	[-0.8–1.0]	**1.2**	**[0.7–1.8]**	0.7	[0.4–1.1]
HBSI_Left	0.6	[-0.1–1.3]	0.2	[-0.7–0.9]	**1.2**	**[0.4–5.9]**	0.7	[-0.1–1.4]	0.9	[0.3–1.4]
HBSI_Right	0.9	[0.5–1.2]	0.5	[-0.3–1.4]	-0.1	[-1.0–0.8]	0.9	[0.6–1.3]	0.7	[0.3–1.1]
CBSI_Left	**1.5**	**[0.8–2.2]**	0.7	[0.4–1.0]	**1.2**	**[0.5–2.4]**	**1.3**	**[0.8–1.8]**	0.9	[0.7–1.2]
CBSI_Right	0.9	[0.2–1.6]	**1.1**	**[0.5–1.7]**	0.7	[0.0–1.5]	**1.1**	**[0.3–1.7]**	**1.0**	**[0.7–1.4]**
TBSI_Left	**1.5**	**[0.9–2.0]**	0.6	[-0.2–1.0]	0.9	[-0.3–3.1]	**1.6**	**[1.0–2.1]**	0.7	[0.2–1.0]
TBSI_Right	**1.0**	**[0.5–1.6]**	0.4	[-0.4–1.2]	0.4	[-1.0–1.5]	**1.2**	**[0.7–1.9]**	0.7	[0.1–1.1]
GIF and STEPS volume difference										
Amygdala_GIF_Left	0.4	[-0.3–0.8]	0.0	[-1.0–0.8]	0.3	[-0.6–1.9]	0.5	[-0.1–0.9]	0.2	[-0.2–0.6]
Amygdala_GIF_Right	0.4	[-0.2–0.9]	0.5	[-0.3–1.0]	-0.3	[-1.2–0.8]	0.6	[0.1–1.1]	0.7	[0.3–1.0]
Hippocampal_STEPS_Left	0.2	[-0.4–0.8]	0.7	[-0.1–1.4]	0.9	[0.2–2.0]	0.4	[-0.2–0.9]	0.8	[0.4–1.1]
Hippocampal_STEPS_Right	0.2	[-0.4–0.7]	**1.0**	[0.2–1.8]	0.3	[-0.6–1.4]	0.2	[-0.3–0.8]	**1.0**	**[0.6–1.4]**
Caudate_STEPS_Left	0.9	[0.3–1.4]	0.5	[-0.3–0.8]	**1.0**	**[0.1–1.8]**	0.9	[0.4–1.4]	0.6	[0.3–0.9]
Caudate_STEPS_Right	0.6	[0.0–0.9]	0.4	[-0.3–1.0]	0.8	[-0.2–1.5]	0.6	[0.2–1.0]	0.6	[0.3–1.0]
Thalamus_STEPS_Left	0.7	[0.0–1.3]	0.7	[-0.1–1.5]	0.9	[-0.5–2.6]	**1.0**	**[0.2–1.7]**	0.9	[0.4–1.3]
Thalamus_STEPS_Right	0.5	[-0.2–1.3]	0.6	[-0.1–1.1]	-0.1	[-1.0–1.8]	0.4	[-0.2–1.0]	0.3	[-0.2–0.6]

* Effect sizes of 1.0 or above are in **bold** to indicate a high degree of group separation from controls

Table 5-4 Sample size estimates with 95% BCa CI to detect a 30% reduction in atrophy rate with 80% power and 5% significance for subcortical measures split by FTD subgroups. Results in **blue and bold** are for measures that required <100 participants per treatment arm, green and underlined demonstrate sample size estimates between 100 and 200 individuals per treatment arm, and those highlighted in red have an infinite or unfeasible upper 95% confidence interval produced using 2000 bootstrap replications.

Measure	Clinical subgroups									
	bvFTD		svPPA		nfvPPA		PPA-NOS		lvPPA	
Subcortical BSI	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)
ABSI_Left	294	[168–701]	109	[67–211]	177	[117–305]	**27**	**[13–54]**	289	[136–1495]
ABSI_Right	454	[235–1273]	**78**	**[48–132]**	274	[165–617]	433	[73–∞]	533	[194–1.9e+07]
HBSI_Left	477	[230–1633]	105	[66–208]	221	[135–471]	**12**	**[6–33]**	200	[95–826]
HBSI_Right	640	[316–2097]	272	[154–720]	218	[118–637]	210	[74–4964]	2223	[439–∞]
CBSI_Left	337	[231–549]	**47**	**[30–80]**	167	[107–305]	**35**	**[15–80]**	**50**	**[29–106]**
CBSI_Right	302	[187–577]	128	[80–250]	136	[88–224]	**70**	**[32–192]**	**71**	**[41–166]**
TBSI_Left	306	[165–876]	**57**	**[38–93]**	161	[102–293]	**56**	**[27–157]**	**66**	**[35–183]**
TBSI_Right	286	[157–803]	**80**	**[50–144]**	127	[59–542]	369	[61–∞]	242	[84–20904]
GIF and STEPS volume difference										
Amygdala_GIF_Left	1011	[370–12649]	1683	[429–∞]	200	[108–517]	**85**	**[43–192]**	182	[95–639]
Amygdala_GIF_Right	608	[287–2536]	132	[71–351]	557	[239–4367]	958	[88–∞]	732	[199–1.0e+08]
Hippocampal_STEPS_Left	523	[286–1440]	121	[67–274]	125	[78–246]	107	[14–44578]	120	[64–358]
Hippocampal_STEPS_Right	427	[246–966]	**73**	**[39–307]**	302	[160–786]	128	[34–13827]	**62**	**[34–123]**
Caudate_STEPS_Left	392	[244–929]	239	[145–484]	196	[113–425]	**39**	**[26–89]**	197	[73–1091]
Caudate_STEPS_Right	671	[306–2710]	334	[154–1423]	488	[215–2448]	1115	[50–∞]	1135	[204–∞]
Thalamus_STEPS_Left	500	[239–1587]	144	[74–356]	157	[89–418]	138	[48–1184]	287	[130–1033]
Thalamus_STEPS_Right	891	[324–8130]	5053000	[1134–∞]	898	[270–65797]	4167	[85–∞]	4022	[368–∞]

Table 1-4 continued Sample size estimates with 95% BCa CI to detect a 30% reduction in atrophy rate with 80% power and 5% significance for subcortical measures split by FTD subgroups. Results in **blue and bold** are for measures that required <100 participants per treatment arm, green and underlined demonstrate sample size estimates between 100 and 200 individuals per treatment arm, and those highlighted in red have an infinite or unfeasible upper 95% confidence interval produced using 2000 bootstrap replications.

Measure	Genetic subgroups						Pathology subgroups			
	MAPT		*C9orf72*		*GRN*		Tau		TDP-43	
Subcortical BSI	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)
ABSI_Left	175	[83–1.8e+05]	590	[177–∞]	**65**	**[14–4609]**	117	[43–788]	277	[163–678]
ABSI_Right	132	[50–471]	576	[109–∞]	409119	[182–∞]	124	[55–344]	315	[156–1160]
HBSI_Left	448	[106–∞]	6607	[225–∞]	115	[5–1203]	321	[95–∞]	212	[94–1674]
HBSI_Right	216	[116–635]	833	[86–∞]	51915	[295–∞]	197	[110–467]	346	[149–2003]
CBSI_Left	**80**	**[37–251]**	370	[169–1120]	118	[31–681]	107	[55–274]	198	[120–366]
CBSI_Right	202	[68–5082]	135	[61–648]	332	[78–90305]	157	[58–2642]	160	[86–403]
TBSI_Left	**80**	**[44–233]**	510	[173–∞]	203	[18–∞]	**69**	**[40–176]**	348	[182–2990]
TBSI_Right	169	[71–783]	1058	[112–∞]	1160	[76–∞]	112	[48–331]	347	[139–11468]
GIF and STEPS volume difference										
Amygdala_GIF_Left	1125	[244–∞]	118701	[308–∞]	1969	[47–∞]	740	[203–∞]	4564	[461–∞]
Amygdala_GIF_Right	1078	[222–∞]	785	[175–∞]	1993	[242–∞]	488	[152–48385]	404	[166–2303]
Hippocampal_STEPS_Left	4156	[299–∞]	385	[88–∞]	206	[42–6517]	1176	[237–∞]	302	[140–1039]
Hippocampal_STEPS_Right	6590	[358–∞]	177	[55–7671]	2027	[83–∞]	2848	[268–∞]	180	[90–496]
Caudate_STEPS_Left	213	[91–2069]	717	[259–∞]	183	[56–7909]	200	[93–1111]	442	[238–1433]
Caudate_STEPS_Right	465	[193–1.6e+05]	1128	[188–∞]	257	[73–∞]	416	[169–6186]	423	[185–2480]
Thalamus_STEPS_Left	352	[96–5.9e+06]	326	[75–∞]	230	[26–∞]	192	[63–5370]	238	[101–1171]
Thalamus_STEPS_Right	670	[103–∞]	550	[143–∞]	36337	[53–∞]	1025	[164–∞]	2371	[433–∞]

5.4 Discussion

The current study investigated fully automated longitudinal atrophy rates in the amygdala, hippocampus, caudate and thalamus across clinical, genetic and pathology-confirmed FTD subgroups. Both the GIF and STEPS methods produced accurate segmentations for these subcortical regions closely matching the boundaries of these structures. For many of the structures and subgroups, both direct and indirect measures of subcortical change provided good group separation from controls, high effect sizes and low feasible sample sizes to detect a 30% reduction in atrophy rate (with 80% power, 5% statistical significance). The overall pattern of subcortical regional involvement differed across the subgroups, but reassuringly for the majority of groups, several measures provided sample size estimates substantially below 100 patients per treatment arm. Overall, measures of caudate and thalamus atrophy tended to perform best across the subgroups. The left CBSI and left TBSI provided the lowest or equally low sample size estimates for the svPPA, PPA-NOS, lvPPA, *MAPT* and Tau subgroups. The right CBSI also performed well across the PPA subgroups. The caudate may perform so well due to its location against the ventricular wall, making automated segmentation at that boundary as well as application of the intensity windows for the BSI most favourable for this structure. Its shape is also less problematic than the hippocampus to discern, particularly when atrophied with easier boundaries to differential than for the amygdala. The good performance of thalamic atrophy measures (particularly on the left) across most FTD subgroups is an interesting finding and in line with the majority of cross-sectional investigations. Increased atrophy in this region may be due to the widely distributed connections it has to all regions of the cortex, midbrain and subcortical circuitry. This assertion is based on the prevailing theory of ‘prion-like’ propagation of pathology to interconnected neighbouring neurons (Goedert, Eisenberg, & Crowther, 2017). Given its role as the hub of the brain connecting a vast array of neuronal populations, it may be uniquely vulnerable to this pathological cascade irrespective of which regions in the brain are initially affected in each of the different FTD subgroups.

5.4.1 Clinical subgroups

Reduced volumes in bvFTD cohorts have been previously reported for the amygdala, hippocampus, caudate and thalamus when compared cross-sectionally with control volumes (Barnes et al., 2006; Boccardi et al., 2002; Bocchetta et al., 2018; Garibotto et al., 2011; Hornberger et al., 2012). Currently, the only longitudinal study to report annual rates in these subcortical regions in bvFTD is a recent study from Landin-Romero and colleagues (2017). Table 5-5 shows the values for the four subcortical structures reported in their study and the current results using the BSI measures for comparison. It is important to note that the disease duration for their bvFTD cohort was shorter than the current cohort (4.1 years compared with 5.9 years in the current cohort) and gender ratio had a slightly more even distribution (23M:14F compared with the current cohort with 52M:14F). Despite the cohort differences and the use of a combined left and right measure, it is interesting to note some similarities in the pattern of atrophy across the structures, specifically the high rates of change in the caudate. Data on the standard deviation for Landin-Romero et al. (2017) annual rates of subcortical change were not available making more in-depth direct comparison difficult.

Table 5-5 Annual rates of subcortical volume loss as previously published in Landin-Romero et al. (2017) and the current BSI measures for each structure split by hemisphere

Structure	Annual percentage volume loss*	Subcortical BSI_left	Subcortical BSI_right
Amygdala	2.1	3.2 (4.1)	2.9 (4.7)
Hippocampus	3.6	2.4 (3.6)	1.7 (3.2)
Caudate	10.0	6.5 (8.5)	5.9 (7.1)
Thalamus	3.7	1.3 (1.5)	0.9 (1.0)

* calculated from Landin-Romero et al. (2017) supplementary material and correspondence with the author.

The relatively high standard deviations for the current annual rates of change in the 16 longitudinal measures resulted in less sensitive measures of change in the bvFTD population, as demonstrated by overall higher sample size estimates. It is important to note that many of the BSI measures did provide estimates close to 300, so not entirely out of the feasible range. Given the characteristic heterogeneity, enrolment based on a clinical diagnosis of bvFTD may be problematic to properly power the detection of a meaningful treatment effect using these subcortical measures. An important issue when considering clinical trial design, particularly with a heterogeneous population like bvFTD, is the potential for enrichment by stratifying participants to include those most likely to demonstrate the best response. In the case of bvFTD participants, that could be to enrol those with measurable rates of change during a predetermined run-in phase, so that individuals are randomised if they demonstrate (among other important factors) measurable volume loss within this initial phase. Theoretically, this should increase the likelihood of detecting a treatment effect, albeit with the associated potential cost to external and internal validity by restricting the enrolment population using this criterion. Such a design would also avoid the confounds of reduced power by the inclusion of bvFTD phenocopies who do not appear to have a degenerative condition (Valente et al., 2019). A short interval and back-to-back imaging study similar to the Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) study would provide valuable information in FTD as to how many months the run-in phase should be to reliably detect a change, what the acceptable range of loss for randomisation should be and the reliability of the segmentation and longitudinal measures by using the back-to-back imaging (Malone et al., 2013).

The results of the current study closely reflect that expected from the svPPA literature. Measures of all eight of these structures in the current study showed high rates of annual change, high effect sizes and low feasible sample sizes using the BSI (and many of the indirect measures), suggesting that all subcortical regions could provide useful measures in upcoming

putative trials. Comparing the HBSI outcomes with those reported in Rohrer et al. (2008) showed good correspondence for the left (104 versus 105 in the current study), but a less favourable result for the right (81 compared with 272 in the current study). Interestingly, the indirect measure derived from the right hippocampal STEPS segmentations did provide a comparable result of 73 patients required per treatment arm. It is unclear from reviewing the registrations why this discrepancy occurred on the right and not the left, other than the left regions having a larger amount of surface area surrounded by CSF due to the substantial atrophying of the regions. This could have resulted in a higher estimate of change from a larger proportion being contributed by the CSF–GM intensity window, which often produces a clearer boundary shift result than the GM–WM boundary. However, the segmentation accuracy was better on the right because these regions were less atrophied and included less mis-segmentation. Further investigation would be useful to clarify this minor discrepancy.

Previous investigations with nfvPPA are more limited and have often proven inconsistent in terms of which subcortical region may provide a signal of longitudinal change. Despite this, the current study demonstrates that both direct and indirect measures across all four structures show some potential as a future biomarker. The lowest sample size was produced by the left STEPS indirect hippocampal measure at 125, closely followed by the right TBSI at 127 patients per treatment arm. However, the upper limit for the latter was 542, again demonstrating the importance of taking not only the point estimate but also the full range of potential values into account when reviewing biomarker options and informing trial design. Whilst the majority of investigations into patterns of atrophy in nfvPPA have focussed on cortical GM loss, the reported spread of volume loss within the temporal lobe in these studies corresponds well to the involvement of the amygdala and hippocampus found in the current study. In addition, the caudate has specifically been shown to be atrophied bilaterally as the disease progresses into the later stages (Gorno-Tempini et al., 2004; Rogalski et al., 2011) in line with the current results. Patients in this cohort were an average

of 4.4 years after symptom onset with a range from 0.9 to 10 years. It is therefore likely that the cohort represents the full range of regional atrophy previously reported to occur in this patient population, which the current results support showing regional involvement from both earlier and later stages previous reported.

Except for investigations into thalamic atrophy (Bocchetta, Gordon et al., 2018, Bocchetta et al., 2020), subcortical regional involvement in PPA-NOS has not previously been studied. The current results in this subgroup provide strong evidence for subcortical involvement across the amygdala, hippocampus, caudate and thalamus. A clear left-sided dominance across all these subcortical regions was evident. These results replicate and extend the cross-sectional findings reported in Bocchetta, Gordon and colleagues (2018), where an asymmetric pattern of thalamic volume reduction also demonstrated a clear left bias with 12% reduction on the left and 2% on the right compared with control volumes. It is important to note that all seven of the participants in this longitudinal cohort contributed data to both this study and the following Bocchetta et al. (2020) cross-sectional study investigating the separate nuclei of the thalamus. The current results demonstrate that not only the presentation but also the progression over time maintain this asymmetric bias and that this asymmetric atrophy can be measured across several key subcortical structures. It is important to reiterate that these results are based on a small cohort (n = 7), who demonstrated remarkably homogeneous rates of subcortical atrophy. Caution is therefore advised in interpreting the wider application of these results, particularly in terms of the estimated BCa CIs. The left HBSI provides the smallest sample size estimate in this population for all the subcortical measures across all subgroups at 12 [6–33]. All other left BSIs and the right CBSI provide estimates well below 100, providing not only additional information about the neuroanatomical distribution within this newly emergent population but also that the rate of progression in these patients appears to be high and relatively uniform.

Studies into lvPPA have focused on the cortical pattern of atrophy and distribution of amyloid pathology, showing that it closely reflects that of AD patients centred around the temporoparietal junction, a region that also reflects the particular language deficits seen in this population (Henry & Gorno-Tempini, 2010). Investigations into subcortical atrophy, however, are limited and inconsistent, particularly with regard to involvement of the hippocampus. Despite this, the current study has demonstrated the potential of several subcortical measures, most notably bilateral CBSI, left TBSI and bilateral hippocampal STEPS measures in this cohort to detect longitudinal change.

5.4.2 Genetic subgroups

Taking into account the smaller cohort sizes and likely reduced power to detect an effect, the results for the genetic groups were still slightly disappointing and somewhat in conflict with those expected from the literature. In *MAPT*, clear atrophy in the amygdala, hippocampus and thalamus has been previously reported (Bocchetta, Iglesias et al., 2018, Bocchetta, Gordon et al.,2018; Bocchetta, Iglesias et al., 2019; Cash et al., 2018; Mahoney et al., 2012; Whitwell et al., 2012), including presymptomatically (Rohrer et al., 2015). The marked symmetrical loss of hippocampal volume frequently seen in *MAPT* can result in misdiagnosis as early-onset AD (Ghetti et al., 2015; Spina et al., 2008), suggesting that this marker should have performed well at detecting a change in this population. Unfortunately, both the left HBSI and bilateral hippocampal STEPS measures produced low rates of change relative to standard deviation, low effect sizes and high sample size estimates, where an upper confidence interval could not be estimated. Visual review of these cases demonstrated striking bilateral hippocampal atrophy in these patients already present at baseline, which may limit the amount of atrophy able to be detected at follow-up. The disease duration of the *MAPT* patients in the current cohort was an average of 6.3 (3.9) years. Because hippocampal atrophy has been detected up to 15 years before symptom onset (Rohrer et al., 2015) and is already substantial at the point of diagnosis, this may be a better marker in

presymptomatic or early-disease stage trials, which will be explored in the following chapter. Encouragingly though, despite limited literature on caudate involvement in *MAPT*, both the CBSI and TBSI measures (particularly on the left) performed well at detecting longitudinal change and produced low sample size estimates of 80 participants per treatment arm for both left measures. The ABSI showed some promise with point estimates of sample sizes below 200; however, the left ABSI upper confidence interval is prohibitively large, indicating a high level of uncertainty for trial design. Visual assessment of the *MAPT* patients demonstrated they presented with substantially reduced amygdala at baseline (the left more reduced than the right), which may provide some indication as to why the longitudinal rates for the left ABSI and both indirect amygdala measures did not perform as well.

With high standard deviations relative to mean annual rates and low subsequent effect sizes, the longitudinal subcortical measures in the *C9orf72* population performed relatively poorly across the regions. Figure 5-5 demonstrates this with a clear overlap of 95% CI with control rates. This is except for the right CBSI measure. Based on the literature, this is somewhat in line with what would be expected. Compared with the other genetic subgroups, several studies have reported that subcortical rates of change in the *C9orf72* subgroup are lowest, or not significantly different to controls in the hippocampus (Bocchetta, Iglesias et al., 2018; Mahoney et al., 2012), amygdala (Bocchetta, Iglesias, Cash et al., 2019) and caudate (Cash et al., 2018); although thalamic atrophy would be expected because it is most associated with the *C9orf72* expansion both symptomatically (Bocchetta, Gordon et al., 2018; Cash et al., 2018) and presymptomatically (Cash et al., 2018; Rohrer et al., 2015). A more detailed investigation into thalamic markers in *C9orf72* would be of value in the future. Growing evidence suggests that reductions in the medial pulvinar region are exclusive to *C9orf72* patients (Bocchetta et al., 2020; Lee et al., 2014, 2017). Application of a probabilistic atlas of the thalamic nuclei, such as that described in Iglesias et al. (2018), would allow for investigating changes in

this particular region in isolation and may prove valuable as future work to assess whether this provides a more specific and sensitive marker in this population.

As would be expected from the literature, the subcortical markers in the *GRN* subgroup showed an asymmetric pattern, with the left markers proving more sensitive and reliable at detecting longitudinal change. Again, only a few markers provided feasible sample sizes of under 200 participants, including the left ABSI, left HBSI, left CBSI and left caudate STEPS volume-difference measure. Unfortunately, the majority of these also demonstrated wide 95% BCa CI with large potential upper limits to the number of patients that may be required using these markers to find a 30% reduction in the atrophy rate. These wide confidence intervals cannot be explained by different patterns of asymmetry increasing variability as all the *GRN* patients had a left dominant atrophy profile (which is not always the case). Further investigations with a larger cohort would be of use to clarify the utility of these subcortical markers in *GRN* patients and will form a key focus of the analysis in the following genetic FTD chapter.

5.4.3 Pathology subgroups

Finally, the results for the larger pathology-confirmed subgroups showed a stronger signal of change across many of the subcortical regions, particularly after application of the BSI. The Tau group closely reflected the pattern of the *MAPT* results, with the encouraging addition of substantially reduced upper limits for the 95% BCa CI across the promising BSI measures. Of these BSI measures, only the left HBSI failed to provide a feasible sample size estimate. As with the *MAPT*, this is likely due to the substantial volume loss already evident as the baseline. The results for the TDP-43 showed fewer stand out measures (i.e. those producing sample size estimates below 100), but instead showed a more distributed pattern of potential across the measures, particularly for the CBSI bilaterally. A somewhat disappointing result was for the thalamus because patients with TDP-43 pathology have been shown to present with substantial involvement

of the thalamus. Depending on the subtype, TDP-43 patients show distinctly different patterns of thalamic nuclei involvement, with the mediodorsal region the only nuclei overlapping between Types A and C (Bocchetta et al., 2020). Again, a more targeted automated segmentation of the thalamic nuclei may prove a more useful marker of change for this population.

Longitudinal investigations of subcortical change are lacking in the FTD population. The current study has contributed to closing these knowledge gaps by demonstrated that automated measures of atrophy in the amygdala, hippocampus, caudate and thalamus show significant promise as potential biomarkers to detect treatment effects of disease-modifying therapies across the FTD spectrum. This was particularly true for measures of the caudate and thalamus. Given the reduced cohort sizes in the genetic subgroups, it is possible that results for the subcortical measures were underpowered to detect a meaningful effect in this group. The following chapter aims to address this by applying the best whole-brain measure from Chapter 3 and these current subcortical measures in a large genetic FTD cohort to assess their utility in both symptomatic and presymptomatic individuals.

6 Automated global and subcortical measures in genetic FTD

The three previous data chapters have explored the utility of global, lobar and subcortical measures of change across the clinical, genetic and pathology symptomatic FTD spectrum. Results from all three chapters have demonstrated that automated measures of atrophy across these regions have shown promise as potential biomarkers, demonstrating good group separation from controls and low sample size estimates to detect a meaningful treatment effect. The utility of these biomarkers has been shown to vary across FTD subgroup, brain region and method of measurement (i.e. direct or indirect calculation of longitudinal change). Currently, several treatment trials in genetic FTD are in the early stages, with several promising candidate molecules in development. Initial interventions are likely to target specific proteinopathies and thus enrolling participants that have (or are likely to develop) that particular pathological substrate will increase the potential for detecting a positive result. The widely documented failure to find treatment effects in several promising AD trials (Cummings, Lee, Ritter, Sabbagh, & Zhong, 2019) may have been exacerbated by enrolment based on clinical diagnosis, given that a substantial number of clinically diagnosed AD patients do not have underlying AD pathology at post-mortem (Beach, Monsell, Phillips, & Kukull, 2012). The use of genetic cohorts could help avoid this pitfall (Bateman et al., 2012; Morris et al., 2012). By enrolling participants with a confirmed genetic mutation, then the overall type of proteinopathy (Tau versus TDP-43) would be known and assure that an appropriate therapy was being used for that group. The use of genetic cohorts has another important advantage. It allows for investigations involving presymptomatic individuals, thus providing a unique opportunity to better understand and hopefully intervene in the earliest stages of the disease process. Ideally, interventions should be targeted to slow or halt the progression of the disease before symptom onset and the extensive neurological damage that has already occurred by this point. Structural imaging biomarkers show promise in detecting volume differences between mutation carriers and non-carrier controls in several regions, several years before symptom onset (Borroni et al., 2008; Cash et al., 2018; Dopper et al.,

2014; Lee et al., 2017; Panman et al., 2019; Papma et al., 2017; Popuri et al., 2018; Rohrer et al., 2015). It remains unclear whether these early volume differences in presymptomatic carriers translate to quantifiable longitudinal changes over time that can be detected using automated volumetric measures in *MAPT*, *C9orf72* or *GRN* carriers and thus provide viable biomarkers for presymptomatic trials. In addition, even within the symptomatic individuals, it remains unclear which longitudinal measures may provide the most sensitive markers for each of the genetic FTD subgroups. To address this, the current study investigates both direct and indirect automated measures of whole-brain and subcortical volume change in a longitudinal genetic FTD cohort.

6.1 Introduction

6.1.1 Neuroanatomical changes in *MAPT*

In presymptomatic *MAPT* mutation carriers, GM volume reductions have been reported 10 years before symptom onset in the anterior and medial temporal lobes, orbitofrontal lobe and insula (Cash et al., 2018; Panman et al., 2019; Rohrer et al., 2015). In terms of subcortical involvement, reduced hippocampal volumes have been observed bilaterally, 15 years before expected symptom onset (Rohrer et al., 2015) and a longitudinal VBM study demonstrated reduced annual GM density in the right hippocampus (p = 0.049; cluster size of 905 voxels) in a presymptomatic cohort (Panman et al., 2019). However, several studies have also failed to find any volumetric differences between *MAPT* carriers and non-carriers (Dopper et al., 2014; Whitwell, Josephs et al., 2011). It is important to note that these studies all employed different methods, variable cohort sizes and presymptomatic individuals at varying stages on the preclinical continuum, so further comprehensive investigation in well-characterised presymptomatic cohorts is required to address these inconsistencies. By the time *MAPT* patients become symptomatic, they present primarily with focal symmetrical medial temporal and orbitofrontal lobe atrophy, with involvement of the insula and anterior cingulate frequently reported. Subcortically, bilateral reductions in

hippocampal volumes are a striking feature in *MAPT* patients and likely to provide a promising marker in the current genetic cohort. Reductions in the caudate, amygdala and thalamus have also been demonstrated cross-sectionally and may prove useful as a longitudinal measure (Bocchetta, Gordon et al., 2018, Bocchetta, Iglesias, Cash et al., 2018; Cash et al., 2018; Gorno-Tempini et al., 2004; Mahoney et al., 2012; Whitwell et al., 2012). Comprehensive longitudinal investigations are currently lacking in *MAPT* and are required to assess the utility of volumetric regional measures as potential future biomarkers for therapeutic interventions in this population.

6.1.2 Neuroanatomical changes in *C9orf72*

C9orf72 expansion carriers demonstrate early GM volume loss, reported in individuals younger than 40 years of age (Bertrand et al., 2018), with brain volume loss evident potentially more than 25 years before symptom onset (Rohrer et al., 2015). Previous VBM studies have demonstrated reduced GM in the thalamus, cerebellum, cingulate, insula, left frontal, parietal and temporal lobes (Cash et al., 2018; Lee et al., 2017; Panman et al., 2019; Papma et al., 2017), albeit not uniformly across all publications. One finding that was consistently reported was involvement of the thalamus. Cortical thickness investigations have also demonstrated reduced cortical GM in *C9orf72* carriers compared with non-carriers in the frontal, temporal, parietal and occipital cortices (Panman et al., 2019; Popuri et al., 2018; Walhout et al., 2015). Despite showing reduced GM volumes and cortical thinning in several regions cross-sectionally, a recent study failed to show any difference between *C9orf72* carriers and non-carriers in their longitudinal VBM and cortical thickness measures using the same cohort (Panman et al., 2019). By the time they show symptoms, *C9orf72* patients typically present with a distributed symmetric pattern of atrophy, primarily involving dorsolateral, medial frontal and orbitofrontal lobes. Additional volume loss has also been reported in the anterior temporal, parietal and occipital lobes, as well as the insula, cingulate and cerebellum (Cash et al., 2018; Mahoney et al., 2012; Sha et al., 2012). Subcortically, bilateral thalamic atrophy has been particularly emphasised as a feature, with cross-sectional analyses

demonstrating significantly reduced thalamic volumes compared with controls in symptomatic (Bocchetta, Gordon et al., 2018; Cash et al., 2018; Lee et al., 2014), and presymptomatic individuals (Cash et al., 2018; Lee et al., 2017; Rohrer et al., 2015). Longitudinal investigations are lacking and with such a wide distribution of neuroanatomical involvement, it remains unclear which regional measures may provide the most sensitive marker of change in this population, requiring further investigation.

6.1.3 Neuroanatomical changes in *GRN*

In *GRN* carriers, presymptomatic volume loss has been observed in frontal, parietal and insular cortices as well as the striatum 10 years before symptom onset (Cash et al., 2018; Rohrer et al., 2015). Cortical thickness investigations have reported reduced cortical GM at baseline in bilateral orbitofrontal, insular, and anterior temporal cortices. In addition, the same study also showed longitudinal cortical GM changes in the right orbitofrontal and left occipital cortices compared with controls (Olm et al., 2018). However, results remain inconsistent, with several investigations comparing *GRN* carriers with non-carriers failing to show any GM volume or cortical thickness differences both cross-sectionally and longitudinally (Borroni et al., 2008; Dopper et al., 2014; Panman et al., 2019; Popuri et al., 2018). Symptomatic *GRN* patients present with a marked asymmetric pattern of atrophy throughout the temporal, inferior frontal and inferior parietal lobes (Cash et al., 2018; Premi et al., 2014; Rohrer et al., 2010; Whitwell et al., 2015) and reported having the fastest global rate of progression compared with the other genetic subgroups (Whitwell et al., 2015). Subcortical involvement of the caudate has been particularly highlighted as a feature of patients with the *GRN* mutation (Cash et al., 2018; Premi et al., 2014) as well as atrophy in the hippocampus and the thalamus both cross-sectionally (Bocchetta, Gordon et al., 2018, Bocchetta et la., 2020) and longitudinally (Rohrer et al., 2010). Taken together, the *GRN* literature suggests that longitudinal measures of the whole brain, caudate, hippocampus and thalamus in the current cohort may prove useful for differentiating patients

from controls and provide useful biomarkers of potential disease-modifying effects.

The current project aims to investigate these previous cross-sectional findings further and to address the current lack of comprehensive longitudinal investigations in genetic FTD. The longitudinal investigations aimed to assess the utility of automated volumetric MRI measures as potential biomarkers for genetic FTD trials. To do this, direct and indirect longitudinal measures of the whole brain, amygdala, hippocampus, caudate and thalamus were applied to a large genetic cohort, including individuals in both the presymptomatic and symptomatic stages of the disease for the three main genetic causes of FTD: *MAPT*, *C9orf72* and *GRN*.

6.2 Methods

6.2.1 Participants

The genetic cohort included symptomatic and presymptomatic carriers of a mutation in either the *MAPT*, *C9orf72* or *GRN* gene and non-carrier first-degree relatives enrolled in the international GENetic Frontotemporal dementia Initiative (GENFI) study (https://www.genfi.org/). Individuals were included if they had undergone longitudinal T1-weighted MRI imaging on the same scanner, with a minimum interval of six months. Following an extensive visual assessment to ensure the quality of all images, the final resulting cohort consisted of 382 participants. The selection, QC and baseline comparisons are discussed in detail in Section 2.2.2 and summarised in Table 2-2. Scanner information and key imaging acquisition parameters for the GENFI longitudinal imaging cohort are summarised in

Table ***2-4***. To investigate potential presymptomatic changes more fully at the most proximal stage to symptom onset, the presymptomatic groups were split into two categories: early presymptomatic (>10 years prior to expected age of symptom onset) and late presymptomatic (< 10 years to the expected age of onset). EYO for each individual was calculated as the difference between age at assessment and mean age of symptom onset within their family.

6.2.2 Image analysis

6.2.2.1 Segmentation

The global whole-brain measure was derived from application of the GIF segmentation method (Cardoso et al., 2015). As discussed in detail in Section 2.3.3.3, this involves atlas propagation and label fusion to derive segmentation for the target image. This was applied successfully to all 764 images (382 baseline-repeat pairs). For the subcortical analyses, all 764 images were transformed into MNI space (along with the GIF segmentation because the amygdala was again derived from the MNI-transformed GIF segmentation as described in Section 5.2.2.2). All 764 MNI images and GIF transformations were visually inspected and completed successfully. The 1528 amygdala GIF masks and volumes (764 left and right) were then extracted for analysis. The hippocampus, caudate and thalamus were produced using the STEPS methodology with the three manually segmented template libraries in MNI space as described in Sections 5.2.2.1. All 1528 STEPS regions were successfully propagated for the hippocampus, caudate and thalamus. Including the amygdala GIF regions, this resulted in 6112 subcortical segmentations for the cohort. The total intracranial volume (TIV) for the baseline comparisons was derived from the SPM12 standard pipeline as described in Malone et al. (2015). All subcortical and whole-brain volumes were extracted from the corresponding segmentations and used to calculate the indirect longitudinal volume-difference measures of change.

6.2.2.2 Registration

To calculate the direct global and subcortical measures of change, the BSI was applied to the paired images using the appropriate masks to calculate the XOR over which the volume change was quantified. For the whole-brain rate of change, an affine registration was applied to the 382 image pairs in native space and the GBSI (single window) used as described in Section 2.3.4.1. The subcortical MNI image pairs and regions underwent an affine registration followed by an addition 6 dof local registration with application of the double-window BSI (Hobbs et al., 2009) as described in Section 5.2.2.2 and Figure 5-2. This allows for calculating the boundary shift across both the CSF–GM and GM–WM borders. This was applied separately for the left and right subcortical regions resulting in 3056 local subcortical BSI registrations.[1]

[1] Because of an unforeseen and unavoidable migration of the UCL computer science clusters used to run the computationally demanding image analysis for this thesis (and subsequent loss of access to vital packages, pipelines and scripts), I was unable to complete running the PieBSI for the GENFI cohort. Therefore, the investigations of lobar rates of change in genetic FTD will form part of future work and leaving the global whole-brain and subcortical analyses presented in this chapter.

6.2.2.3 Quality control

All 382 GBSI whole-brain registration pairs were visually assessed. Final numbers of passing longitudinal pairs are listed in the observations (Obs) column in Table 6-3. Before the STEPS segmentations were run, all 1528 MNI-transformed GIF amygdalar masks were loaded with their accompanying 764 MNI images to ensure both the MR image and left and right amygdalar mask realignment had completed successfully. As described in detail in the previous chapter (Section 5.2.2.3), the following range checking was performed for the subcortical QC. All segmentations and registrations with an annual negative BSI or GIF/STEPS volume change (i.e. growth) of 10% or above were excluded as biologically implausible. This removed 34 of the full set of 6112 subcortical longitudinal results (3056 BSI and 3056 volume-difference measures). The remaining datasets for each of the eight subcortical BSI and eight GIF/STEPS volume measures were then modelled separately and all values falling outside a 1.5 upper and lower IQR were flagged for review.

This was also done for image pairs where the absolute difference between the indirect GIF/STEPS volume measure and the equivalent BSI measure for that individual fell outside the allowed IQR values indicating indirect and direct measures for the same individual and region were substantially different. An additional random selection of 30 image pairs (~10%) for each of the 16 subcortical longitudinal measures was also reviewed. In this process, 842 image pairs were flagged based on these criteria for review. As with the subcortical reviews in the previous chapter, this list of 842 image pairs was automatically loaded allowing for review and QC blinded to scan pair and diagnosis. The local registered image pairs with region overlays were used for the review, with additional checks of the unregistered regions if the segmentation quality was questionable.

6.2.3 Statistics

Linear regression analyses, correcting for age, gender and scanner type were performed to investigate the following pairwise comparisons. For the baseline analysis, the baseline regions (volume expressed as a % of TIV) for each of the three presymptomatic (early, late and combined) and symptomatic groups were compared with non-carrier controls. This was run for the *MAPT*, *C9orf72* and *GRN* groups separately. The same regression analyses were performed for the 18 whole-brain and subcortical longitudinal measures (expressed as an annual percentage change from baseline volume) for each of the 12 genetic subgroups. These 18 longitudinal measures are summarised in Table 6-3. Longitudinal rates of change were used to calculate effect sizes and associated sample sizes with BCa CIs using the bootstrap methodology discussed in detail in Section 3.3. For measures where the mean annual rate of change was faster in the control group than the FTD subgroup the sample size estimate was not run, as it would be calculating the sample size required to detect a 30% increase in the atrophy rate of the FTD subgroup in these instances. The values for these particular measures are omitted in Table 6-6 and tended to be for right subcortical indirect measures of change in the early presymptomatic subgroups. All analyses were performed in STATA 14.

6.3 Results

6.3.1 Cross-sectional baseline comparisons

The mean (sd) baseline whole-brain and subcortical volumes are summarised in Table 6-1. Volumes are expressed as a percentage of TIV to correct for differences in head size and because it provides a good proxy of maximum pre-morbid brain volume in neurodegenerative patients. Linear regression analyses of these baseline volumes between each genetic subgroup and non-carrier controls are summarised in Table 6-2. The percentage difference reported was calculated by subtracting the mean baseline volume (% TIV) in the FTD subgroup from the mean control volume (% TIV) for that region and expressed as a percentage of the control baseline volume. Negative percentage difference results demonstrate a larger baseline volume in the FTD subgroup and a positive % difference indicates a reduced volume in the FTD subgroup relative to controls.

For the *MAPT* comparisons, none of the baseline volumes were significantly different in any of the presymptomatic subgroups compared with non-carrier controls. In the symptomatic *MAPT* group, the whole-brain measure and all subcortical regions except the left and right caudate were significantly smaller than controls $p < 0.05$. This was particularly evident for the amygdala and hippocampus bilaterally (all $p < 0.001$) ranging from 31.5% to 35.2% difference compared with controls. The left thalamus was also particularly atrophied at baseline ($p < 0.001$) with a 13.9% smaller volume compared with non-carrier controls.

Early volume reductions were found in the *C9orf72* carriers. In the early presymptomatic comparisons, many of the subcortical regions demonstrated smaller volumes. This was particularly evident for the amygdala bilaterally, with a 5.0% and 4.4% difference (left and right; $p < 0.001$) and the thalamus bilaterally, with a 4.9% and 3.7% reduction (left and right; $p < 0.001$) when compared with non-carrier controls. Hippocampal volumes were also bilaterally smaller, with a 5.0% and 4.5% difference from

controls (both $p < 0.002$). In the late presymptomatic group, the pattern somewhat overlapped with that found in the early presymptomatic, albeit at the $p < 0.05$ level of significance most likely due to a substantially smaller cohort size (n = 15 compared with n = 35). Despite a reasonable 3.2% and 3.8% volumetric difference for the left and right amygdala, this result was no longer significant. It is possible that the smaller cohort size in the late presymptomatic group compared with the early presymptomatic group contributed to this non-significant result. Interestingly, smaller thalami bilaterally were now evident in this later presymptomatic stage, with a 6.2% and 6.8% smaller volume for the left and right, respectively, than non-carrier controls ($p < 0.05$). Given the larger number of early presymptomatic individuals contributing to the overall presymptomatic group, the combined presymptomatic *C9orf72* group results reflected the earlier pattern. In the symptomatic *C9orf72* group, there was evidence of widespread smaller subcortical regions at baseline (all $p < 0.001$, except for the right caudate at $p < 0.004$). Evidence of whole-brain volume loss at baseline was also found with a 9.5% reduction compared with controls ($p < 0.001$).

In *GRN* carriers, both the combined presymptomatic group and the early presymptomatic group did not show any significant lower volumes than non-carrier controls. The late presymptomatic group showed smaller thalamic volumes than the controls (left: 6.2%, $p = 0.049$; right: 6.3%, $p = 0.012$). The symptomatic *GRN* group presented with significantly smaller whole-brain volumes (7.0%; $p < 0.001$) and the characteristic asymmetric pattern of reduced volumes across all subcortical measures. The current cohort has a predominantly left-sided pattern of reduced subcortical volumes (all $p < 0.001$, except the amygdala with $p < 0.002$), with volumes ranging from 8.8% (left amygdala) to 15.3% (left caudate) smaller than non-carrier controls. The right hippocampus and right caudate also demonstrated 10.4% ($p = 0.005$) and 5.3% ($p = 0.033$) smaller regions, respectively.

Table 6-1 Mean (sd) baseline volumes for the whole-brain, amygdala, hippocampus, caudate and thalamus expressed as a percentage of TIV*

Measure	Non-carrier controls (n=159)	*MAPT*			
		Presymptomatic early (>10 EYO) (n=17)	Presymptomatic late (<10 EYO) (n=17)	Presymptomatic (all) (n=34)	Symptomatic (n=11)
	Mean (sd)	Mean (sd)	Mean (sd)	Mean (sd)	Mean (sd)
Whole-brain (as % of TIV)	80.17 (2.90)	81.84 (2.22)	80.00 (2.11)	80.84 (2.33)	73.7 (3.09)
Amygdala_GIF_Left (as % of TIV)	0.12 (0.01)	0.13 (0.01)	0.12 (0.01)	0.12 (0.01)	0.08 (0.02)
Amygdala_GIF_Right (as % of TIV)	0.12 (0.01)	0.12 (0.01)	0.12 (0.02)	0.12 (0.01)	0.08 (0.02)
Hippocampus_STEPS_Left (as % of TIV)	0.22 (0.02)	0.22 (0.02)	0.22 (0.02)	0.22 (0.03)	0.14 (0.04)
Hippocampus_STEPS_Right (as % of TIV)	0.23 (0.02)	0.23 (0.02)	0.22 (0.03)	0.22 (0.03)	0.15 (0.04)
Caudate_STEPS_Left (as % of TIV)	0.25 (0.03)	0.25 (0.03)	0.25 (0.03)	0.25 (0.03)	0.23 (0.04)
Caudate_STEPS_Right (as % of TIV)	0.24 (0.03)	0.24 (0.04)	0.25 (0.03)	0.24 (0.03)	0.23 (0.03)
Thalamus_STEPS_Left (as % of TIV)	0.46 (0.04)	0.49 (0.04)	0.47 (0.04)	0.48 (0.05)	0.4 (0.04)
Thalamus_STEPS_Right (as % of TIV)	0.47 (0.04)	0.50 (0.05)	0.47 (0.04)	0.48 (0.05)	0.43 (0.04)

* TIV = Total intercranial volume derived from SPM12

Table 6-1 continued: Mean (sd) baseline volumes for the whole-brain, amygdala, hippocampus, caudate and thalamus expressed as a percentage of TIV**

Measure	*C9orf72*			
	Presymptomatic early (>10 EYO) (n=35)	**Presymptomatic late (<10 EYO) (n=15)**	**Presymptomatic (all) (n=50)**	**Symptomatic (n=25)**
	Mean (sd)	Mean (sd)	Mean (sd)	Mean (sd)
Whole-brain (as % of TIV)	80.16 (2.85)	77.42 (3.62)	79.34 (3.31)	72.57 (2.75)
Amygdala_GIF_Left (as % of TIV)	0.12 (0.01)	0.12 (0.00)	0.12 (0.01)	0.10 (0.02)
Amygdala_GIF_Right (as % of TIV)	0.11 (0.01)	0.12 (0.01)	0.11 (0.01)	0.10 (0.01)
Hippocampus_STEPS_Left (as % of TIV)	0.21 (0.02)	0.21 (0.02)	0.21 (0.02)	0.17 (0.03)
Hippocampus_STEPS_Right (as % of TIV)	0.22 (0.02)	0.21 (0.02)	0.22 (0.02)	0.19 (0.03)
Caudate_STEPS_Left (as % of TIV)	0.25 (0.02)	0.23 (0.04)	0.24 (0.03)	0.20 (0.03)
Caudate_STEPS_Right (as % of TIV)	0.24 (0.03)	0.23 (0.03)	0.24 (0.03)	0.22 (0.03)
Thalamus_STEPS_Left (as % of TIV)	0.44 (0.04)	0.43 (0.05)	0.44 (0.04)	0.36 (0.04)
Thalamus_STEPS_Right (as % of TIV)	0.45 (0.04)	0.44 (0.04)	0.44 (0.04)	0.38 (0.04)
Measure	***GRN***			
	Presymptomatic early (>10 EYO) (n=40)	**Presymptomatic late (<10 EYO) (n=41)**	**Presymptomatic (all) (n=81)**	**Symptomatic (n=22)**
Whole-brain (as % of TIV)	81.63 (2.59)	78.51 (2.70)	80.05 (3.06)	74.56 (4.28)
Amygdala_GIF_Left (as % of TIV)	0.13 (0.01)	0.12 (0.01)	0.12 (0.01)	0.11 (0.01)
Amygdala_GIF_Right (as % of TIV)	0.12 (0.01)	0.12 (0.01)	0.12 (0.01)	0.11 (0.01)
Hippocampus_STEPS_Left (as % of TIV)	0.23 (0.02)	0.21 (0.03)	0.22 (0.03)	0.19 (0.04)
Hippocampus_STEPS_Right (as % of TIV)	0.23 (0.02)	0.22 (0.03)	0.22 (0.03)	0.20 (0.04)
Caudate_STEPS_Left (as % of TIV)	0.25 (0.02)	0.25 (0.03)	0.25 (0.03)	0.21 (0.05)
Caudate_STEPS_Right (as % of TIV)	0.24 (0.03)	0.25 (0.04)	0.24 (0.04)	0.23 (0.04)
Thalamus_STEPS_Left (as % of TIV)	0.49 (0.03)	0.44 (0.04)	0.46 (0.05)	0.40 (0.05)
Thalamus_STEPS_Right (as % of TIV)	0.49 (0.04)	0.44 (0.04)	0.46 (0.05)	0.44 (0.04)

Table 6-2 Linear regression results comparing baseline whole-brain and subcortical volumes between genetic subgroups and controls (non-carriers). Negative% difference indicates volumes are larger in the genetic subgroups and positive% difference indicates volumes are smaller than non-carrier controls.

Baseline volumes (as % of TIV)	***MAPT***											
	Presymptomatic early (>10 EYO)			**Presymptomatic late (<10 EYO)**			**Presymptomatic (all)**			**Symptomatic**		
	***p*-value**	**95% CI**	**% difference**	***p*-value**	**95% CI**	**% difference**	***p*-value**	**95% CI**	**% difference**	***p*-value**	**95% CI**	**% difference**
Whole-brain GIF	0.779	[-1.435 – 1.076]	-2.1	0.770	[-1.397 – 1.036]	0.2	0.623	[-1.114 – 0.669]	-0.8	**<0.001**	[-6.405 – -3.492]	8.1
Amygdala_GIF_Left	0.748	[-0.005 – 0.006]	-2.3	0.345	[-0.008 – 0.003]	2.1	0.294	[-0.006 – 0.002]	0.8	**<0.001**	[-0.048 – -0.034]	34.6
Amygdala_GIF_Right	0.913	[-0.005 – 0.005]	-1.5	0.090	[-0.01 – 0.001]	3.6	0.071	[-0.007 – 0.000]	1.9	**<0.001**	[-0.042 – -0.030]	31.5
Hippocampus_STEPS_Left	0.240	[-0.02 – 0.005]	1.0	0.915	[-0.013 – 0.012]	0.2	0.153	[-0.016 – 0.003]	1.8	**<0.001**	[-0.089 – -0.058]	35.1
Hippocampus_STEPS_Right	0.490	[-0.017 – 0.008]	-0.9	0.464	[-0.017 – 0.008]	1.9	0.111	[-0.017 – 0.002]	1.7	**<0.001**	[-0.090 – -0.060]	35.2
Caudate_STEPS_Left	0.578	[-0.019 – 0.011]	0.3	0.622	[-0.011 – 0.019]	-1.6	0.752	[-0.013 – 0.009]	-0.1	0.093	[-0.034 – 0.003]	7.6
Caudate_STEPS_Right	0.463	[-0.023 – 0.010]	2.8	0.648	[-0.013 – 0.02]	-1.5	0.656	[-0.015 – 0.009]	1.3	0.386	[-0.029 – 0.011]	3.5
Thalamus_STEPS_Left	0.322	[-0.009 – 0.027]	-6.5	0.753	[-0.015 – 0.021]	-0.9	0.667	[-0.011 – 0.016]	-3.2	**<0.001**	[-0.069 – -0.025]	13.9
Thalamus_STEPS_Right	0.087	[-0.002 – 0.035]	-7.1	0.455	[-0.012 – 0.026]	-1.8	0.265	[-0.006 – 0.022]	-3.8	*0.030*	[-0.048 – -0.002]	8.5

Table 6-2 continued: Linear regression results comparing baseline whole-brain and subcortical volumes between genetic subgroups and controls (non-carriers). Negative% difference indicates volumes are larger in the genetic subgroups and positive% difference indicates volumes are smaller than non-carrier controls.

Baseline volumes (as % of TIV)	*C9orf72*											
	Presymptomatic early (>10 EYO)			Presymptomatic late (<10 EYO)			Presymptomatic (all)			Symptomatic		
	p-value	95% CI	% difference	*p*-value	95% CI	% difference	*p*-value	95% CI	% difference	*p*-value	95% CI	% difference
Whole-brain GIF	0.505	[-3.694 – 2.822]	0.0	0.544	[-5.159 – 2.725]	3.4	**<0.001**	[-2.212 – -0.681]	1.0	**<0.001**	[-5.485 – -3.352]	9.5
Amygdala_GIF_Left	**<0.001**	[-0.012 – -0.004]	5.0	0.393	[-0.008 – 0.003]	3.2	*0.001*	[-0.010 – -0.003]	4.5	**<0.001**	[-0.022 – -0.012]	17.0
Amygdala_GIF_Right	**<0.001**	[-0.011 – -0.003]	4.4	0.251	[-0.008 – 0.002]	3.8	**<0.001**	[-0.009 – -0.003]	4.2	**<0.001**	[-0.017 – -0.008]	13.4
Hippocampus_STEPS_Left	*0.002*	[-0.024 – -0.005]	5.0	*0.038*	[-0.027 – -0.001]	6.8	**<0.001**	[-0.022 – -0.007]	5.5	**<0.001**	[-0.050 – -0.027]	22.1
Hippocampus_STEPS_Right	*0.002*	[-0.023 – -0.005]	4.5	*0.036*	[-0.027 – -0.001]	6.9	**<0.001**	[-0.022 – -0.007]	5.2	**<0.001**	[-0.038 – -0.016]	17.4
Caudate_STEPS_Left	0.762	[-0.009 – 0.013]	-1.5	*0.043*	[-0.032 – -0.001]	6.2	0.424	[-0.013 – 0.006]	0.9	**<0.001**	[-0.050 – -0.023]	18.2
Caudate_STEPS_Right	0.659	[-0.009 – 0.015]	-0.5	*0.021*	[-0.038 – -0.003]	6.8	0.411	[-0.015 – 0.006]	1.7	0.004	[-0.036 – -0.007]	9.3
Thalamus_STEPS_Left	**<0.001**	[-0.051 – -0.025]	4.9	*0.017*	[-0.042 – -0.004]	7.3	**<0.001**	[-0.046 – -0.023]	5.6	**<0.001**	[-0.085 – -0.052]	23.1
Thalamus_STEPS_Right	**<0.001**	[-0.043 – -0.015]	3.7	*0.025*	[-0.042 – -0.003]	6.4	**<0.001**	[-0.039 – -0.016]	4.5	**<0.001**	[-0.067 – -0.033]	17.7

* All analyses adjusted for age, gender and scanner type: **$p < 0.001$ bolded**, $p < 0.05$ italicised

Table 6-2 continued: Linear regression results comparing baseline whole-brain and subcortical volumes between genetic subgroups and controls (non-carriers). Negative% difference indicates volumes are larger in the genetic subgroups and positive% difference indicates volumes are smaller than non-carrier controls.

Baseline volumes (as % of TIV)	*GRN*											
	Presymptomatic early (>10 EYO)			Presymptomatic late (<10 EYO)			Presymptomatic (all)			Symptomatic		
	p-value	95% CI	% difference	*p*-value	95% CI	% difference	*p*-value	95% CI	% difference	*p*-value	95% CI	% difference
Whole-brain GIF	0.688	[-2.370 – 2.912]	-1.8	0.767	[-2.216 – 3.004]	2.1	0.996	[-0.643 – 0.638]	0.1	**<0.001**	[-4.932 – -2.662]	7.0
Amygdala_GIF_Left	0.391	[-0.002 – 0.006]	-2.7	0.085	[-0.007 – 0.000]	4.1	0.554	[-0.004 – 0.002]	0.7	0.002	[-0.013 – -0.002]	8.8
Amygdala_GIF_Right	0.782	[-0.003 – 0.004]	-1.9	0.125	[-0.006 – 0.001]	3.7	0.398	[-0.004 – 0.002]	1.0	0.284	[-0.007 – 0.002]	4.5
Hippocampus_STEPS_Left	0.887	[-0.008 – 0.009]	-2.3	0.291	[-0.013 – 0.004]	3.7	0.538	[-0.009 – 0.005]	0.7	**<0.001**	[-0.038 – -0.015]	14.5
Hippocampus_STEPS_Right	0.884	[-0.009 – 0.008]	-2.1	0.079	[-0.016 – 0.001]	5.3	0.204	[-0.011 – 0.002]	1.7	0.005	[-0.028 – -0.005]	10.4
Caudate_STEPS_Left	0.651	[-0.008 – 0.013]	-2.2	0.474	[-0.007 – 0.014]	-0.8	0.434	[-0.005 – 0.011]	-1.5	**<0.001**	[-0.049 – -0.022]	15.3
Caudate_STEPS_Right	0.739	[-0.014 – 0.010]	1.1	0.816	[-0.010 – 0.013]	-1.3	0.953	[-0.009 – 0.008]	-0.1	0.033	[-0.031 – -0.001]	5.3
Thalamus_STEPS_Left	0.286	[-0.006 – 0.020]	-5.4	*0.049*	[-0.025 – 0.000]	6.2	0.535	[-0.013 – 0.007]	0.5	**<0.001**	[-0.054 – -0.021]	13.9
Thalamus_STEPS_Right	0.360	[-0.007 – 0.019]	-4.3	*0.012*	[-0.03 – -0.004]	6.3	0.275	[-0.015 – 0.004]	1.0	0.396	[-0.025 – 0.010]	6.3

* All analyses adjusted for age, gender and scanner type: **p *< 0.001* bolded**, p < 0.05 italicised

6.3.1 Longitudinal analysis

6.3.1.1 Rates of change and regression results

Mean (sd) annual rates of whole-brain and subcortical change are summarised in Table 6-3. Figure 6-1 shows these atrophy rates with 95% CIs produced by the 18 longitudinal measures for the non-carrier controls and each of the genetic FTD subgroup separately. Note the differences in the x-axis for the subgroups. All presymptomatic (early, late and combined) have an upper limit on the x-axis of 5% of baseline volume loss per year. The symptomatic *MAPT* and *C9orf72* graphs have an upper x-axis limit of 10% annual atrophy, whereas the *GRN* symptomatic graph required an extension to 20% annual volume loss due to the high rates of caudate atrophy in this group.

Table 6-4 presents the results from the linear regression analysis with 95% BCa CIs comparing annual rates of whole-brain and subcortical change for each genetic subgroup with non-carrier controls. The regression coefficient (Coef.) representing the adjusted mean difference in the annual rate of change for each measure and genetic subgroup relative to the control rate when correcting for age, gender and scanner is also included. All comparisons with a *p*-value of <0.001 are in bold and those with a statistically different annual rate of change compared with non-carrier controls at the $p < 0.05$ level are italicised.

Non-carrier control rates (blue in Figure 6-1) were relatively homogeneous across all direct BSI and indirect GIF/STEPS volume-difference measures of whole-brain and subcortical atrophy, demonstrating no evidence of volume change over a year. This ranged from 0.0 (2.0)% for the right thalamus indirect STEPS measure of change to 0.5 (2.6)% for the left amygdala GIF volume-difference measure.

The *MAPT* early presymptomatic group showed similar results to the non-carrier controls with the whole-brain and subcortical measures indicating minimal annual volume change. None of the comparisons reached statistical significance in the regression analyses. In the late presymptomatic *MAPT* group, the highest annual rates of change were derived from the left and right hippocampal STEPS volume difference at 1.4 (2.6)% and 1.1 (2.6)%, respectively. Atrophy rates derived using this indirect left hippocampal STEPS measure were significantly higher ($p = 0.015$) with a mean adjusted annual difference of 1.3% faster annual hippocampal atrophy than non-carriers once correcting for age, gender and scanner type in this late presymptomatic group. In the combined presymptomatic group, the highest adjusted mean difference rates were derived from the indirect volume-difference measures for the right caudate (1.4%; $p = 0.045$) and left hippocampus (1.1%; $p = 0.004$). The left ABSI also provided annual rates of change that were significantly higher than controls ($p = 0.013$), with an adjusted mean difference of a 0.9%. The symptomatic *MAPT* patients showed substantially higher mean-to-sd ratios using the raw atrophy rates and significantly higher rates of change for all measures than the non-carrier controls ($p < 0.05$), with the exception of the indirect GIF left amygdala measure ($p = 0.187$). Application of the BSI reduced variability improved the longitudinal signal with the majority of measures demonstrating significantly higher rates of change at the $p < 0.001$ significance level (Table 6-4).

In the *C9orf72* early presymptomatic groups, annual rates of change were overall close to zero with high standard deviations relative to the mean. The left and right HBSI had the best (highest) mean-to-sd ratio in the subgroup based on the raw mean rates of change data, with the left HBSI producing significantly higher annual rates of change than non-carrier controls ($p = 0.012$). In the late *C9orf72* presymptomatic group, the majority of measures had higher rates of annual change and improved mean-to-sd ratios. The regression results reveal that the right STEPS thalamic measure ($p = 0.003$), left TBSI ($p = 0.019$) and right TBSI ($p = 0.050$) all demonstrated significantly higher annual rates of change than non-carrier controls. When combining

the late presymptomatic group with the larger early presymptomatic group, the full presymptomatic mean rates of change were lowered, along with the mean-to-sd ratios across all measures except the HBSI. Therefore, the left HBSI was the only measure to remain significant in this combined group (p = 0.030). The symptomatic *C9orf72* group demonstrated high annual rates of change and mean-to-sd ratios, particularly for the left and right CBSI. Both whole-brain measures and the majority of subcortical longitudinal results demonstrated significantly higher rates of change than non-carrier controls.

Finally, mean rates of change in the early presymptomatic *GRN* subgroup failed to demonstrate annual volume changes for any of the 18 longitudinal measures. All were close to zero with low mean-to-sd ratios and no significant differences from non-carrier controls. The late presymptomatic group had comparatively higher annual rates of change, particularly for the CBSI and caudate STEPS measures bilaterally. Standard deviations relative to the mean also improved but were still close to or under 0.5 for multiple measures. Despite this, the left HBSI, bilateral CBSI and right TBSI measures demonstrated significantly increased rates of change in these presymptomatic carriers compared with non-carrier controls at $p < 0.05$. The symptomatic *GRN* cohort showed high rates of annual change across all measures, with the characteristically asymmetric predominance of volume loss. The left CBSI had the highest rate of change with 17.7 (12.1)% annual atrophy and the whole-brain GBSI had the highest mean-to-sd ratio of 2.3. Sixteen of the 18 longitudinal whole-brain and subcortical measures demonstrated higher rates of annual atrophy at the $p < 0.001$ level of significance, with the exception of the indirect GIF volume-difference measures of amygdala atrophy (p = 0.058 and p = 0.384; left and right).

Table 6-3 Mean (sd) annual rates of whole-brain and subcortical change (BSI and volume-difference measures), expressed as an annual percentage change from baseline volume for non-carrier controls and genetic FTD subgroups

Measure	Non-carrier controls (n = 159)		*MAPT*							
			Early Presymptomatic (>10 EYO) (n = 17)		Late Presymptomatic (<10 EYO) (n = 17)		Presymptomatic (all) (n = 34)		Symptomatic (n = 11)	
BSI	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**
Whole-brain GBSI	125	0.2 (0.5)	14	0.4 (0.4)	14	0.2 (0.2)	28	0.3 (0.4)	10	0.9 (0.7)
ABSI_Left	153	0.1 (1.8)	17	0.4 (1.1)	16	0.8 (2.4)	33	0.9 (2.4)	9	1.8 (2.8)
ABSI_Right	156	0.3 (1.5)	16	0.6 (1.7)	15	0.1 (2.2)	31	0.6 (2.3)	10	2.6 (2.7)
HBSI_Left	155	0.0 (1.4)	17	0.0 (0.7)	16	0.5 (1.2)	33	0.5 (1.7)	9	1.2 (2.0)
HBSI_Right	154	0.2 (1.2)	17	0.3 (1.1)	15	0.4 (1.3)	32	0.5 (1.7)	11	1.2 (2.4)
CBSI_Left	155	0.3 (1.0)	17	0.3 (0.7)	16	0.8 (1.3)	33	0.8 (1.7)	11	5.9 (4.9)
CBSI_Right	158	0.3 (1.1)	17	0.5 (0.8)	16	0.4 (0.9)	33	0.7 (1.5)	11	4.6 (4.6)
TBSI_Left	156	0.2 (0.6)	17	0.0 (0.3)	16	0.4 (0.7)	33	0.2 (0.6)	11	1.1 (1.3)
TBSI_Right	156	0.1 (0.6)	17	0.1 (0.9)	16	0.3 (0.6)	33	0.2 (0.8)	11	1.0 (1.2)
STEPS_GIF_Volume difference										
Whole-brain GIF	158	0.1 (0.6)	16	0.1 (0.7)	16	0.3 (0.5)	32	0.3 (0.7)	11	1.2 (0.8)
Amygdala_GIF_Left	152	0.5 (2.6)	14	0.9 (4.1)	16	1.1 (4.1)	30	1.3 (4.3)	11	2.1 (3.0)
Amygdala_GIF_Right	152	0.1 (2.0)	15	-0.3 (2.8)	15	0.4 (2.6)	30	0.4 (3.2)	11	2.3 (4.2)
Hippocampus_STEPS_Left	156	0.2 (1.6)	17	0.3 (2.5)	16	1.4 (2.6)	33	1.1 (3.1)	10	1.6 (2.2)
Hippocampus_STEPS_Right	155	0.3 (1.6)	16	0.1 (2.1)	16	1.1 (2.6)	32	0.8 (2.6)	11	2.0 (2.1)
Caudate_STEPS_Left	151	0.4 (2.3)	14	-0.7 (2.1)	15	0.6 (1.3)	29	0.3 (2.3)	11	6.3 (8.1)
Caudate_STEPS_Right	147	0.2 (2.5)	13	1.9 (4.4)	15	1.1 (2.1)	28	1.6 (3.4)	11	3.6 (5.7)
Thalamus_STEPS_Left	155	0.2 (1.7)	17	-0.4 (1.6)	15	0.4 (1.2)	32	0.0 (1.5)	10	2.0 (2.2)
Thalamus_STEPS_Right	154	0.0 (2.0)	16	-0.2 (1.7)	15	0.7 (1.4)	31	0.2 (1.6)	11	1.8 (2.1)

Table 6-3 continued: Mean (sd) annual rates of whole-brain and subcortical change (BSI and volume-difference measures), expressed as an annual percentage change from baseline volume for non-carrier controls and genetic FTD subgroups

Measure	*C9orf72*							
	Early Presymptomatic (>10 EYO) (n = 35)		Late Presymptomatic (<10 EYO) (n = 15)		Presymptomatic (all) (n = 50)		Symptomatic (n = 25)	
BSI	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**
Whole-brain GBSI	30	0.2 (0.7)	15	0.5 (0.6)	45	0.3 (0.7)	24	0.8 (0.8)
ABSI_Left	33	0.2 (1.6)	15	0.8 (1.3)	48	0.4 (1.5)	23	1.0 (2.6)
ABSI_Right	33	0.2 (1.4)	15	0.9 (1.3)	48	0.4 (1.4)	24	1.6 (2.7)
HBSI_Left	34	0.6 (1.7)	15	0.4 (1.6)	49	0.5 (1.7)	24	1.1 (2.4)
HBSI_Right	34	0.6 (1.4)	15	0.4 (1.9)	49	0.5 (1.5)	22	0.8 (1.6)
CBSI_Left	35	0.2 (1.0)	15	1.9 (2.7)	50	0.7 (1.9)	23	6.2 (4.1)
CBSI_Right	35	0.2 (0.9)	15	1.1 (1.9)	50	0.5 (1.3)	25	4.4 (3.4)
TBSI_Left	34	0.1 (0.6)	14	0.8 (0.8)	48	0.3 (0.7)	25	1.3 (1.4)
TBSI_Right	32	0.2 (0.6)	15	0.6 (0.8)	47	0.3 (0.7)	24	0.7 (1.1)
STEPS_GIF_Volume difference								
Whole-brain GIF	35	0.2 (0.7)	15	0.4 (0.7)	50	0.3 (0.7)	25	1.1 (0.8)
Amygdala_GIF_Left	33	-0.3 (2.0)	15	0.7 (2.2)	48	0 (2.1)	23	2.1 (2.7)
Amygdala_GIF_Right	33	-0.4 (1.9)	15	-0.8 (1.7)	48	-0.5 (1.8)	23	1.3 (3.0)
Hippocampus_STEPS_Left	33	0.2 (1.5)	15	0.4 (1.5)	48	0.3 (1.5)	24	1.7 (2.7)
Hippocampus_STEPS_Right	35	-0.1 (1.8)	14	0.0 (2.3)	49	-0.1 (1.9)	24	2.1 (2.1)
Caudate_STEPS_Left	34	0.7 (2.5)	14	1.0 (1.9)	48	0.8 (2.3)	21	2.4 (3.5)
Caudate_STEPS_Right	33	-1.0 (2.8)	15	0.0 (3.1)	48	-0.7 (2.3)	23	2.3 (3.8)
Thalamus_STEPS_Left	35	0.1 (1.6)	15	0.5 (2.0)	50	0.2 (1.7)	23	2.2 (2.5)
Thalamus_STEPS_Right	34	-0.1 (1.9)	14	1.6 (2.3)	48	0.4 (2.2)	23	1.3 (2.1)

Table 6-3 continued: Mean (sd) annual rates of whole-brain and subcortical change (BSI and volume-difference measures), expressed as an annual percentage change from baseline volume for non-carrier controls and genetic FTD subgroups

Measure	*GRN*							
	Early Presymptomatic (>10 EYO) (n = 40)		Late Presymptomatic late (<10 EYO) (n = 41)		Presymptomatic (all) (n = 81)		Symptomatic (n = 22)	
BSI	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**	**Obs**	**Mean (sd)**
Whole-brain GBSI	31	0.2 (0.5)	36	0.4 (0.7)	67	0.3 (0.6)	21	1.6 (0.7)
ABSI_Left	38	0.4 (1.8)	39	0.9 (2.0)	77	0.7 (1.9)	19	4.8 (3.8)
ABSI_Right	37	0.3 (1.5)	39	0.7 (1.7)	76	0.5 (1.6)	19	3.3 (3.1)
HBSI_Left	39	0.4 (1.5)	39	0.7 (1.3)	78	0.6 (1.4)	20	3.4 (2.3)
HBSI_Right	39	0.5 (1.3)	39	0.7 (1.5)	78	0.6 (1.4)	20	2.4 (2.5)
CBSI_Left	39	0.4 (0.8)	40	1.6 (3.7)	79	1.0 (2.8)	22	17.7 (12.1)
CBSI_Right	40	0.2 (0.9)	40	1.9 (4.4)	80	1.0 (3.3)	21	13.8 (9.8)
TBSI_Left	40	0.1 (0.5)	39	0.4 (0.9)	79	0.3 (0.7)	21	3.7 (2.1)
TBSI_Right	40	0.0 (0.5)	39	0.6 (0.9)	79	0.3 (0.8)	22	1.5 (1.5)
STEPS_GIF_Volume difference								
Whole-brain GIF	40	0.0 (0.8)	41	0.3 (0.6)	81	0.2 (0.7)	21	2.1 (1.5)
Amygdala_GIF_Left	39	0.6 (2.6)	38	0.6 (2.7)	77	0.6 (2.6)	22	2.2 (2.8)
Amygdala_GIF_Right	37	-0.1 (2.1)	41	0.8 (2.5)	78	0.4 (2.3)	22	0.9 (2.2)
Hippocampus_STEPS_Left	40	-0.1 (1.5)	40	0.9 (1.8)	80	0.4 (1.7)	22	3.1 (3.3)
Hippocampus_STEPS_Right	38	0.4 (1.2)	40	0.5 (1.6)	78	0.5 (1.4)	22	2.7 (4.2)
Caudate_STEPS_Left	39	0.5 (2.6)	40	1.8 (3.3)	79	1.2 (3.1)	20	12 (11.7)
Caudate_STEPS_Right	39	0.3 (2.6)	40	1.3 (3.2)	79	0.8 (3.0)	21	10.6 (8.4)
Thalamus_STEPS_Left	39	0.4 (1.1)	39	0.4 (1.4)	78	0.4 (1.3)	20	6.1 (3.9)
Thalamus_STEPS_Right	40	-0.1 (1.4)	38	0.4 (1.8)	78	0.2 (1.6)	20	2.5 (2.9)

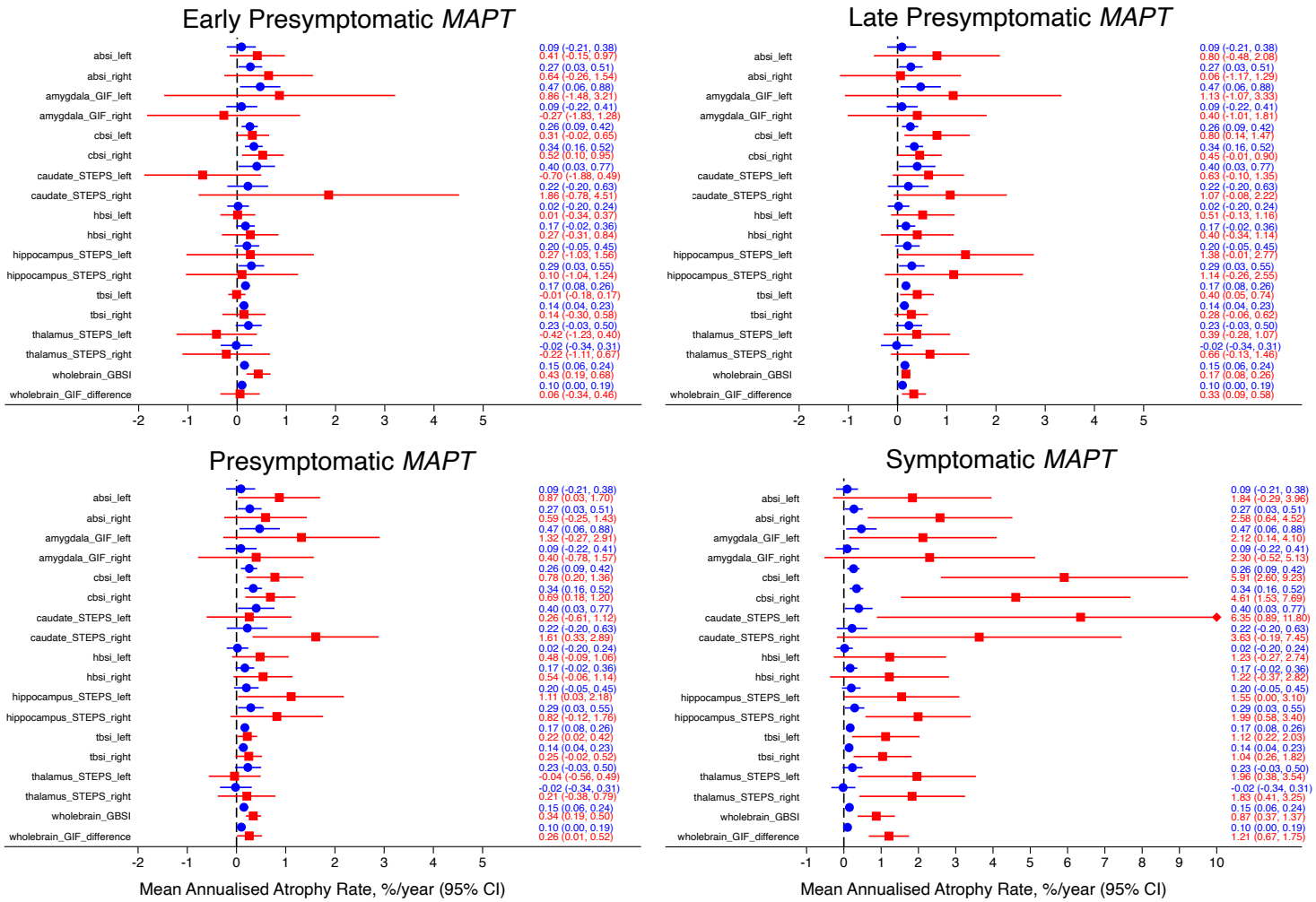

Figure 6-1 Mean whole-brain and subcortical atrophy rates with 95% confidence intervals for each genetic subgroup (red) and the non-carrier controls (blue)

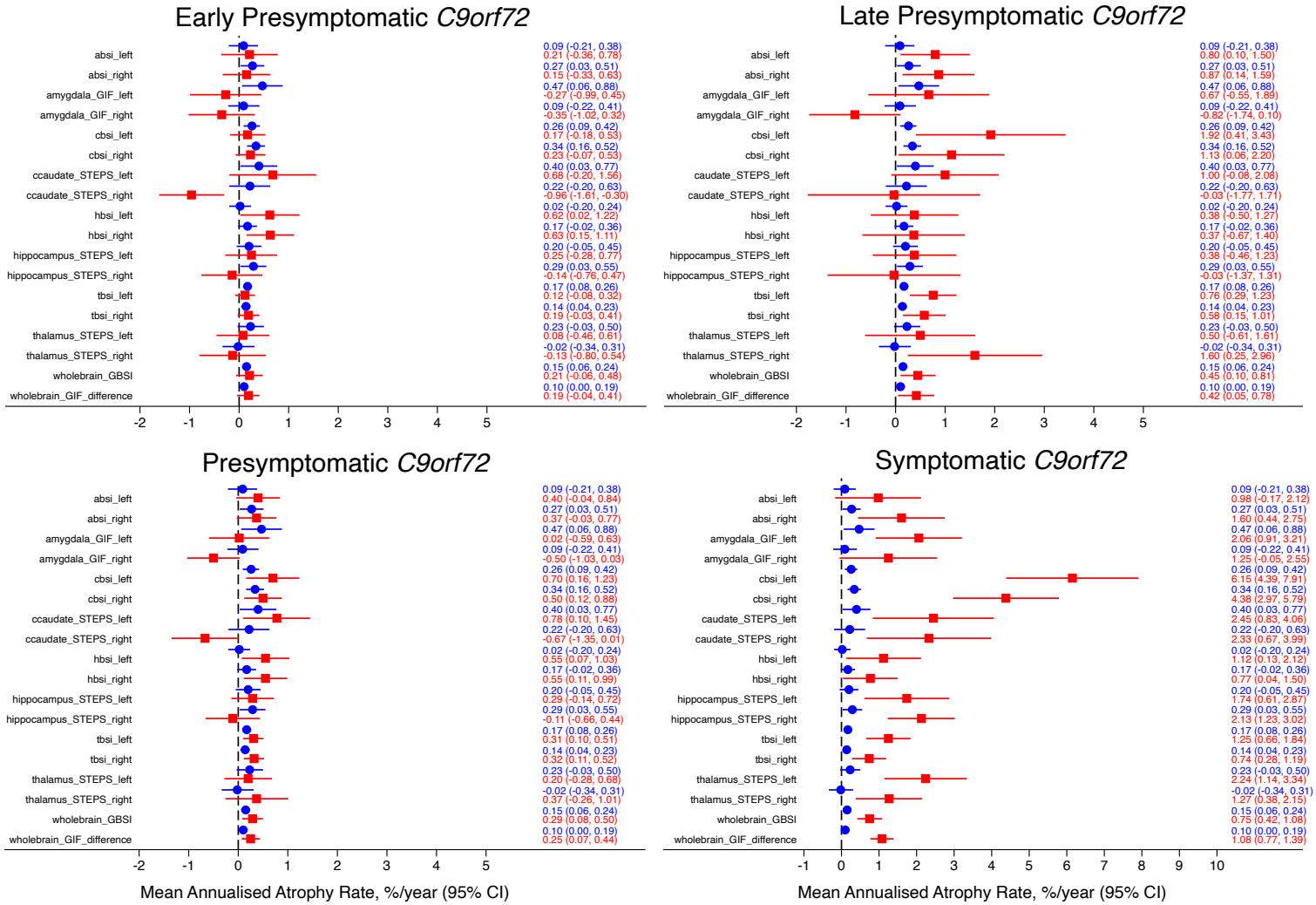

Figure 6-1 continued: Mean whole-brain and subcortical atrophy rates with 95% confidence intervals for each genetic subgroup (red) and the non-carrier controls (blue)

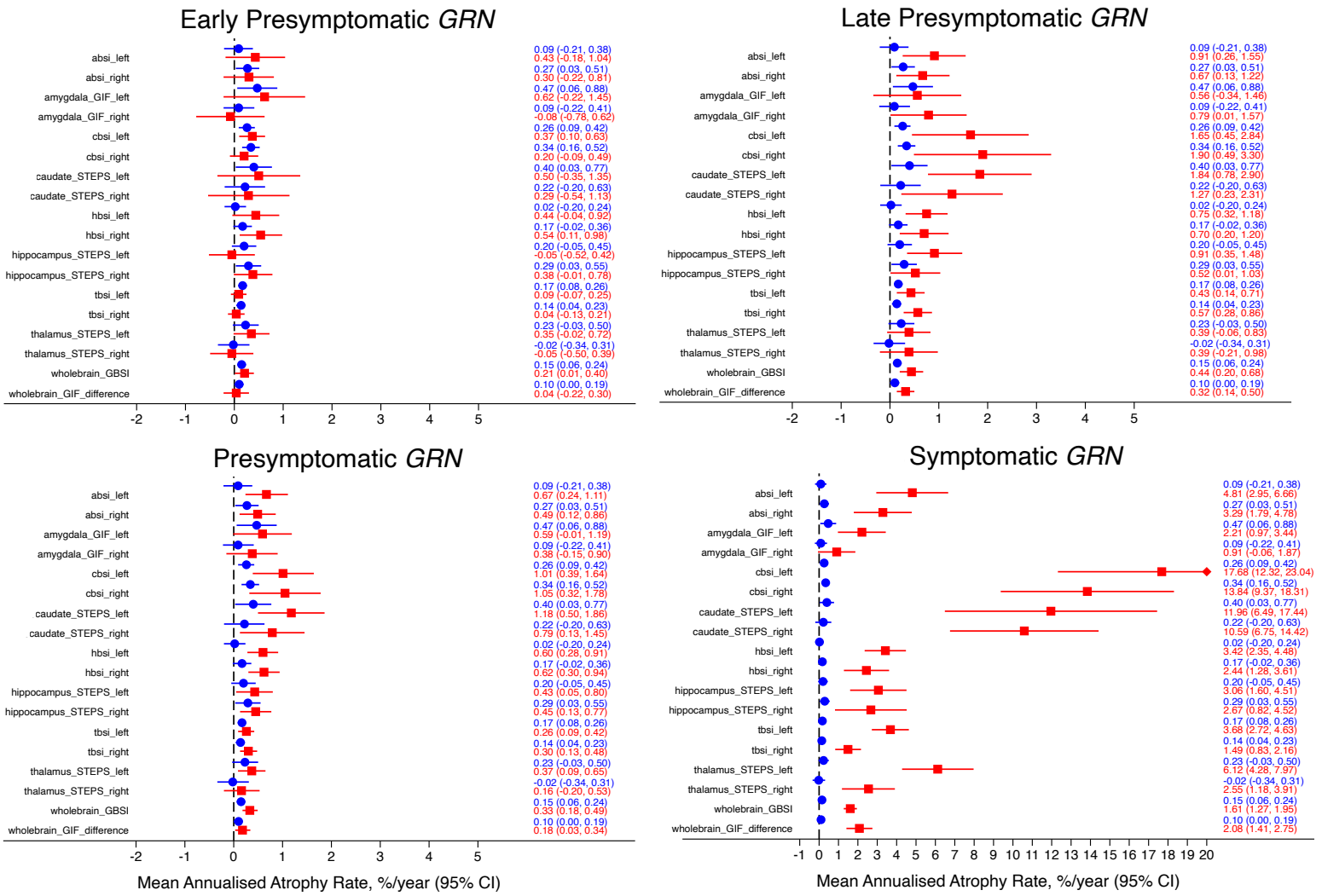

Figure 6-1 continued: Mean whole-brain and subcortical atrophy rates with 95% confidence intervals for each genetic subgroup (red) and the non-carrier controls (blue)

Table 6-4 Linear regression analysis results with adjusted mean difference and 95% CIs comparing annual rates of whole-brain and subcortical atrophy for each genetic subgroup with non-carrier controls. *

Measure	*MAPT*							
	Presymptomatic early (>10 EYO)		Presymptomatic late (<10 EYO)		Presymptomatic (all)		Symptomatic	
BSI	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**
Whole-brain GBSI	0.3 [0.0–0.6]	0.069	0.0 [-0.3–0.4]	0.815	0.2 [0.0–0.5]	0.053	0.7 [0.3–1.1]	*0.001*
ABSI_Left	0.7 [-0.4–1.8]	0.220	0.7 [-0.3–1.8]	0.173	1.0 [0.2–1.8]	*0.013*	1.7 [0.3–3.1]	*0.020*
ABSI_Right	0.7 [-0.3–1.7]	0.158	-0.1 [-1.2–0.8]	0.685	0.5 [-0.2–1.2]	0.154	2.2 [1.0–3.4]	**<0.001**
HBSI_Left	0.1 [-0.7–0.9]	0.850	0.6 [-0.2–1.4]	0.169	0.6 [0.0–1.2]	*0.030*	1.3 [0.2–2.4]	*0.007*
HBSI_Right	0.2 [-0.6–0.9]	0.693	0.2 [-0.6–1.0]	0.603	0.4 [-0.2–1.0]	0.180	1.1 [0.2–2.1]	*0.016*
CBSI_Left	0.0 [-1.8–1.9]	0.963	0.7 [-1.2–2.5]	0.472	0.7 [-0.6–2.1]	0.305	5.8 [3.6–8.1]	**<0.001**
CBSI_Right	0.5 [-1.1–2.1]	0.518	0.1 [-1.5–1.7]	0.841	0.6 [-0.5–1.8]	0.504	4.5 [2.6–6.4]	**<0.001**
TBSI_Left	-0.1 [-0.6–0.3]	0.585	0.3 [-0.2–0.7]	0.248	0.1 [-0.2–0.5]	0.445	0.9 [0.4–1.5]	**<0.001**
TBSI_Right	0.0 [-0.4–0.4]	0.872	0.2 [-0.2–0.6]	0.415	0.2 [-0.1–0.5]	0.251	0.9 [0.4–1.4]	**<0.001**
STEPS_GIF_Volume Difference								
Whole-brain GIF	0.1 [-0.3–0.5]	0.648	0.3 [-0.1–0.6]	0.171	0.3 [0.0–0.5]	0.068	1.1 [0.7–1.6]	**<0.001**
Amygdala_GIF_Left	0.5 [-1.0–2.1]	0.496	0.4 [-0.8–2.0]	0.635	0.9 [-0.2–1.9]	0.115	1.1 [-0.6–2.8]	0.187
Amygdala_GIF_Right	-0.3 [-1.6–0.9]	0.595	0.2 [-1.0–1.5]	0.709	0.3 [-0.6–1.2]	0.487	1.9 [0.5–3.4]	*0.009*
Hippocampus_STEPS_Left	0.3 [-0.8–1.3]	0.621	1.3 [0.2–2.3]	*0.015*	1.1 [0.3–1.8]	*0.004*	1.1 [-0.2–2.4]	0.093
Hippocampus_STEPS_Right	0.1 [-0.9–1.1]	0.804	0.9 [-0.1–1.9]	0.092	0.7 [-0.1–1.4]	0.076	1.5 [0.2–2.7]	*0.020*
Caudate_STEPS_Left	-0.7 [-2.9–1.5]	0.540	0.1 [-2.0–2.2]	0.817	0.1 [-1.5–1.6]	0.947	5.9 [3.5–8.4]	**<0.001**
Caudate_STEPS_Right	1.8 [-0.3–3.8]	0.093	0.8 [-1.1–2.7]	0.405	1.4 [0.0–2.9]	*0.045*	3.4 [1.2–5.6]	*0.003*
Thalamus_STEPS_Left	-0.6 [-1.6–0.4]	0.219	0.2 [-0.8–1.1]	0.761	-0.3 [-1.0–0.5]	0.540	1.9 [0.7–3.1]	*0.002*
Thalamus_STEPS_Right	-0.2 [-1.3–0.8]	0.706	0.7 [-0.4–1.7]	0.222	0.3 [-0.5–1.1]	0.451	1.8 [0.6–3.1]	*0.005*

** All analyses adjusted for age, gender and scanner type:* ***p < 0.001 bolded****, p < 0.05 italicised*

Table 6-4 continued: Linear regression analysis results with adjusted mean difference and 95% CIs comparing annual rates of whole-brain and subcortical atrophy for each genetic subgroup with non-carrier controls. *

Measure	*C9orf72*							
	Presymptomatic early (>10 EYO)		Presymptomatic late (<10 EYO)		Presymptomatic (all)		Symptomatic	
BSI	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**
Whole-brain GBSI	0.1 [-0.1–0.3]	0.584	0.2 [-0.1–0.6]	0.145	0.1 [-0.1–0.3]	0.207	0.5 [0.2–0.8]	**<0.001**
ABSI_Left	0.2 [-0.6–1.1]	0.587	0.6 [-0.6–1.7]	0.331	0.3 [-0.4–1.0]	0.334	0.5 [-0.5–1.5]	0.306
ABSI_Right	0.1 [-0.6–0.8]	0.840	0.5 [-0.6–1.4]	0.467	0.2 [-0.4–0.8]	0.576	1.0 [0.1–1.8]	*0.026*
HBSI_Left	0.7 [0.1–1.4]	*0.020*	0.3 [-0.6–1.1]	0.558	0.6 [0.1–1.1]	*0.030*	1.0 [0.3–1.8]	*0.007*
HBSI_Right	0.5 [-0.1–1.1]	0.078	0.1 [-0.7–0.9]	0.760	0.4 [-0.1–0.9]	0.114	0.6 [-0.1–1.3]	0.091
CBSI_Left	0.0 [-1.3–1.4]	0.944	1.6 [-0.3–3.6]	0.099	0.5 [-0.6–1.8]	0.335	5.7 [4.0–7.4]	**<0.001**
CBSI_Right	0.0 [-1.2–1.2]	0.974	0.5 [-1.2–2.2]	0.554	0.2 [-0.8–1.2]	0.729	3.7 [2.3–5.1]	**<0.001**
TBSI_Left	0.1 [-0.3–0.4]	0.772	0.6 [0.1–1.1]	*0.019*	0.2 [-0.1–0.5]	0.137	0.9 [0.5–1.3]	**<0.001**
TBSI_Right	0.1 [-0.2–0.4]	0.619	0.4 [0.0–0.9]	*0.050*	0.2 [-0.1–0.5]	0.139	0.5 [0.1–0.8]	*0.013*
STEPS_GIF_Volume Difference								
Whole-brain GIF	0.1 [-0.2–0.4]	0.383	0.3 [-0.2–0.6]	0.252	0.2 [-0.1–0.4]	0.184	0.8 [0.6–1.2]	**<0.001**
Amygdala_GIF_Left	-0.5 [-1.6–0.5]	0.314	0.1 [-1.4–1.5]	0.947	-0.4 [-1.3–0.5]	0.443	1.2 [0.0–2.5]	0.055
Amygdala_GIF_Right	-0.3 [-1.3–0.6]	0.466	-1.0 [-2.3–0.2]	0.101	-0.5 [-1.3–0.2]	0.156	0.9 [-0.2–2.0]	0.112
Hippocampus_STEPS_Left	0.5 [-0.2–1.3]	0.181	0.1 [-1.0–1.1]	0.909	0.4 [-0.3–1.0]	0.255	0.9 [0.0–1.8]	*0.048*
Hippocampus_STEPS_Right	0.0 [-0.8–0.7]	0.937	-0.6 [-1.7–0.5]	0.288	-0.2 [-0.0–0.4]	0.526	1.4 [0.5–2.3]	*0.003*
Caudate_STEPS_Left	0.3 [-1.2–1.8]	0.702	0.3 [-1.9–2.5]	0.782	0.3 [-1.0–1.7]	0.622	1.7 [-0.3–3.6]	*0.088*
Caudate_STEPS_Right	-1.1 [-2.4–0.4]	0.149	-0.4 [-2.3–1.5]	0.705	-0.8 [-2.0–0.4]	0.185	1.7 [0.1–3.4]	*0.039*
Thalamus_STEPS_Left	0.0 [-0.7–0.7]	0.963	0.3 [-0.7–1.3]	0.605	0.1 [-0.5–0.7]	0.813	2.0 [1.1–2.9]	**<0.001**
Thalamus_STEPS_Right	-0.2 [-1.0–0.6]	0.619	1.7 [0.6–2.8]	*0.003*	0.4 [-0.3–1.0]	0.254	1.1 [0.2–2.1]	*0.018*

** All analyses adjusted for age, gender and scanner type:* ***p* < 0.001 *bolded***, *p < 0.05 italicised*

Table 6-4 continued: Linear regression analysis results with adjusted mean difference and 95% CIs comparing annual rates of whole-brain and subcortical atrophy for each genetic subgroup with non-carrier controls. *

measure	*GRN*							
	Presymptomatic early (>10 EYO)		Presymptomatic late (<10 EYO)		Presymptomatic (all)		Symptomatic	
BSI	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**	**Coef. [95% CI]**	***p*-value**
Whole-brain GBSI	0.1 [-0.1–0.3]	0.450	0.2 [0.0–0.5]	0.054	0.2 [0.0–0.4]	0.070	1.4 [1.1–1.7]	**<0.001**
ABSI_Left	0.5 [-0.3–1.2]	0.245	0.6 [-0.1–1.4]	0.104	0.5 [0.0–1.1]	0.069	4.6 [3.6–5.7]	**<0.001**
ABSI_Right	0.2 [-0.5–0.9]	0.579	0.2 [-0.5–0.9]	0.561	0.2 [-0.3–0.7]	0.454	2.9 [2.0–3.8]	**<0.001**
HBSI_Left	0.5 [-0.1–1.1]	0.102	0.7 [0.1–1.3]	*0.021*	0.6 [0.1–1.0]	*0.011*	3.4 [2.6–4.1]	**<0.001**
HBSI_Right	0.4 [-0.2–0.9]	0.184	0.5 [0.0–0.9]	*0.050*	0.4 [0.0–0.9]	*0.034*	2.5 [1.8–3.2]	**<0.001**
CBSI_Left	0.2 [-1.1–1.5]	0.789	1.5 [0.2–2.8]	*0.028*	0.8 [-0.2–1.8]	0.108	17.4 [15.6–19.1]	**<0.001**
CBSI_Right	0.0 [-1.1–1.1]	0.964	1.4 [0.3–2.5]	*0.015*	0.7 [-0.2–1.5]	0.115	13.6 [12.1–15.1]	**<0.001**
TBSI_Left	0.0 [-0.3–0.3]	0.880	0.3 [-0.1–0.6]	0.177	0.1 [-0.1–0.3]	0.441	3.5 [3.0–3.9]	**<0.001**
TBSI_Right	-0.1 [-0.4–0.2]	0.683	0.4 [0.1–0.7]	*0.005*	0.2 [0.0–0.4]	0.124	1.3 [0.9–1.7]	**<0.001**
STEPS_GIF_Volume Difference								
Whole-brain GIF	0.0 [-0.3–0.2]	0.775	0.2 [-0.1–0.4]	0.222	0.1 [-0.1–0.3]	0.555	2.0 [1.6–2.3]	**<0.001**
Amygdala_GIF_Left	0.2 [-0.8–1.2]	0.685	0.0 [-1.0–1.0]	0.938	0.1 [-0.7–0.8]	0.846	1.5 [0.0–2.5]	0.058
Amygdala_GIF_Right	-0.1 [-1.0–0.8]	0.863	0.6 [-0.3–1.4]	0.170	0.3 [-0.4–0.9]	0.437	0.5 [-0.6–1.6]	0.384
Hippocampus_STEPS_Left	0.1 [-0.6–0.8]	0.822	0.6 [-0.1–1.3]	0.104	0.3 [-0.2–0.9]	0.241	2.5 [1.5–3.4]	**<0.001**
Hippocampus_STEPS_Right	0.3 [-0.4–1.0]	0.419	0.0 [-0.7–0.7]	0.994	0.1 [-0.4–0.7]	0.618	2.2 [1.2–3.1]	**<0.001**
Caudate_STEPS_Left	0.1 [-1.3–1.6]	0.854	1.2 [-0.2–2.6]	0.104	0.3 [-0.4–1.8]	0.235	11.5 [9.6–13.5]	**<0.001**
Caudate_STEPS_Right	0.3 [-1.0–1.6]	0.640	1.0 [-0.3–2.2]	0.139	0.6 [-0.3–1.6]	0.191	10.4 [8.7–12.2]	**<0.001**
Thalamus_STEPS_Left	0.1 [-0.6–0.7]	0.852	0.2 [-0.5–0.9]	0.554	0.1 [-0.4–0.6]	0.624	6.1 [5.1–7.0]	**<0.001**
Thalamus_STEPS_Right	-0.1 [-0.8–0.6]	0.779	0.4 [-0.3–1.2]	0.253	0.2 [-0.4–0.7]	0.595	2.5 [1.5–3.4]	**<0.001**

* All analyses adjusted for age, gender and scanner type: **p *< 0.001* bolded**, p < 0.05 italicised

6.3.1.2 Effect size and sample size results

Effect size results with 95% BCa CIs are summarised in Table 6-5. Effect sizes of one or above are presented in bold, indicating a large effect with good group separation between FTD subgroups and controls based on the rates of annual volume change for that measure. Table 6-6 reports the sample size calculations with 95% BCa CI to detect a 30% reduction in atrophy rates with 80% power at a 5% statistical significance level as the treatment effect. These sample sizes are adjusted for the longitudinal atrophy rates in the non-carrier controls for each measure. Empty cells represent measures where the annual rate of change was faster in the non-carrier controls than in the FTD subgroup. Therefore, applying the same calculations would derive a sample size result demonstrating the number of individuals required to detect a 30% increase in atrophy in the FTD subgroup as a treatment effect and therefore not appropriate to include in the Table summary.

Sample size estimates under 100 participants required per treatment arm to detect a 30% reduction in atrophy rates are in blue and bold. Those in green and underlined required between 100 and 200 individuals to detect the specified treatment effect. Measures with an infinite upper limit indicate those where an accurate calculation of the distribution could not be determined from the bootstrap method employing 2000 repetitions due to insufficient difference between the genetic subgroup and non-carrier controls.

In *MAPT* carriers, whole-brain and subcortical longitudinal measures failed to result in effect sizes over 1 for any of the presymptomatic subgroups, indicating a lack of group separation from the non-carrier controls. This can be seen in Figure 6-1 where the 95% CI lines for the all presymptomatic *MAPT* carriers (in red) substantially overlap with those of the non-carrier controls (in blue) across all 18 measures. The resulting sample sizes were universally high with infinite or unfeasible upper 95% BCa CI estimates. In

the *MAPT* symptomatic group, whole-brain GBSI, left CBSI and the whole-brain GIF volume-difference measure all produced effect sizes at or above 1. Several other measures suggested reasonable group separation with effect sizes of 0.8 and 0.9, including the right CBSI and measures of thalamic atrophy. The resulting sample size calculations suggest whole-brain measures (both direct BSI and indirect volume-difference) and direct measures of caudate atrophy (particularly left CBSI) show promise as potential biomarkers in this population. The indirect GIF whole-brain volume-loss measure had the lowest point estimate of 91 [30–586]. In addition, the reduced upper CI limit for the left CBSI indicates that it could be a good additional choice: 133 [63–399].

The effect size calculations in the *C9orf72* presymptomatic groups also failed to reach one or above. However, there are several results indicating potential in the late presymptomatic group. The left CBSI (0.6 [-0.3–1.5]), left TBSI (0.7 [0.3–1.1]), right TBSI (0.6 [0.0–1.1]) and right thalamus STEPS volume-difference (0.7 [0.1–1.3]) results suggest a reasonable level of group separation using these measures, which is supported by the regression analysis above. The sample size calculations for this group demonstrate that the left TBSI (331 [142–1372]) and right STEPS thalamus volume difference (367 [153–53339]) provide the best point estimate but with the previously mentioned issues when considering the information provided by the CIs. Effect sizes were highest in the symptomatic *C9orf72* cohort for the CBSI bilaterally (1.4 and 1.2; left and right) and the indirect GIF whole-brain volume-difference measure at 1.3. The sample size estimates derived from these measures were 83 [50–189] for the left CBSI, 125 [73–241] for the right CBSI and 102 [52–300] for the GIF whole-brain measure. The narrow CIs suggest all three could provide robust biomarkers of change in a putative trial. Although slightly above 200, sample size results for measures of the left thalamus (both direct and indirect) and the right hippocampal STEPS measures imply that automated measures of these subcortical regions may also demonstrate potential for detecting treatment effects.

Finally, in the *GRN* presymptomatic groups, all effect sizes were 0.5 or below. The highest effect size was for the left HBSI in the late presymptomatic group with 0.5 [0.2–0.9]. This translated into the lowest sample size across the presymptomatic subgroups with an estimate of 583 [215–5194]. Although all sample sizes were high in the *GRN* presymptomatic groups, it is encouraging to note that there were fewer estimates with an infinite upper limit, indicating an improvement in detecting group separation using the bootstrapping method with 2000 repetitions, which is supported by the regression results discussed in Section 6.3.1.1. The larger cohort numbers in the *GRN* subgroup contributed to these more reliable estimates with narrower CIs and increasing the size of the cohorts for the other genetic subgroups will be similarly useful as the GENFI study progresses with continued enrolment and increased longitudinal data collection. Effect sizes for the symptomatic *GRN* group were high across all the BSI and most of the indirect volume-difference measures, except for the amygdala GIF measures bilaterally. The resulting sample size calculations revealed that multiple whole-brain and subcortical measures would require less than 100 participants per treatment arm to detect a 30% treatment effect (80% power, 5% significance). The lowest estimate was provided by the whole-brain GBSI measure with 45 [25–145] individuals. The other measures providing estimates below 100 were the left TBSI (63 [27–216]), left CBSI (84 [44–166]) and whole-brain GIF volume-difference measure (96 [40–392]). The left HBSI, right CBSI, bilateral caudate STEPS volume difference and left thalamus STEPS longitudinal measure also demonstrated promise with estimates between 100 and 200 individuals required per treatment arm.

Table 6-5 Effect size calculations with 95% BCa confidence intervals for the whole-brain and subcortical longitudinal measures for each of the genetic FTD subgroups*.

Measure	*MAPT*							
	Early Presymptomatic (>10 EYO)		**Late Presymptomatic (<10 EYO)**		**Presymptomatic (all)**		**Symptomatic**	
BSI	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)
Whole-brain GBSI	0.7	[0.1–1.3]	0.1	[-0.8–1.0]	0.5	[0.1–0.8]	**1.0**	**[0.4–1.7]**
ABSI_Left	0.3	[-0.4–1.0]	0.3	[-0.2–0.8]	0.3	[0.0–0.6]	0.6	[0.0–1.3]
ABSI_Right	0.2	[-0.4–0.8]	-0.1	[-0.8–0.6]	0.1	[-0.2–0.5]	0.9	[0.1–1.6]
HBSI_Left	0.0	[-0.6–0.6]	0.4	[-0.1–0.9]	0.3	[0.0–0.6]	0.6	[0.0–1.3]
HBSI_Right	0.1	[-0.4–0.6]	0.2	[-0.4–0.7]	0.2	[-0.1–0.6]	0.4	[-0.2–1.0]
CBSI_Left	0.1	[-0.5–0.7]	0.4	[0.0–0.9]	0.3	[0.0–0.6]	**1.1**	**[0.5–1.8]**
CBSI_Right	0.2	[-0.3–0.8]	0.1	[-0.5–0.7]	0.2	[-0.1–0.6]	0.9	[0.3–1.6]
TBSI_Left	-0.5	[-1.2–0.1]	0.3	[-0.2–0.9]	0.1	[-0.3–0.5]	0.7	[0.3–1.1]
TBSI_Right	0.0	[-0.5–0.5]	0.2	[-0.3–0.8]	0.1	[-0.2–0.5]	0.8	[0.1–1.5]
STEPS_GIF_ volume difference								
Whole-brain GIF	-0.1	[-0.6–0.5]	0.5	[0.0–1.0]	0.2	[-0.1–0.6]	**1.4**	**[0.2–2.6]**
Amygdala_GIF_Left	0.1	[-0.5–0.7]	0.2	[-0.4–0.7]	0.2	[-0.2–0.5]	0.6	[-0.2–1.3]
Amygdala_GIF_Right	-0.1	[-0.8–0.6]	0.1	[-0.5–0.7]	0.1	[-0.3–0.5]	0.5	[-0.1–1.2]
Hippocampus_STEPS_Left	0.0	[-0.5–0.6]	0.5	[0.0–0.9]	0.3	[0.0–0.6]	0.6	[-0.4–1.7]
Hippocampus_STEPS_Right	-0.1	[-0.6–0.5]	0.3	[-0.2–0.8]	0.2	[-0.1–0.5]	0.8	[-0.1–1.8]
Caudate_STEPS_Left	-0.5	[-1.1–0.0]	0.2	[-0.5–0.8]	-0.1	[-0.5–0.4]	0.7	[0.1–1.4]
Caudate_STEPS_Right	0.4	[-0.2–0.9]	0.4	[-0.1–0.9]	0.4	[0.1–0.8]	0.6	[0.0–1.2]
Thalamus_STEPS_Left	-0.4	[-0.9–0.1]	0.1	[-0.5–0.8]	-0.2	[-0.6–0.2]	0.8	[-1.1–2.7]
Thalamus_STEPS_Right	-0.1	[-0.7–0.4]	0.5	[-0.1–1.0]	0.1	[-0.3–0.6]	0.9	[-0.1–1.8]

* Effect sizes of 1.0 or larger are in **bold** to indicate a high degree of group separation from non-carrier controls for that longitudinal measure

Table 6-5 continued: Effect size calculations with 95% BCa confidence intervals for the whole-brain and subcortical longitudinal measures for each of the genetic FTD subgroups*.

Measure	*C9orf72*							
	Early Presymptomatic (>10 EYO)		Late Presymptomatic (<10 EYO)		Presymptomatic (all)		Symptomatic	
BSI	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)
Whole-brain GBSI	0.1	[-0.3–0.5]	0.5	[0.0–0.9]	0.2	[-0.1–0.5]	0.8	[0.4–1.2]
ABSI_Left	0.1	[-0.3–0.5]	0.6	[-0.1–1.2]	0.2	[-0.2–0.6]	0.3	[-0.1–0.8]
ABSI_Right	-0.1	[-0.5–0.3]	0.5	[-0.2–1.1]	0.1	[-0.3–0.4]	0.5	[0.1–0.9]
HBSI_Left	0.3	[0.0–0.7]	0.2	[-0.3–0.8]	0.3	[0.0–0.6]	0.5	[0.1–0.8]
HBSI_Right	0.3	[0.0–0.7]	0.1	[-0.4–0.6]	0.2	[0.0–0.5]	0.4	[0.0–0.7]
CBSI_Left	-0.1	[-0.5–0.3]	0.6	[-0.3–1.5]	0.2	[0.0–0.5]	**1.4**	**[1.0–1.9]**
CBSI_Right	-0.1	[-0.5–0.3]	0.4	[-0.1–0.9]	0.1	[-0.2–0.4]	**1.2**	**[0.8–1.6]**
TBSI_Left	-0.1	[-0.5–0.3]	0.7	[0.3–1.1]	0.2	[-0.1–0.5]	0.8	[0.3–1.2]
TBSI_Right	0.1	[-0.3–0.5]	0.6	[0.0–1.1]	0.3	[-0.1–0.6]	0.6	[0.1–1.0]
STEPS_GIF_volume difference								
Whole-brain GIF	0.1	[-0.2–0.5]	0.5	[0.0–1.0]	0.2	[-0.1–0.5]	**1.3**	**[0.7–1.9]**
Amygdala_GIF_Left	-0.4	[-0.8–0.1]	0.1	[-0.5–0.7]	-0.2	[-0.6–0.1]	0.6	[0.1–1.1]
Amygdala_GIF_Right	-0.2	[-0.6–0.1]	-0.5	[-1.1–0.1]	-0.3	[-0.7–0.0]	0.4	[0.0–0.8]
Hippocampus_STEPS_Left	0.0	[-0.4–0.4]	0.1	[-0.5–0.7]	0.1	[-0.3–0.4]	0.6	[0.1–1.1]
Hippocampus_STEPS_Right	-0.2	[-0.6–0.2]	-0.1	[-0.7–0.4]	-0.2	[-0.5–0.1]	0.9	[0.3–1.4]
Caudate_STEPS_Left	0.1	[-0.3–0.5]	0.3	[-0.2–0.9]	0.2	[-0.2–0.5]	0.6	[0.1–1.0]
Caudate_STEPS_Right	-0.6	[-1.1–-0.2]	-0.1	[-0.6–0.5]	-0.4	[-0.7–0.0]	0.6	[0.2–0.9]
Thalamus_STEPS_Left	-0.1	[-0.5–0.3]	0.1	[-0.5–0.7]	0.0	[-0.4–0.3]	0.8	[0.3–1.2]
Thalamus_STEPS_Right	-0.1	[-0.4–0.3]	0.7	[0.1–1.3]	0.2	[-0.1–0.5]	0.6	[0.2–1.1]

* Effect sizes of 1.0 or larger are in bold to indicate a high degree of group separation from non-carrier controls for that longitudinal measure

Table 6-5 continued: Effect size calculations with 95% BCa confidence intervals for the whole-brain and subcortical longitudinal measures for each of the genetic FTD subgroups*.

Measure	*GRN*							
	Early Presymptomatic (>10 EYO)		**Late Presymptomatic (<10 EYO)**		**Presymptomatic (all)**		**Symptomatic**	
BSI	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)	Effect Size	95% CI (BCa)
Whole-brain GBSI	0.1	[-0.3–0.5]	0.4	[0.1–0.7]	0.3	[0.0–0.5]	**2.0**	**[1.1–2.8]**
ABSI_Left	0.2	[-0.2–0.6]	0.4	[0.1–0.7]	0.3	[0.0–0.6]	**1.2**	**[0.2–2.3]**
ABSI_Right	0.0	[-0.4–0.4]	0.2	[-0.1–0.6]	0.1	[-0.1–0.4]	**1.0**	**[0.5–1.4]**
HBSI_Left	0.3	[0.0–0.6]	0.5	[0.2–0.9]	0.4	[0.1–0.7]	**1.5**	**[0.7–2.2]**
HBSI_Right	0.3	[0.0–0.6]	0.3	[0.1–0.6]	0.3	[0.1–0.5]	0.9	[0.4–1.4]
CBSI_Left	0.1	[-0.2–0.5]	0.4	[0.2–0.6]	0.3	[0.1–0.4]	**1.4**	**[0.9–2.0]**
CBSI_Right	-0.2	[-0.5–0.2]	0.4	[0.2–0.5]	0.2	[0.1–0.4]	**1.4**	**[0.9–1.8]**
TBSI_Left	-0.2	[-0.5–0.2]	0.3	[0.0–0.6]	0.1	[-0.1–0.4]	**1.7**	**[0.7–2.6]**
TBSI_Right	-0.2	[-0.6–0.2]	0.5	[0.2–0.7]	0.2	[0.0–0.4]	0.9	[0.3–1.5]
STEPS_GIF_Volume difference								
Whole-brain GIF	-0.1	[-0.5–0.3]	0.4	[0.0–0.7]	0.1	[-0.2–0.4]	**1.3**	**[0.5–2.2]**
Amygdala_GIF_Left	0.1	[-0.3–0.4]	0.0	[-0.3–0.4]	0.0	[-0.2–0.3]	0.6	[0.1–1.1]
Amygdala_GIF_Right	-0.1	[-0.4–0.3]	0.3	[0.0–0.6]	0.1	[-0.1–0.4]	0.4	[0.0–0.8]
Hippocampus_STEPS_Left	-0.2	[-0.5–0.2]	0.4	[0.1–0.7]	0.1	[-0.1–0.4]	0.9	[0.4–1.3]
Hippocampus_STEPS_Right	0.1	[-0.3–0.5]	0.1	[-0.2–0.5]	0.1	[-0.2–0.4]	0.6	[0.2–1.0]
Caudate_STEPS_Left	0.0	[-0.3–0.4]	0.4	[0.2–0.7]	0.3	[0.0–0.5]	**1.0**	**[0.6–1.4]**
Caudate_STEPS_Right	0.0	[-0.3–0.4]	0.3	[0.0–0.6]	0.2	[-0.1–0.4]	**1.2**	**[0.9–1.6]**
Thalamus_STEPS_Left	0.1	[-0.3–0.5]	0.1	[-0.3–0.5]	0.1	[-0.2–0.4]	**1.5**	**[0.9–2.1]**
Thalamus_STEPS_Right	0.0	[-0.4–0.4]	0.2	[-0.2–0.6]	0.1	[-0.2–0.4]	0.9	[0.3–1.4]

* Effect sizes of 1.0 or larger are in bold to indicate a high degree of group separation from non-carrier controls for that longitudinal measure

Table 6-6 Sample size estimates with 95% BCa CI to detect a 30% reduction in atrophy rate with 80% power and 5% significance. Results in **blue** and bold are for measures that required <100 participants per treatment arm, green and underlined demonstrate sample size estimates between 100 and 200 individuals per treatment arm*

Measure	*MAPT*							
	Early Presymptomatic (>10 EYO)		Late Presymptomatic (<10 EYO)		Presymptomatic (all)		Symptomatic	
BSI	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)
Whole-brain GBSI	400	[114–3.5e+05]	12916	[195–∞]	781	[268–44127]	162	[74–557]
ABSI_Left	1992	[241–∞]	2003	[336–∞]	1639	[523–∞]	435	[119–∞]
ABSI_Right	3619	[274–∞]	19427	[638–∞]	9230	[796–∞]	240	[74–2679]
HBSI_Left	858488	[452–∞]	1049	[250–∞]	2235	[654–∞]	453	[135–∞]
HBSI_Right	22350	[454–∞]	5834	[315–∞]	3639	[653–∞]	882	[172–∞]
CBSI_Left	22281	[445–∞]	913	[244–∞]	1757	[572–∞]	133	[63–399]
CBSI_Right	3669	[338–∞]	11818	[396–∞]	3014	[685–∞]	200	[79–1389]
TBSI_Left	654	[10207–∞]	1444	[273–∞]	20629	[837–∞]	348	[131–2011]
TBSI_Right			3348	[327–∞]	8382	[702–∞]	286	[83–8766]
STEPS_GIF_Volume Difference								
Whole-brain volume difference	69147	[601–∞]	684	[219–∞]	3376	[570–∞]	**91**	**[30–586]**
Amygdala_GIF_Left	18880	[416–∞]	6884	[371–∞]	4567	[656–∞]	561	[99–∞]
Amygdala_GIF_Right			12185	[394–∞]	19789	[848–∞]	632	[132–∞]
Hippocampus_STEPS_Left	271902	[616–∞]	849	[202–∞]	2025	[477–∞]	449	[123–∞]
Hippocampus_STEPS_Right	22483	[745–∞]	1664	[287–∞]	4302	[615–∞]	267	[56–35288]
Caudate_STEPS_Left			5688	[264–∞]	47363	[1371–∞]	325	[131–9266]
Caudate_STEPS_Right	1236	[193–∞]	1029	[248–∞]	1024	[310–1.4e+05]	483	[127–∞]
Thalamus_STEPS_Left			9990	[250–∞]	5179	[4037–∞]	285	[27–∞]
Thalamus_STEPS_Right			780	[180–∞]	8996	[516–∞]	228	[58–∞]

* Empty cells denote values for which the non-carrier control rate was faster than the FTD cohort and thus not appropriate for inclusion for the specified treatment effect of a reduction in 30% atrophy (see Section 6.3.1.2).

Table 6-7 continued: Sample size estimates with 95% BCa CI to detect a 30% reduction in atrophy rate with 80% power and 5% significance. Results in **blue** and bold are for measures that required <100 participants per treatment arm, green and underlined demonstrate sample size estimates between 100 and 200 individuals per treatment arm*

Measure	*C9ORF72*							
	Early Presymptomatic (>10 EYO)		Late Presymptomatic (<10 EYO)		Presymptomatic (all)		Symptomatic	
BSI	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)
Whole-brain GBSI	22804	[941–∞]	774	[233–∞]	4153	[846–∞]	291	[134–1694]
ABSI_Left	28624	[717–∞]	545	[112–∞]	4190	[581–∞]	1540	[330–∞]
ABSI_Right	22426	[1917–∞]	852	[139–∞]	30803	[1088–∞]	740	[231–22911]
HBSI_Left	1433	[402–∞]	3441	[342–∞]	1757	[523–1.6e+06]	804	[269–26325]
HBSI_Right	1591	[401–∞]	16061	[464–∞]	2888	[613–∞]	1301	[354–∞]
CBSI_Left	26679	[2003–∞]	470	[247–4492]	3156	[886–∞]	**83**	**[50–189]**
CBSI_Right	10905	[2411–∞]	1040	[289–∞]	12266	[1046–∞]	125	[73–241]
TBSI_Left	25197	[1981–∞]	331	[142–1372]	4640	[728–∞]	307	[132–1999]
TBSI_Right	21350	[691–∞]	522	[156–∞]	2522	[495–∞]	554	[182–68423]
STEPS_GIF_Volume Difference								
Whole-brain volume difference	9893	[727–∞]	741	[217–∞]	3088	[603–∞]	102	[52–300]
Amygdala_GIF_Left			21521	[371–∞]	3800	[8849–∞]	491	[152–12549]
Amygdala_GIF_Right					1636	[1.1e+07–∞]	1177	[282–∞]
Hippocampus_STEPS_Left	193205	[963–∞]	12682	[412–∞]	50967	[1224–∞]	528	[149–9744]
Hippocampus_STEPS_Right			9307	[1009–∞]	4077	[13004–∞]	231	[99–1838]
Caudate_STEPS_Left	13611	[641–∞]	1697	[237–∞]	6680	[674–∞]	522	[191–86770]
Caudate_STEPS_Right			28045	[734–∞]	1213	[99086–∞]	573	[215–6472]
Thalamus_STEPS_Left	17380	[1901–∞]	10031	[312–∞]	576450	[1834–∞]	279	[131–1875]
Thalamus_STEPS_Right			367	[153–53339]	5483	[744–∞]	450	[163–5252]

* Empty cells denote values for which the non-carrier control rate was faster than the FTD cohort and thus not appropriate for inclusion for the specified treatment effect of a reduction in 30% atrophy (see Section 6.3.1.2).

Table 6-8 continued: Sample size estimates with 95% BCa CI to detect a 30% reduction in atrophy rate with 80% power and 5% significance. Results in **blue** and bold are for measures that required <100 participants per treatment arm, green and underlined demonstrate sample size estimates between 100 and 200 individuals per treatment arm*

Measure	*GRN*							
	Early Presymptomatic (>10 EYO)		**Late Presymptomatic (<10 EYO)**		**Presymptomatic (all)**		**Symptomatic**	
BSI	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)	Sample Size	95% CI (BCa)
Whole-brain GBSI	16197	[662–∞]	1024	[356–34594]	2140	[614–4.6e+05]	**45**	**[25–145]**
ABSI_Left	5123	[555–∞]	1025	[344–37001]	1893	[572–110805]	225	[89–1580]
ABSI_Right	550246	[1124–∞]	3002	[499–∞]	9311	[1105–∞]	299	[120–3822]
HBSI_Left	2143	[497–∞]	583	[215–5194]	1044	[403–11020]	142	[57–1183]
HBSI_Right	2266	[505–∞]	1498	[575–∞]	1787	[652–32859]	306	[120–3381]
CBSI_Left	9917	[681–∞]	1253	[586–7961]	2335	[1028–11714]	**84**	**[44–166]**
CBSI_Right	7393	[2751–∞]	1399	[676–10018]	3734	[1449–1.1e+05]	114	[60–247]
TBSI_Left	7403	[3484–∞]	2043	[451–∞]	11944	[1314–∞]	**63**	**[27–216]**
TBSI_Right	5404	[4220–∞]	744	[336–4456]	3768	[942–∞]	213	[79–1477]
STEPS_GIF_Volume Difference								
Whole-brain volume difference	31982	[1785–∞]	1121	[333–2.1e+05]	12261	[978–∞]	**96**	**[40–392]**
Amygdala_GIF_Left	56672	[1195–∞]	159044	[1073–∞]	88678	[1795–∞]	454	[139–9891]
Amygdala_GIF_Right			2207	[553–∞]	11698	[1320–∞]	1255	[283–∞]
Hippocampus_STEPS_Left			1075	[334–1.7e+05]	9539	[1192–∞]	232	[107–906]
Hippocampus_STEPS_Right	26820	[810–∞]	8412	[710–∞]	12721	[1054–∞]	536	[172–8389]
Caudate_STEPS_Left	105863	[1048–∞]	924	[362–9371]	2650	[762–2.8e+05]	178	[90–432]
Caudate_STEPS_Right	192547	[1155–∞]	1658	[522–∞]	4661	[988–∞]	115	[71–198]
Thalamus_STEPS_Left	15707	[690–∞]	13540	[675–∞]	14409	[889–∞]	116	[55–504]
Thalamus_STEPS_Right			3507	[429–∞]	14294	[996–∞]	376	[123–8832]

* Empty cells denote values for which the non-carrier control rate was faster than the FTD cohort and thus not appropriate for inclusion for the specified treatment effect of a reduction in 30% atrophy (see Section 6.3.1.2).

6.4 Discussion

The current study collated structural imaging in a large longitudinal genetic FTD cohort to investigate the potential of automated measures of whole-brain and subcortical atrophy as non-invasive biomarkers for therapeutic interventions. Both presymptomatic and symptomatic individuals were included to 1) evaluate the potential for detecting preclinical neuroanatomical change and 2) to contribute to the limited and inconsistent longitudinal structural neuroimaging findings in genetic FTD. With the majority of publications in genetic FTD being cross-sectional, baseline comparisons in the cohort were also investigated to compare with the available literature in this population.

6.4.1 Baseline comparisons

At presentation, the baseline neuroanatomical profile of the genetic cohort was largely consistent with current cross-sectional literature in genetic FTD but with some notable exceptions across the genetic spectrum.

6.4.1.1 MAPT

Although not statistically significant, there was a suggestion of lower right amygdalar volumes at baseline in the presymptomatic *MAPT* group (1.9%; $p = 0.071$), which is in line with previous research demonstrating reduced volumes 5–10 years before symptom onset (Cash et al., 2018; Rohrer et al., 2015). A recent case-study also demonstrated the only structural change evident in a presymptomatic *MAPT* carrier was reduced volume in the right amygdala (Bevan-Jones et al., 2019). The current results failed to find any presymptomatic hippocampal volume reduction, which is consistent with the findings reported in Dopper et al. (2014) and Whitwell, Josephs et al. (2011); however, in contrast to Rohrer et al. (2015) where bilateral hippocampal volume reductions were reported five years before symptom onset. It is important to note that the current late presymptomatic group included individuals up to 10 years before symptom onset. Therefore, the lack of significant bilateral hippocampal reductions may be explained by this difference in the EYO window and inclusion of individuals at an earlier preclinical stage.

Further investigations to clarify the preclinical timings at which hippocampal atrophy becomes evident will be of use in presymptomatic genetic FTD research. The symptomatic *MAPT* group presented with the classic medial temporal lobe atrophy frequently reported and showed severely reduced bilateral amygdala and hippocampi (all $p < 0.001$) ranging from 31.5% to 35.2% reduction compared with non-carrier controls. The left thalamus also demonstrated a 13.9% lower volume, similar to previous publications (Bocchetta et al., 2018).

6.4.1.2 C9orf72

The *C9orf72* presymptomatic groups presented with early subcortical volume reduction in the amygdala, hippocampus and thalamus bilaterally compared with non-carrier controls. This was evident more than 10 years before symptom onset in line with previous reports of a very early profile of presymptomatic volume loss in *C9orf72* (Bertrand et al., 2018; Cash et al., 2018; Popuri et al., 2018; Rohrer et al., 2015). It is important to note that only 40 of the 382 GENFI participants included in this current analysis overlapped with those in the seminal Rohrer and colleagues (2015) publication; therefore, the current presymptomatic subcortical volume-difference results can be considered a predominantly independent replication of these previous findings. Significantly reduced volumes in the left and right caudate become evident in the late presymptomatic group. This provides additional findings to the current inconsistent literature with regard to presymptomatic caudate involvement. The current results are in line with a single previous study that also reported bilateral caudate reductions (Lee et al., 2017). However, most previous investigations have failed to report smaller caudate volumes (Cash et al., 2018; Floeter et al., 2016; Panman et al., 2019; Papma et al., 2017), whilst others have reported substantial but only left-sided volume differences (Popuri et al., 2018; Walhout et al., 2015) in presymptomatic *C9orf72* carriers compared with controls. Further focused investigations of caudate volumetry in presymptomatic *C9orf72* individuals would be of value to clarify these inconsistencies. At presentation, symptomatic *C9orf72* patients demonstrated significantly smaller whole-brain and subcortical volumes across all regions than non-carrier

controls. The largest difference was found in the left thalamus (23.1% smaller than controls), consistent with thalamic involvement being a key feature of *C9orf72* neuroanatomical presentation (Bocchetta, Gordon et al., 2018; Cash et al., 2018; Lee et al., 2014, 2017; Rohrer et al., 2015).

6.4.1.3 GRN

In the late presymptomatic *GRN* group, baseline comparisons showed carriers presented with significantly smaller thalami bilaterally (6.2% and 6.3%; left and right) than non-carriers. This is in contrast with the literature, which has consistently failed to report any significant volumetric differences in presymptomatic *GRN* carriers in the thalamic region (Cash et al., 2018; Dopper et al., 2014; Olm et al., 2018; Panman et al., 2019; Popuri et al., 2018). However, there is some recent evidence that converters (individuals who transition from an asymptomatic to symptomatic status during a study) do demonstrate smaller thalami than non-carrier controls (Jiskoot et al., 2019). Thalamic atrophy may, therefore, constitute a marker for proximity to symptom onset in presymptomatic *GRN* individuals. Finally, the symptomatic *GRN* group presented with significantly smaller whole-brain and subcortical regions (predominantly left) across all areas. The largest difference was in the left caudate (15.3% smaller than controls), which is in line with a consensus of caudate atrophy being predominantly associated with the *GRN* mutations (Cash et al., 2018; Premi et al., 2014).

6.4.2 Longitudinal results

6.4.2.1 MAPT

The current results demonstrated significantly increased annual rates of change in the left hippocampus (STEPS), left amygdala (BSI) and right caudate (STEPS) in presymptomatic *MAPT* carriers. Longitudinal investigations in presymptomatic genetic FTD are limited but a recent longitudinal VBM analysis also found presymptomatic reductions in GM density in the hippocampus as their only longitudinal finding (Panman et al., 2019). There are currently no published subcortical rates of change or sample size estimates in this group

for comparison. However, the current study found these significantly increased annual rates of subcortical change did not translate into feasible sample sizes in the presymptomatic *MAPT* cohort. The results of the symptomatic *MAPT* group suggest that whole-brain measures perform well and could be complemented with the addition of bilateral caudate atrophy biomarkers. The hippocampal and amygdala longitudinal measures failed to provide reasonable sample size estimates in the symptomatic group. This is most likely because they were already severely atrophied at baseline in the current cohort with approximately a third of the structures already lost at presentation (Table 6-1). The mean disease duration for this group was 5.8 years and given the early involvement of these regions they may only prove useful with newly symptomatic individuals. In contrast, the caudate appears to become involved later in the disease process with baseline volumes not significantly differing from non-carrier controls at presentation. Based on the longitudinal data, atrophy in this region subsequently accelerates after symptom onset and may, therefore, provide a more sensitive marker in symptomatic *MAPT* trials.

6.4.2.2 C9orf72

Results for the presymptomatic *C9orf72* carriers demonstrate significant widespread atrophy, both cross-sectionally and longitudinally for several subcortical structures (Table 6-2, Table 6-4 and Figure 6-1). Longitudinally, the left HBSI demonstrated significantly increased rates of atrophy in the early presymptomatic group, with significantly increased bilateral thalamic atrophy in the late presymptomatic group. There are currently no published longitudinal rates of change or sample size calculations of subcortical atrophy in this population for comparison; however, the current sample size calculations for the late presymptomatic groups show some promise for the left TBSI and right STEPS thalamus volume difference with 331 [142–1372] and 367 [153–53339] presymptomatic carriers required per treatment arm to detect a 30% reduction in these early rates of change. As previously mentioned, the confidence intervals provide crucial information alongside the point estimate and need to be strongly considered with regard to biomarker choice, suggesting that the right STEPS thalamus longitudinal measure may not be suitable. It is important

to note that these calculations were based on the data of only 14 presymptomatic individuals. Increasing these numbers may assist in producing a more reliable estimate with additional data points, lower the point estimate and narrow the confidence intervals for these potential markers of preclinical change and make the results more generalisable. Overall, the initial early indicators of the disease process evident in the presymptomatic cross-sectional comparisons appear to evolve slowly over several decades and thus did not translate to longitudinal markers able to detect significant change over a year. In the symptomatic *C9orf72* cohort, the results suggest that a whole-brain measure accompanied by the bilateral CBSI measures could perform well at detecting a positive treatment effect given the substantially slower rates of annual change in the other subcortical structures.

6.4.2.3 GRN

Finally, the current longitudinal results suggest that several measures show promise within the *GRN* population. In the late presymptomatic group, annual BSI atrophy rates were significantly higher for bilateral hippocampus, bilateral caudate and right thalamus than for non-carrier controls. The associated sample size calculations were still relatively high for these measures, with the left HBSI producing the lowest estimate of 583 presymptomatic individuals required per treatment arm. Despite this, it does demonstrate the utility of longitudinal volumetric measures in detecting increased rates subcortical atrophy in presymptomatic carriers compared with non-carrier controls up to 10 years prior to expected symptom onset. Significantly increased longitudinal rates of change in the amygdala, hippocampus, caudate and thalamus have not previously been published and provide valuable insights into the presymptomatic subcortical evolution of the disease within a single year for *GRN* carriers. As discussed in the previous chapter (Section 6.4), the pattern of subnuclei involvement for the amygdala, hippocampus and thalamus has been shown to vary depending on the genes involved (Bocchetta, Iglesias et al., 2018; Bocchetta, Iglesias, Cash., 2019; Bocchetta et al., 2020). Employing a more targeted subnuclei segmentation technique may provide a more sensitive marker of presymptomatic change and more feasible sample size

estimates. The utility of these subnuclei markers has yet to be assessed longitudinally and at the presymptomatic stages and could be of great value in future research modelling longitudinal presymptomatic neuroanatomical change.

In the symptomatic *GRN* group, both direct and indirect whole-brain measures produced small sample sizes under 100 in line with longitudinal studies suggesting that the highest rate of global atrophy is evident in *GRN* mutation patients (Jiskoot et al., 2019; Whitwell, Weigand et al., 2011; Whitwell et al., 2015). Measurable GM atrophy of the left middle frontal gyrus has been observed over just eight weeks (Sha et al., 2017). Subcortical longitudinal rates were high and associated sample size estimates particularly low for the left TBSI, left CBSI, right CBSI and left HBSI with 63, 84, 114 and 142 individuals required per treatment arm, respectively. This suggests that combining a global measure with several subcortical measures may prove useful for capturing treatment effects of disease-modifying interventions in symptomatic *GRN* trials.

The current study has several limitations to consider. First, correctly estimating years from symptom onset is challenging. Age at symptom onset is variable in genetic FTD, with intrafamilial variability of at least a decade in some families (Greaves & Rohrer, 2019). Therefore, it is possible that individuals within the same presymptomatic (i.e. early or late) subgroup were at considerably different preclinical stages despite a similar EYO, introducing additional variability and error. Recent work investigating an extensive cohort of several thousand genetic FTD patients with either a *MAPT*, *C9orf72* or *GRN* mutation demonstrated that only the *MAPT* mutation group had a strong correlation between their age at symptom onset and their parental and mean family ages at symptom onset. This relationship was less evident in the *GRN* and *C9orf72* genetic groups, making it likely that the estimates are less reliable in these presymptomatic subgroups (Moore et al., 2020). Despite this, some estimate is required to begin to stage presymptomatic individuals to investigate changes at the proximal stage of symptom onset. EYO, coupled with changes that may indicate conversion, such as the thalamic atrophy in the *GRN* presymptomatic

group suggested here, could provide valuable prognostic information for mutation carriers and their families.

It is also worth noting that, although the full genetic cohort included a considerable number of individuals, the distribution was not even amongst the subgroups. This resulted in relatively small numbers for the *MAPT* subgroups, particularly the symptomatic group (n = 11) as well as the late presymptomatic *C9orf72* group (n = 15). It is, therefore, possible that certain patterns of atrophy as measured by the automated methods fell below the detectable threshold providing only part of the overall picture of neuroanatomical change for these individuals. Results from these smaller cohorts also mean that some caution is required when generalising to the full population. As additional GENFI participants return for longitudinal visits, these numbers will increase; revisiting these analyses in the future with larger cohorts will be of value in building on the insights provided by the current results.

An additional consideration when interpreting the *GRN* results is that I did not calculate an asymmetry index for the presymptomatic and symptomatic individuals and flip any images accordingly so that the most affected hemispheres were aligned. Previous longitudinal analysis of *GRN* symptomatic patients has demonstrated asymmetric thalamic atrophy only once images were flipped to match the most affected hemispheres (Rohrer et al., 2010). This could be particularly interesting in the late presymptomatic group where the bilateral pattern of significantly increased subcortical atrophy may reflect a combination of individuals with left- and right-dominant patterns of atrophy and not an early bilateral stage of atrophy before asymmetry becomes evident during the symptomatic stage. If this is the case, aligning the dominant sides would make the longitudinal signal stronger and potentially result in more feasible sample size calculations for the presymptomatic individuals. This would form a valuable next phase of analysis for the longitudinal presymptomatic *GRN* GENFI cohort.

Additional future work that could prove valuable is investigating PieBSI and longitudinal lobar rates of change in this genetic FTD cohort. Previous research has demonstrated that all three presymptomatic carrier subgroups demonstrated significant decreases in the anterior insula suggesting a potential for a shared presymptomatic biomarker (Cash et al., 2018). In addition, many previous findings in symptomatic individuals have reported results in terms of lobes affected for each of the genetic groups and this additional analysis will be useful to confirm and provide a longitudinal extension to these findings. Whilst each genetic mutation group exhibits unique areas of atrophy, they have also demonstrated reduced volumes in a shared network involving the insula, orbitofrontal lobe, and anterior cingulate (Cash et al., 2018). Investigating these regional changes at the lobar level, may provide valuable insights into the differential and shared patterns of regional atrophy in genetic FTD, and provide information about the potential of these measures as non-invasive markers for each of the genetic groups both symptomatically and presymptomatically.

And finally, it is important to note that these sample sizes are for a year long intervention trial. By including more timepoints over a longer period, the between subject variability and the within-subject measurement error would be more reliably modelled. This is because the longer participants are followed-up the smaller the contribution of the within subject noise and the higher the contribution of the between subject variability is for a measure. Calculations for a longer trial based on this model would reduce the current sample size estimates given the improved signal relative to noise for these longitudinal atrophy measures.

The work presented in this chapter confirms much of the current cross-sectional literature and highlights future work that may prove useful in untangling some of the remaining inconsistencies. More importantly, it contributes to the currently limited investigations of longitudinal change, in particular information about longitudinal subcortical atrophy in symptomatic and presymptomatic genetic FTD. Fully automated measures of whole-brain and several measures of subcortical longitudinal change demonstrate promise

as potential non-invasive biomarkers for detecting positive effects of disease-modifying therapies in all symptomatic genetic groups. This was consistently true for measures of caudate atrophy, which have previously not been investigated longitudinally in genetic FTD and could be valuable in upcoming symptomatic treatment trials. Finally, presymptomatic trials would likely employ mixed-effects modelling of multiple biomarkers of early change, including structural imaging outcomes alongside other imaging modalities and fluid markers. I believe the current data demonstrate the potential of automated measures in capturing presymptomatic subcortical change and am optimistic that applying improved subnuclei measures and future work investigating lobar atrophy in this population could be of value in determining the added utility of automated longitudinal measures of structural change for early intervention trials in genetic FTD.

7 General Conclusions

This final chapter will provide a discussion of the key findings presented in the thesis, address the question of whether global or more focal measures of volumetric change demonstrate the most potential as markers in FTD and address the limitations and potential avenues for future investigations arising from this work.

7.1 Summary of findings

The work presented in this thesis provides a comprehensive investigation of multiple global, lobar and subcortical measures of atrophy across the varied clinical, genetic and pathology-confirmed FTD spectrum with the aim of assessing their utility as potential biomarkers in FTD disease-modifying trials. As increasing numbers of promising candidate interventions become available for therapeutic trials in FTD, the need for validated non-invasive biomarkers becomes ever more pressing. The work presented here has demonstrated the clear utility of multiple fully automated measures of longitudinal change in large well-characterised sporadic and genetic FTD cohorts. More specifically, direct BSI measures of different brain regions produced high effect sizes and feasible sample size estimates, with well below 100 individuals required per treatment arm in trials to detect a disease-modifying effect. The current work further highlights that the most promising measures vary in important ways depending on the patient subgroup to be enrolled (Table 7-1).

Taken together, the results presented here contribute valuable insights to the currently limited body of longitudinal investigations across FTD and are crucial to inform the design of clinical trials. In particular, the thesis addresses the lack of published rates of change and sample size estimate data as well as demonstrating the clear need for inclusion of confidence intervals in any sample size reporting, so as to facilitate properly informed biomarker choice and analysis for these trials. Additional key contributions of this work include: expanding on current knowledge concerning subcortical atrophy profiles and rates of progression across the FTD spectrum; providing the first neuroimaging

investigation of the global, lobar and subcortical presentation and progression of PPA-NOS patients and investigating the feasibility of measuring longitudinal annual volumetric change in presymptomatic mutation carriers.

7.1.1 Comparison of automated measures of whole-brain atrophy in FTD

In Chapter 3, I investigated the comparative performance of six widely used fully automated segmentation techniques and associated indirect and direct measures of longitudinal whole-brain atrophy across 10 clinical, genetic and pathology-confirmed subgroups. The size and scope of the longitudinal cohort collated for this thesis and the ability to perform a direct head-to-head comparison of techniques was a key strength of this study, aimed at addressing inconsistencies in previously published rates of change across FTD. Segmentation accuracy was good cross-sectionally (with a few noted exceptions); however, the longitudinal analyses revealed considerable variability in the accuracy of the whole-brain atrophy measurement. This was particularly evident from the direct paired comparison of effect size ratios, demonstrating significantly different sample size estimates required to detect a meaningful disease-modifying effect depending on the technique utilised. As expected from the literature, direct measures of atrophy almost universally outperformed indirect measures of volume change. For a trial enrolling any of the patient subgroups, application of the BSI to the BMAPS, GIF or SPM segmentations overall produced the lowest sample size estimates. All of these three pipelines worked without additional troubleshooting and would, therefore, be most appropriate for application in large multicentre trials where levels of imaging expertise may vary. These findings support the first three hypotheses of the thesis outlined in Section 1.11.

Hypothesis 1: Cross-sectionally, widely used fully automated whole-brain segmentation pipelines will perform well at delineating brain from non-brain

regions across the varied FTD spectrum using structural imaging as evaluated by comprehensive visual assessment of segmentation accuracy.

Hypothesis 2: Longitudinally, indirect volume-difference measures derived from these automated segmentations will reveal the impact of segmentation errors by producing significantly different effect size and sample size calculations based on the mean and standard deviations of the annual rate of global atrophy these methods produce in head-to-head paired comparisons.

Hypothesis 3: Direct methods of measuring whole-brain atrophy will provide more robust biomarkers of longitudinal change across the FTD spectrum than the indirect volume-difference measures, resulting in significantly lower sample sizes required to detect a putative treatment effect.

7.1.2 Automated measures of lobar atrophy across the FTD spectrum

Applying the best segmentation and direct longitudinal measures from the whole-brain investigations, the findings I present in Chapter 4 are largely in line with the current literature on differential distribution and rates of lobar atrophy across the clinical, genetic and pathology-confirmed FTD subgroups (Binney et al., 2017; Cash et al., 2018; Jiskoot et al., 2018; Mahoney et al., 2012; Montembeault et al., 2018; Rohrer et al., 2008, 2009; Schroeter et al., 2007; Staffaroni et al., 2019; Whitwell, Boeve et al., 2015). Regional atrophy was successfully measured across all lobar regions, with a clear anterior–posterior gradient and differing profiles across the subgroups largely in line with the specific hypotheses outlined in 4.1. Again, the direct BSI method provided the best overall measure of lobar atrophy. Large effect sizes and small sample sizes under 100 were reported for many of the PieBSI lobar regions, which produced largely equivalent or improved estimates compared with previously published results (Binney et al., 2017; Pankov et al., 2016; Rohrer et al., 2008; Staffaroni et al., 2019; Whitwell, Boeve et al., 2015). The temporal lobe PieBSI performed particularly well across all clinical, genetic and pathology-confirmed

FTD subgroups and would thus be a good candidate biomarker for consideration in the upcoming FTD trials.

7.1.3 Automated measures of subcortical atrophy in FTD

Chapter 5 presents results for the first longitudinal investigation into subcortical changes in the amygdala, hippocampus, caudate and thalamus across the full FTD spectrum. Automated segmentation of subcortical structures has several challenges due in part to the difficulty in delineating boundaries between neighbouring subcortical GM structures, such as the amygdala and hippocampus, as well as the wide neuroanatomical variability of shape and size, particularly with neurodegenerative populations at differing stages of the disease. Despite this, the GIF and STEPS techniques produced high-quality segmentations across all eight regions even in the presence of significant atrophy and poor GM/WM contrast across the MR image. The key longitudinal findings were significantly increased rates of atrophy detected for all eight subcortical regions compared with controls in varied patterns across the subgroups, largely in line with those predicted and discussed in 5.1 and 5.4. Application of the BSI to the amygdala and thalamus has not previously been done and the groundwork completed in this chapter to ensure it runs successfully will be of value in future studies. Of particular interest was the strong signal found for caudate atrophy across all clinical, genetic and pathology-confirmed FTD subgroups. Except for Landin and colleagues' (2017) study of 20 bvFTD patients, caudate atrophy has not previously been investigated longitudinally in any of the FTD subgroups and demonstrates substantial promise as a potential marker for FTD trials. In addition, the current results support the use of longitudinal automated measures of thalamic atrophy across multiple subgroups in upcoming intervention trials.

7.1.4 Longitudinal investigations of automated whole-brain and subcortical atrophy in symptomatic and presymptomatic genetic FTD

Finally, Chapter 6 provided an in-depth longitudinal investigation of global and subcortical rates of volume change in a large genetic FTD cohort. This allowed

for the investigation of longitudinal rates of change in presymptomatic individuals at more distant and proximal stages to symptom onset and the opportunity to confirm the genetic subgroup findings from the previous chapters in a symptomatic cohort who had undergone more uniform image acquisition in GENFI. Increased global and subcortical rates of change in the symptomatic *MAPT*, *C9orf72* and *GRN* subgroups were found universally across all direct BSI measures. The caudate and thalamus BSI results again demonstrated the most promise in tracking disease progression and potential detection of treatment effects.

Longitudinal investigations of volume change in presymptomatic mutation carriers are currently lacking and urgently required as we move towards presymptomatic interventions in genetic FTD. In fact, at the time of writing, only a single study had reported longitudinal results comparing structural imaging across the three main presymptomatic carriers with non-carrier controls, using a VBM and cortical thickness analysis (Panman et al., 2019). Caroppo et al. (2015) and Olm et al. (2018) are the only other longitudinal structural MRI studies to investigate volume change in a presymptomatic FTD cohort, both focussing only on *GRN* carriers. Together, these studies report some GM reduction for discrete regions within the orbitofrontal, temporal and insular cortices, but did not report any rates of change or find significant whole-brain or subcortical atrophy in *MAPT*, *C9orf72* or *GRN* carriers. This is except for Panman et al. (2019) reporting significant GM density loss of the right hippocampus in *MAPT* using VBM ($p < 0.05$).

In contrast, the results from the current investigation contribute several key findings to this previously limited picture of presymptomatic neuroanatomical change. In *MAPT*, significantly increased annual rates of change in the left hippocampus (STEPS), left amygdala (BSI) and right caudate (STEPS) were observed presymptomatically. Results for the *C9orf72* carriers demonstrate significant presymptomatic atrophy in the left hippocampus and bilateral thalami, with the increased left HBSI rates evident more than 10 years before symptom onset. Finally, several measures showed promise within the

presymptomatic *GRN* group up to 10 years before symptom onset. Whole-brain, left amygdala, bilateral hippocampus, bilateral caudate and right thalamic rates of atrophy as measures by the BSI were significantly faster in carriers than in non-carrier controls.

Significantly increased longitudinal rates of change in the amygdala, hippocampus, caudate and thalamus have not previously been published and provide valuable insights into the presymptomatic subcortical evolution of FTD within a single year across the different genetic groups. These results provide a measure of support for the hypothesis below, with direct measures of whole-brain and subcortical measures shown to perform best in the symptomatic population showing promise of detecting volume change presymptomatically. Future work including additional structures and direct comparisons between measures would be of value in building on the current contributions.

Hypothesis 6: Measures of caudate, thalamus and hippocampal atrophy, which showed the most promise as biomarkers in the FTD symptomatic subgroups, will demonstrate potential for detecting structural change in the preclinical stage, by producing significantly higher longitudinal rates of atrophy in presymptomatic mutation carriers compared with non-carrier controls.

7.2 Biomarker choice in FTD: global or focal?

The results presented in Chapters 3–5 provide useful insights into the relative utility of global versus more focal automated measures of atrophy as potential non-invasive biomarkers across the FTD spectrum, aimed at addressing the following two thesis hypotheses.

Hypothesis 4: Adaptations of these direct global longitudinal measures will yield low sample size estimates below 100 participants when applied to more focal regional volumetric measures relevant to FTD (i.e. frontal, temporal, insular, cingulate and subcortical regions such as the thalamus, caudate)

providing additional valuable biomarker options for upcoming therapeutic interventions.

Table 7-1 provides a summary of the sample size estimates for the best-performing global, lobar and subcortical regions for comparison. Where available, the direct BSI measure was used given its improved performance over indirect measures. The current results provide some support for the predictions put forward in this hypothesis. Direct measures of particularly temporal lobar, as well as frontal and parietal atrophy performed well across multiple FTD subgroups providing small feasible sample sizes below 100 using the BSI. In contrast, the insular and cingulate longitudinal measures were derived from the indirect volume-difference calculation and were therefore subject to a greater degree of segmentation error with the resulting increase in the sample size estimate. In addition, the performance of the subcortical measurements varied, although the caudate and thalamus did consistently produce sample sizes below 100 across many of the FTD subgroups. One reason for this relates to the challenges faced in accurately delineating smaller structures, particularly in subcortical regions where labelling neighbouring GM structures is difficult because there is limited contrast and intensity information available to differentiate the boundaries of such structures. In addition, there is the issue of the substantial difference in shape and size, particularly in patient populations as well as the disproportionate effect of minor segmentation errors relative to the overall smaller volumes of these structures. Despite this, multiple lobar and subcortical longitudinal measures showed equivalent performance to the more established global measures in terms of reliably tracking regional volume change, which produced large effect sizes and small sample size estimates providing additional valuable biomarker options for upcoming therapeutic interventions. Chapters 4 and 5 highlighted several additional analyses that may provide improvements on the indirect insular and cingulate cortex results with the application of direct longitudinal techniques and improve those for the hippocampus and thalamus by including subnuclei investigations. Overall, these findings suggest that a combination of an established global whole-brain measure with the addition of several promising lobar and

subcortical measures may provide the most complementary outcomes in symptomatic trials.

Hypothesis 5: Due to different neuroanatomical presentations across the FTD spectrum, the most robust regional measures for detecting volumetric change and resulting lowest sample sizes will differ between the patient subgroups reflecting the different disease processes underlying these presentations.

The work presented in this thesis supported the differential neuroanatomical profiles of atrophy across the subgroups at all three levels of volumetric analyses. The individual chapters described these different presentations in each clinical, genetic and pathology-confirmed FTD subgroup in detail. Overall, the documented heterogeneity and more distributed profile of atrophy resulted in the bvFTD, *C9orf72* and TDP-43 subgroups producing the highest sample size estimates. PPA, *MAPT* and Tau subgroups produced rates of change with a higher mean relative to standard deviation resulting in more favourable sample size estimates for the global, lobar and subcortical structures, often demonstrating an asymmetric left-sided dominance. This asymmetry was particularly evident in the PPA-NOS and *GRN* subgroups. The pattern of best-performing regional measures for each of the subgroups can be visualized based on the colour coding in Table 7-1. Bright green represents measures that perform particularly well at measuring atrophy in the different subgroups with estimates generally well below 100 participants required per arm to detect a disease-modifying treatment effect of reducing atrophy by 30% (80% power, 5% significance). Those highlighted in light green also show substantial promise as imaging biomarkers with point estimates of 100–200 and BCa CIs that often overlap with those producing estimates <100. Those in orange require 200–300 individuals and denote measures that show some potential that with a larger cohort and improved measurement may prove useful in future trials; however, the orange-designated results should be regarded with some caution. The results for measures highlighted in red suggest the limited utility of these measures to detect disease progression and more importantly a

treatment effect on these rates of change. Their point estimates were 300 and above, with many providing unfeasibly large or infinite upper 95% BCa CI limits.

Overall the results from this thesis support the use of an established global measure of whole-brain atrophy in combination with the best-performing targeted regional measure (i.e. temporal lobe or caudate atrophy) for each particular subgroup, as has been done in AD, MS and HD trials.

Table 7-1 Sample size estimates for the global, lobar and subcortical longitudinal measures across the FTD subgroups. The whole-brain measure is derived from the GIF GBSI results, the lobar is from the PieBSI with the addition of the insula and cingulate results from the STEPS volume difference measure and the subcortical results reflect those derived from the application of the local BSI on the GIF and STEPS subcortical segmentations *

Measure		Whole-brain (GIF GBSI)	PieBSI									STEPS volume difference			
			Frontal left	Frontal right	Temporal left	Temporal right	Parietal left	Parietal right	Occipital left	Occipital right	Infratentorial	Insula left	Insula right	Cingulate left	Cingulate right
Clinical subgroups	bvFTD	127	245	227	161	212	283	327	307	360	270	1028	1614	1256	897
	svPPA	35	120	193	23	43	60	143	135	370	730	157	664	243	307
	nfvPPA	58	93	100	76	100	113	88	176	221	183	194	471	346	949
	PPA-NOS	31	74	63	37	72	146	238	56	185	135	51	229	60	369
	lvPPA	45	116	196	37	69	63	93	140	218	484	291	1308	1366	2343
Genetic subgroups	*MAPT*	55	72	84	68	83	548	376	552	523	72	329	495	242	288
	C9orf72	152	239	175	202	261	366	423	509	1249	589	184	538	3623	452
	GRN	48	127	213	109	285	128	223	53	150	297	137	10430	985	10250
Pathology subgroup	Tau	50	85	67	79	87	287	199	547	676	77	298	2343	780	236
	TDP-43	76	174	209	105	125	215	284	134	282	468	143	786	797	907

* Cells highlighted bright green provide sample size estimates < 100, light green report estimates of 100–200, yellow highlight measures requiring 201–300 individuals per treatment arm and any estimate over 300 is highlighted red.

Table 7-1 continued: Sample size estimates for the global, lobar and subcortical longitudinal measures across the FTD subgroups. The whole-brain measure is derived from the GIF GBSI results, the lobar is from the PieBSI with the addition of the insula and cingulate results from the STEPS volume difference measure and the subcortical results reflect those derived from the application of the local BSI on the GIF and STEPS subcortical segmentations *

Measure		Whole brain (GIF GBSI)	Subcortical BSI measures							
			ABSI left	ABSI right	HBSI left	HBSI right	CBSI left	CBSI right	TBSI left	TBSI right
Clinical subgroups	bvFTD	127	294	454	477	640	337	302	306	286
	svPPA	35	109	78	105	272	47	128	57	80
	nfvPPA	58	177	274	221	218	167	136	161	127
	PPA-NOS	31	27	433	12	210	35	70	56	369
	lvPPA	45	289	533	200	2223	50	71	66	242
Genetic subgroups	*MAPT*	55	175	132	448	216	80	202	80	169
	C9orf72	152	590	576	6607	833	370	135	510	1058
	GRN	48	65	4.10E+05	115	51915	118	332	203	1160
Pathology subgroups	Tau	50	117	124	321	197	107	157	69	112
	TDP-43	76	277	315	212	346	198	160	348	347

* Cells highlighted bright green provide sample size estimates < 100, light green report estimates of 100–200, yellow highlight measures requiring 201–300 individuals per treatment arm and any estimate over 300 is highlighted red.

7.3 Limitations

The current findings are subject to several caveats. First, whilst the full longitudinal sporadic and genetic cohorts were large, some of the individual subgroups were relatively small and potentially underpowered to detect the full picture of volumetric change. Patterns of change at the individual level may not be well captured in the group as can be seen from Figure 4-3 and Figure 4-4 for the PPA-NOS (n = 7) and *GRN* (n = 8) subgroups. As discussed, this may have also affected the efficacy of the bootstrapping methodology for accurate estimation of the BCa CIs in these cases. Therefore, some caution is advised in interpreting these results.

Second, the main direct lobar measure of volumetric change (PieBSI) is a recently developed technique that will need validation of its reproducibility and reliability in another longitudinal cohort with atrophied and healthy control images. However, based on the qualitative and quantitative review of the PieBSI segmentations and longitudinal results, it was shown to demonstrate substantial promise based on the results presented here.

Third, because of the inevitable time constraints of a PhD, it was not feasible to visually assess all 5426 subcortical STEPS segmentations or 3357 subcortical registrations for the retrospective (Chapter 5) and genetic cohorts (Chapter 6) combined. To account for this, substantial range-checking was performed and all longitudinal data points that fell outside a reasonable upper and lower interquartile range or demonstrated a substantial difference between the direct and indirect measures of the same structure were visually assessed for errors. A random selection of STEPS data points was also reviewed, as were all the 1288 amygdala GIF regions as part of the MNI image and GIF mask transformation checks. Despite the targeted QC checks, it is possible that some data points were included that would have been failed and excluded had the full dataset been reviewed, introducing potential error. Ideally, all automated segmentations should be visually assessed by a trained analyst to ensure the pipeline was completed

successfully and accurately delineated the structure of interest. This process would be highly recommended in any trial image-analysis plan.

Fourth, age at symptom onset is variable within genetic FTD (Greaves & Rohrer, 2019; Moore et al., 2020), the distinction between presymptomatic and symptomatic is difficult to define and improved ways of describing and predicting the conversion from one to the other is an active area of research in FTD. Therefore, using the estimated years from onset to group the presymptomatic carriers into early and late subgroups may mean participants were not well matched in terms of preclinical disease stage within these groups, introducing additional variability and error. As these participants become symptomatic in the coming years, it will be valuable to revisit and match their presymptomatic imaging time-points to confirm and extend the current findings. In addition to structural imaging, multiple biomarkers including other imaging modalities, fluid markers (such as neurofilament levels (NFL)) and a combination of cognitive and clinical scores will be needed to better identify the presymptomatic stages for upcoming genetic FTD trials.

Finally, asymmetry was not directly investigated in either sporadic or genetic cohorts, but analyses for the lobes and subcortical regions were split by hemisphere and investigated independently. An asymmetry index calculation in itself would provide a useful future investigation in the FTD subgroups in these cohorts, but the omission of it may have disproportionately affected the *GRN* results. In the late presymptomatic *GRN* group, the bilateral pattern of significantly increased subcortical atrophy may reflect both left- and right-dominant patterns of atrophy, potentially diluting the signal of change that would have be obtained if images were flipped to have the most affected hemisphere on the same side. By matching the most affected hemisphere, a stronger signal of change and more feasible sample size calculations for the presymptomatic individuals may be possible. This could form a valuable future phase of analysis within the longitudinal presymptomatic *GRN* GENFI cohort.

7.4 Future work

There are several additional potential avenues to build on the current results to improve non-invasive imaging markers that may better support determining the efficacy of upcoming disease-modifying therapies in FTD.

First and foremost, replication of the current novel longitudinal findings in an independent large well-characterised FTD cohort will be of great value. This will be particularly important to confirm the significant results of increased rates and specific patterns of subcortical atrophy found across the different symptomatic patient groups as well as the presymptomatic carriers that have previously not been reported. In addition, as the PPA-NOS subgroup becomes more established in the literature and better recognised in the clinic, it will be important to confirm and extend the current first investigations reported here demonstrating the fast progressing, homogeneous, asymmetric longitudinal profile of volumetric change reported in this population.

The subcortical atrophy measures performed well across most symptomatic subgroups and showed higher rates in the presymptomatic carriers than in non-carriers. However, for the presymptomatic groups, these significantly increased rates of change did not translate to feasible sample size estimates. One potential improvement to increase the sensitivity of these subcortical measures would be to focus on the particular subnuclei recently shown to demonstrate differential patterns of involvement across the clinical, genetic and pathology-confirmed individuals. Substantial reductions (16–33%) were reported cross-sectionally in the mediodorsal nucleus across all FTD subgroups compared with controls. In addition, the laterodorsal nucleus was particularly affected in the three genetic groups (28–38% smaller than controls) and the pulvinar only affected in the *C9orf72* group (16% smaller) (Bocchetta et al., 2020), suggesting these may provide a more sensitive marker presymptomatically in the genetic FTD cohort.

An additional avenue that could prove valuable for building on the current findings would be the application of the PieBSI and longitudinal lobar rates of change in genetic FTD, particularly in the presymptomatic cohort. The previous longitudinal investigations in presymptomatic mutation carriers found changes within the orbitofrontal, temporal and insular cortices (Caroppo et al., 2015; Olm et al., 2018; Panman et al., 2019). In addition, cross-sectional investigations have shown that genetic groups exhibit unique areas of atrophy but they also demonstrated reduced volumes in a shared network involving the insula, orbitofrontal lobe and anterior cingulate (Cash et al., 2018). The current study investigated rates of insular and cingulate atrophy using an indirect volume-difference measure derived from the GIF parcellations. There are several ways this may be improved. First, the GIF parcellation provides an anterior and posterior division for the insula and an anterior, middle and posterior division for the cingulate. Particularly in the cingulate, focussing on the anterior portion of the segmentation may provide a better signal in this structure, which traverses a substantial portion of the medial surface of the brain. In addition, using these segmentations to derive a direct measure of change is likely to provide a more robust measure as demonstrated repeatedly in this thesis. Application of the double window BSI (Hobbs et al., 2009) to these regions would provide data consistent with the subcortical regional analyses. In addition, applying the GIF segmentation algorithm to the mid-point average longitudinal SPM output images and extracting the deformation field information for those discrete regions could also provide promising robust measures of change for all or part of the insula and cingulate. This future work would be valuable for added insights into the neuroanatomical progression across the symptomatic and presymptomatic FTD spectrum and provide additional data concerning the utility of automated techniques as potential biomarkers to detect disease-modifying treatment effects.

Finally, the current results were based on rates calculated from only two images and calculated to detect a treatment effect over a single year. Future work including multiple timepoints over a longer period would be valuable to

build on the current work as additional imaging datapoints assist in improving the ability to separate noise related to within subject variability from the detection of between subject variability and to model the measurement error related to the technique being employed to track these imaging changes. Longer follow-up with multiple timepoints could prove particularly helpful to develop on the current results in the presymptomatic subgroups.

In summary, the studies described in this thesis have contributed substantially to the current knowledge of longitudinal neuroanatomical rates of change across the full FTD spectrum and confirmed the potential of fully automated measures of volumetric change across global, lobar and subcortical structures. These data have important implications for non-invasive biomarker choice in clinical trials for putative therapies in sporadic and genetic FTD.

8 References

Agosta, F., Canu, E., Sarro, L., Comi, G., & Filippi, M. (2012). Neuroimaging findings in frontotemporal lobar degeneration spectrum of disorders. *Cortex*, *48*(4), 389–413. https://doi.org/10.1016/j.cortex.2011.04.012

Anderson, V. M., Schott, J. M., Bartlett, J. W., Leung, K. K., Miller, D. H., & Fox, N. C. (2012). Gray matter atrophy rate as a marker of disease progression in AD. *Neurobiology of Aging*, *33*(7), 1194–1202. https://doi.org/10.1016/j.neurobiolaging.2010.11.001

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Barnes, J., Whitwell, J. L., Frost, C., Josephs, K. A., Rossor, M., & Fox, N. C. (2006). Measurements of the amygdala and hippocampus in pathologically confirmed Alzheimer disease and frontotemporal lobar degeneration. *Archives of Neurology*, *63*(10), 1434–1439. https://doi.org/10.1001/archneur.63.10.1434

Bateman, R. J., Xiong, C., Benzinger, T. L. S., Fagan, A. M., Goate, A., Fox, N. C., … Morris, J. C. (2012). Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *New England Journal of Medicine*, *367*(9), 795–804. https://doi.org/10.1056/NEJMoa1202753

Battaglini, M., Jenkinson, M., & De Stefano, N. (2018). SIENA-XL for improving the assessment of gray and white matter volume changes on brain MRI. *Human Brain Mapping*, *39*(3), 1063–1077. https://doi.org/10.1002/hbm.23828

Beach, T. G., Monsell, S. E., Phillips, L. E., & Kukull, W. (2012). Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. *Journal of Neuropathology and Experimental Neurology*, *71*(4), 266–273. https://doi.org/10.1097/NEN.0b013e31824b211b

Beber, B. C., Kochhann, R., Da Silva, B. M., & Chaves, M. L. F. (2014). Logophenic aphasia or Altzheimer's disease: Different phases of the same disease? *Dementia & Neuropsychologia*, *8*(3), 302–307. https://doi.org/10.1590/S1980-57642014DN83000016

Bertrand, A., Wen, J., Rinaldi, D., Houot, M., Sayah, S., Camuzat, A., … Le Ber, I. (2018). Early cognitive, structural, and microstructural changes in presymptomatic C9orf72 carriers younger than 40 years. *JAMA Neurology*, *75*(2), 236–245. https://doi.org/10.1001/jamaneurol.2017.4266

Bevan-Jones, W. R., Cope, T. E., Jones, P. S., Passamonti, L., Hong, Y. T., Fryer, T., … Rowe, J. B. (2019). In vivo evidence for pre-symptomatic neuroinflammation in a MAPT mutation carrier. *Annals of Clinical and Translational Neurology*, *6*(2), 373–378. https://doi.org/10.1002/acn3.683

Binney, R. J., Pankov, A., Marx, G., He, X., McKenna, F., Staffaroni, A. M., … Rosen, H. J. (2017). Data-driven regions of interest for longitudinal change in three variants of frontotemporal lobar degeneration. *Brain and Behavior*, *7*(4), 1–11. https://doi.org/10.1002/brb3.675

Boccardi, M., Pennanen, C., Laakso, M. P., Testa, C., Geroldi, C., Soininen, H., & Frisoni, G. B. (2002). Amygdaloid atrophy in frontotemporal dementia and Alzheimer's disease. *Neuroscience Letters*, *335*(2), 139–143. https://doi.org/10.1016/S0304-3940(02)01169-2

Bocchetta, M., Cardoso, M. J., Cash, D. M., Ourselin, S., Warren, J. D., & Rohrer, J. D. (2016). Patterns of regional cerebellar atrophy in genetic frontotemporal dementia. *NeuroImage: Clinical*, *11*, 287–290. https://doi.org/10.1016/j.nicl.2016.02.008

Bocchetta, M., Gordon, E., Cardoso, M. J., Modat, M., Ourselin, S., Warren, J. D., & Rohrer, J. D. (2018). Thalamic atrophy in frontotemporal dementia — Not just a C9orf72 problem. *NeuroImage: Clinical*, *18*(February), 675–681. https://doi.org/10.1016/j.nicl.2018.02.019

Bocchetta, M., Gordon, E., Manning, E., Barnes, J., Cash, D. M., Espak, M., … Rohrer, J. D. (2015). Detailed volumetric analysis of the hypothalamus in behavioral variant frontotemporal dementia. *Journal of Neurology*. https://doi.org/10.1007/s00415-015-7885-2

Bocchetta, M., Iglesias, J. E., Cash, D. M., Warren, J. D., & Rohrer, J. D. (2019). Amygdala subnuclei are differentially affected in the different genetic and pathological forms of frontotemporal dementia. *Alzheimer's*

and Dementia: Diagnosis, Assessment and Disease Monitoring, *11*, 136–141. https://doi.org/10.1016/j.dadm.2018.12.006

Bocchetta, M., Iglesias, J. E., Neason, M., Cash, D. M., Warren, J. D., & Rohrer, J. D. (2020). Thalamic nuclei in frontotemporal dementia: Mediodorsal nucleus involvement is universal but pulvinar atrophy is unique to C9orf72. *Human Brain Mapping*, *41*(4), 1006–1016. https://doi.org/10.1002/hbm.24856

Bocchetta, M., Iglesias, J. E., Russell, L. L., Greaves, C. V., Marshall, C. R., Scelsi, M. A., … Rohrer, J. D. (2019). Segmentation of medial temporal subregions reveals early right-sided involvement in semantic variant PPA. *Alzheimer's Research and Therapy*, *11*(1), 1–9. https://doi.org/10.1186/s13195-019-0489-9

Bocchetta, M., Iglesias, J. E., Scelsi, M. A., Cash, D. M., Cardoso, M. J., Modat, M., … Rohrer, J. D. (2018). Hippocampal Subfield Volumetry: Differential Pattern of Atrophy in Different Forms of Genetic Frontotemporal Dementia. *Journal of Alzheimer's Disease*, *64*(2), 497–504. https://doi.org/10.3233/JAD-180195

Borroni, B., Alberici, A., Premi, E., Archetti, S., Garibotto, V., Agosti, C., … Padovani, A. (2008). Brain magnetic resonance imaging structural changes in a pedigree of asymptomatic progranulin mutation carriers. *Rejuvenation Research*, *11*(3), 585–595. https://doi.org/10.1089/rej.2007.0623

Borroni, B., Brambati, S. M., Agosti, C., Gipponi, S., Bellelli, G., Gasparotti, R., … Padovani, A. (2007). Evidence of white matter changes on diffusion tensor imaging in frontotemporal dementia. *Archives of Neurology*, *64*(2), 246–251. https://doi.org/10.1001/archneur.64.2.246

Brambati, S. M., Rankin, K. P., Narvid, J., Seeley, W. W., Dean, D. L., Rosen, H. J., … Gorno-Tempini, M. L. (2009). Atrophy progression in semantic dementia with asymmetric temporal involvement: A tensor-based morphometry study. Retrieved November 12, 2014, from http://ac.els-cdn.com/S0197458007002187/1-s2.0-S0197458007002187-main.pdf?_tid=47b1343e-6a8a-11e4-a99e-00000aacb35f&acdnat=1415810467_eb7f2437b812e7e1c1e2a37fbf1

6201f

Broe, M., Hodges, J. R., Schofield, E., Shepherd, C. E., Kril, J. J., & Halliday, G. M. (2003). Staging disease severity in pathologically confirmed cases of frontotemporal dementia. *Neurology*, *60*(6), 1005–1011. https://doi.org/10.1212/01.WNL.0000052685.09194.39

Cardenas, V. A., Boxer, A. L., Chao, L. L., Gorno-Tempini, M. L., Miller, B. L., Weiner, M. W., & Studholme, C. (2007). Deformation-based morphometry reveals brain atrophy in frontotemporal dementia. *Archives of Neurology*, *64*(6), 873–877. https://doi.org/10.1001/archneur.64.6.873

Cardoso, M. J., Leung, K., Modat, M., Keihaninejad, S., Cash, D., Barnes, J., … Ourselin, S. (2013). STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation. *Medical Image Analysis*, *17*(6), 671–684. https://doi.org/10.1016/j.media.2013.02.006

Cardoso, M. J., Modat, M., Wolz, R., Melbourne, A., Cash, D., Rueckert, D., & Ourselin, S. (2015). Geodesic Information Flows: Spatially-Variant Graphs and Their Application to Segmentation and Fusion. *IEEE Transactions on Medical Imaging*, *34*(9), 1976–1988. https://doi.org/10.1109/TMI.2015.2418298

Caroppo, P., Habert, M. O., Durrleman, S., Funkiewiez, A., Perlbarg, V., Hahn, V., … Le Ber, I. (2015). Lateral Temporal Lobe: An Early Imaging Marker of the Presymptomatic GRN Disease? *Journal of Alzheimer's Disease*, *47*(3), 751–759. https://doi.org/10.3233/JAD-150270

Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, *19*(9), 1141–1164. https://doi.org/10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F

Cash, D. M., Bocchetta, M., Thomas, D. L., Dick, K. M., van Swieten, J. C., Borroni, B., … Rohrer, J. D. (2018). Patterns of gray matter atrophy in genetic frontotemporal dementia: results from the GENFI study. *Neurobiology of Aging*, *62*, 191–196. https://doi.org/10.1016/j.neurobiolaging.2017.10.008

Cash, D. M., Frost, C., Iheme, L. O., Ünay, D., Kandemir, M., Fripp, J., … Ourselin, S. (2015). Assessing atrophy measurement techniques in dementia: Results from the MIRIAD atrophy challenge. *NeuroImage*, *123*, 149–164. https://doi.org/10.1016/j.neuroimage.2015.07.087

Cash, D. M., Rohrer, J. D., Ryan, N. S., Ourselin, S., & Fox, N. C. (2014). Imaging endpoints for clinical trials in Alzheimer's disease. *Alzheimer's Research & Therapy*, *6*(9), 87–97. https://doi.org/10.1186/s13195-014-0087-9

Chan, D, Fox, N. C., Jenkins, R., Scahill, R. I., Crum, W. R., & Rossor, M. N. (2001). Rates of global and regional cerebral atrophy in AD and frontotemporal dementia. *Neurology*, *57*(10), 1756–1763. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11723259

Chan, D, Fox, N., Scahill, R., Crum, W., Whitwell, J., Leschziner, G., … Rossor, M. (2001). Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Annals of Neurology*, *49*(2), 433–442. Retrieved from http://discovery.ucl.ac.uk/79528/

Chan, Dennis, Anderson, V., Pijnenburg, Y., Whitwell, J., Barnes, J., Scahill, R., … Fox, N. C. (2009). The clinical profile of right temporal lobe atrophy. *Brain*, *132*(5), 1287–1298. https://doi.org/10.1093/brain/awp037

Charles, P., Giraudeau, B., Dechartres, A., Baron, G., & Ravaud, P. (2009). Reporting of sample size calculation in randomised controlled trials: Review. *BMJ (Online)*, *338*(7705), 1256. https://doi.org/10.1136/bmj.b1732

Clarkson, M. J., Ourselin, S., Nielsen, C., Leung, K. K., Barnes, J., Whitwell, J. L., … Fox, N. C. (2009). Comparison of phantom and registration scaling corrections using the ADNI cohort. *NeuroImage*, *47*(4), 1506–1513. https://doi.org/10.1016/j.neuroimage.2009.05.045.Comparison

Convery, R., Mead, S., & Rohrer, J. D. (2019). Review: Clinical, genetic and neuroimaging features of frontotemporal dementia. *Neuropathology and Applied Neurobiology*, *45*(1), 6–18. https://doi.org/10.1111/nan.12535

Cosseddu, M., Benussi, A., Gazzina, S., Alberici, A., Dell'Era, V., Manes,

M., … Padovani, A. (2020). Progression of behavioural disturbances in frontotemporal dementia: a longitudinal observational study. *European Journal of Neurology*, *27*(2), 265–272. https://doi.org/10.1111/ene.14071

Cummings, J. L. (2007). The Neuropsychiatric Inventory : Assessing psychopathology in dementia patients. *Neurology*, *48*, S10–S16. https://doi.org/10.1212/WNL.48.5

Cummings, J., Lee, G., Ritter, A., Sabbagh, M., & Zhong, K. (2019). Alzheimer's disease drug development pipeline: 2019. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, *5*, 272–293. https://doi.org/10.1016/j.trci.2019.05.008

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. I. Segmentation and surface reconstruction. *Neuroimage*, *9*, 179–194.

Dopper, E. G. P., Rombouts, S. A. R. B., Jiskoot, L. C., Heijer, T. Den, de Graaf, J. R. A., Koning, I. De, … van Swieten, J. C. (2014). Structural and functional brain connectivity in presymptomatic familial frontotemporal dementia. *Neurology*, *83*(2), 814–823. https://doi.org/10.1212/WNL.0b013e31828407bc

Draper, B., Cations, M., White, F., Trollor, J., Loy, C., Brodaty, H., … Withall, A. (2016). Time to diagnosis in young-onset dementia and its determinants: the INSPIRED study. *International Journal of Geriatric Psychiatry*, *31*(11), 1217–1224. https://doi.org/10.1002/gps.4430

Fellhauer, I., Zöllner, F. G., Schröder, J., Degen, C., Kong, L., Essig, M., … Schad, L. R. (2015). Comparison of automated brain segmentation using a brain phantom and patients with early Alzheimer's dementia or mild cognitive impairment. *Psychiatry Research - Neuroimaging*, *233*(3), 299–305. https://doi.org/10.1016/j.pscychresns.2015.07.011

Fischl, B., Salat, D. H., Busa, E., Albert, M. S., Dieterich, M., Haselgrove, C., … Dale, A. M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33*(3), 341–355. https://doi.org/10.1016/S0896-6273(02)00569-X

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical Surface-Based

Analysis. II: Inflation, flattening, and a surface-cased coordinate system. *Neuroimage*, *9*, 195–207.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., … Dale, A. M. (2004). Automatically Parcellating the Human Cerebral Cortex. *Cortex*, *14*, 11–22. https://doi.org/10.1093/cercor/bhg087

Fletcher, P. D., Downey, L. E., Golden, H. L., Clark, C. N., Slattery, C. F., Paterson, R. W., … Warren, J. D. (2015). Pain and temperature processing in dementia: A clinical and neuroanatomical analysis. *Brain*, *138*(11), 3360–3372. https://doi.org/10.1093/brain/awv276

Floeter, M. K., Bageac, D., Danielian, L. E., Braun, L. E., Traynor, B. J., & Kwan, J. Y. (2016). Longitudinal imaging in C9orf72 mutation carriers: Relationship to phenotype. *NeuroImage: Clinical*, *12*, 1035–1043. https://doi.org/10.1016/j.nicl.2016.10.014

Freeborough, P. A., & Fox, N. C. (1997). The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Transactions on Medical Imaging*, *16*(5), 623–629. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9368118

Freeborough, P. A., Fox, N. C., & Kitney, R. I. (1997). Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. *Computer Methods and Programs in Biomedicine*, *53*(1), 15–25. Retrieved from http://discovery.ucl.ac.uk/143964/

Frings, L., Yew, B., Flanagan, E., Lam, B. Y. K., Hüll, M., Huppertz, H.-J., … Hornberger, M. (2014). Longitudinal grey and white matter changes in frontotemporal dementia and Alzheimer's disease. *PloS One*, *9*(3), e90814. https://doi.org/10.1371/journal.pone.0090814

Garibotto, V., Borroni, B., Agosti, C., Premi, E., Alberici, A., Eickhoff, S. B., … Padovani, A. (2011). Subcortical and deep cortical atrophy in Frontotemporal Lobar Degeneration. *Neurobiology of Aging*, *32*(5), 875–884. https://doi.org/10.1016/j.neurobiolaging.2009.05.004

Ghetti, B., Oblak, A. L., Boeve, B. F., Johnson, K. A., Dickerson, B. C., & Goedert, M. (2015). Invited review: Frontotemporal dementia caused

by *microtubule-associated protein tau* gene (*MAPT*) mutations: a chameleon for neuropathology and neuroimaging. *Neuropathology and Applied Neurobiology*, *41*(1), 24–46. https://doi.org/10.1111/nan.12213

Goedert, M., Eisenberg, D. S., & Crowther, R. A. (2017). Propagation of Tau Aggregates and Neurodegeneration. *Annual Review of Neuroscience*, *40*(1), 189–210. https://doi.org/10.1146/annurev-neuro-072116-031153

González-Villà, S., Oliver, A., Valverde, S., Wang, L., Zwiggelaar, R., & Lladó, X. (2016). A review on brain structures segmentation in magnetic resonance imaging. *Artificial Intelligence in Medicine*. https://doi.org/10.1016/j.artmed.2016.09.001

Gordon, E., Rohrer, J. D., & Fox, N. C. (2016). Advances in neuroimaging in frontotemporal dementia. *Journal of Neurochemistry*, *138*, 193–210. https://doi.org/10.1111/jnc.13656

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. https://doi.org/10.1212/WNL.0b013e3181d1a879

Gorno-tempini, M. L., Dronkers, N. F., Rankin, K. P., Ogar, J. M., Phengrasamy, L., Rosen, H. J., … Miller, B. L. (2004). Cognition and anatomy in three variants of primary progressive aphasia. *Annals of Neurology*, *55*(3), 335–346. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2362399&tool=pmcentrez&rendertype=abstract

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M. F., Cappa, S. F., … Grossman, M. (2011). Classification of primary progressive aphasia and its variants. *Neurology*, *76*, 1006–1014. https://doi.org/10.1212/WNL.0b013e31821103e6

Gorno-Tempini, M. L., Rankin, K. P., Woolley, J. D., Rosen, H. J., Phengrasamy, L., & Miller, B. L. (2004). Cognitive and behavioral profile in a case of right anterior temporal lobe neurodegeneration. *Cortex*, *40*(4–5), 631–644. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S001094520870159X

Graff-Radford, J., Williams, L., Jones, D. T., & Benarroch, E. E. (2017). Caudate nucleus as a component of networks controlling behavior. *Neurology*, *89*(21), 2192–2197. https://doi.org/10.1212/WNL.0000000000004680

Greaves, C. V., & Rohrer, J. D. (2019). An update on genetic frontotemporal dementia. *Journal of Neurology*, *266*(8), 2075–2086. https://doi.org/10.1007/s00415-019-09363-4

Grill, J. D., & Karlawish, J. (2010). Addressing the challenges to successful recruitment and retention in Alzheimer's disease clinical trials. *Alzheimer's Research and Therapy*, *2*(6), 34–45. https://doi.org/10.1186/alzrt58

Grill, J. D., Raman, R., Ernstrom, K., Aisen, P., Dowsett, S. A., Chen, Y. F., … Cummings, J. L. (2015). Comparing recruitment, retention, and safety reporting among geographic regions in multinational Alzheimer's disease clinical trials. *Alzheimer's Research and Therapy*, *7*(1), 1–15. https://doi.org/10.1186/s13195-015-0122-5

Gronenschild, E. H. B. M., Habets, P., Jacobs, H. I. L., Mengelers, R., Rozendaal, N., van Os, J., & Marcelis, M. (2012). The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements. *PLoS ONE*, *7*(6), 1–13. https://doi.org/10.1371/journal.pone.0038234

Grossman, M. (2012). The non-fluent/agrammatic variant of primary progressive aphasia. *The Lancet Neurology*, *11*(6), 545–555. https://doi.org/10.1016/S1474-4422(12)70099-6

Haber, S. N. (2016). Corticostriatal circuitry. *Dialogues in Clinical Neuroscience*, *18*(1), 7–21. https://doi.org/10.1007/978-1-4614-6434-1_135-1

Halabi, C., Halabi, A., Dean, D. L., Wang, P. N., Boxer, A. L., Trojanowski, J. Q., … Seeley, W. W. (2013). Patterns of striatal degeneration in frontotemporal dementia. *Alzheimer Disease and Associated Disorders*, *27*(1), 74–83. https://doi.org/10.1097/WAD.0b013e31824a7df4

Harris, J. M., Gall, C., Jennifer Thompson, C. C., Richardson, A. M., David

Neary, C., du Plessis, D., … Jones, M. (2013). *Classification and pathology of primary progressive aphasia*. *Neurology ®* (Vol. 81). Retrieved from www.neurology.org

Heckemann, R. A., Keihaninejad, S., Aljabar, P., Rueckert, D., Hajnal, J. V., & Hammers, A. (2010). Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. *NeuroImage*, *51*(1), 221–227. https://doi.org/10.1016/j.neuroimage.2010.01.072

Heckemann, R. A., Ledig, C., Gray, K. R., Aljabar, P., Rueckert, D., Hajnal, J. V., & Hammers, A. (2015). Brain extraction using label propagation and group agreement: Pincram. *PLoS ONE*, *10*(7), 1–18. https://doi.org/10.1371/journal.pone.0129211

Henry, M. L., & Gorno-Tempini, M.-L. (2010). The logopenic variant of primary progressive aphasia. *Current Opinion in Neurology*, *23*(6), 633–637. https://doi.org/10.1126/scisignal.2001449.Engineering

Herrero, M. T., Barcia, C., & Navarro, J. M. (2002). Functional anatomy of thalamus and basal ganglia. *Child's Nervous System*, *18*(8), 386–404. https://doi.org/10.1007/s00381-002-0604-1

Hobbs, N. Z., Henley, S. M. D., Wild, E. J., Leung, K. K., Frost, C., Barker, R. A., … Fox, N. C. (2009). Automated quantification of caudate atrophy by local registration of serial MRI: Evaluation and application in Huntington's disease. *NeuroImage*, *47*(4), 1659–1665. https://doi.org/10.1016/j.neuroimage.2009.06.003

Hornberger, M., Wong, S., Tan, R., Irish, M., Piguet, O., Kril, J., … Halliday, G. (2012). In vivo and post-mortem memory circuit integrity in frontotemporal dementia and Alzheimer's disease. *Brain*, *135*, 3015–3025. https://doi.org/10.1093/brain/aws239

Hua, X., Hibar, D. P., Ching, C. R. K., Boyle, C. P., Rajagopalan, P., Gutman, B. A., … Thompson, P. M. (2013). Unbiased tensor-based morphometry: Improved robustness and sample size estimates for Alzheimer's disease clinical trials. *NeuroImage*, *66*, 648–661. https://doi.org/10.1016/j.neuroimage.2012.10.086

Iglesias, J. E., Insausti, R., Lerma-Usabiaga, G., Bocchetta, M., Van

Leemput, K., Greve, D. N., … Paz-Alonso, P. M. (2018). A probabilistic atlas of the human thalamic nuclei combining ex vivo MRI and histology. *NeuroImage*, *183*, 314–326. https://doi.org/10.1016/j.neuroimage.2018.08.012

Iglesias, J. E., & Sabuncu, M. R. (2015). Multi-Atlas Segmentation of Biomedical Images: A Survey. *Medical Image Analysis*, *24*(1), 205–219. https://doi.org/10.1016/j.physbeh.2017.03.040

Iscan, Z., Jin, T. B., Kendrick, A., Szeglin, B., Lu, H., Trivedi, M., … Delorenzo, C. (2015). Test-retest reliability of freesurfer measurements within and between sites: Effects of visual approval process. *Human Brain Mapping*, *36*(9), 3472–3485. https://doi.org/10.1002/hbm.22856

Jack, C. R., Shiung, M. M., Gunter, J. L., O'Brien, P. C., Weigand, S. D., Knopman, D. S., … Petersen, R. C. (2004). Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology*, *62*(4), 591–600. https://doi.org/10.1212/01.WNL.0000110315.26026.EF

Jiskoot, L. C., Bocchetta, M., Nicholas, J. M., Cash, D. M., Thomas, D., Modat, M., … Rohrer, J. D. (2018). Presymptomatic white matter integrity loss in familial frontotemporal dementia in the GENFI cohort: A cross-sectional diffusion tensor imaging study. *Annals of Clinical and Translational Neurology*, *5*(9), 1025–1036. https://doi.org/10.1002/acn3.601

Jiskoot, L. C., Panman, J. L., Meeter, L. H., Dopper, E. G. P., Donker Kaat, L., Franzen, S., … Van Swieten, J. C. (2019). Longitudinal multimodal MRI as prognostic and diagnostic biomarker in presymptomatic familial frontotemporal dementia. *Brain*, *142*(1), 193–208. https://doi.org/10.1093/brain/awy288

Johnson, E. B., Gregory, S., Johnson, H. J., Durr, A., Leavitt, B. R., Roos, R. A., … Scahill, R. I. (2017). Recommendations for the use of automated gray matter segmentation tools: Evidence from Huntington's disease. *Frontiers in Neurology*, *8*, 519. https://doi.org/10.3389/fneur.2017.00519

Josephs, K. A., Whitwell, J., Dickson, D. W., Boeve, B. F., Knopman, D. S.,

Petersen, R. C., … Jack, C. R. (2008). Voxel-based morphometry in autopsy proven PSP and CBD. *Neurobiology of Aging*, *29*(2), 280–289. https://doi.org/10.1016/j.neurobiolaging.2006.09.019

Josephs, K. A., Whitwell, J., Knopman, D. S., Boeve, B. F., Vemuri, P., Senjem, M. L., … Jack, C. R. (2009). Two distinct subtypes of right temporal variant frontotemporal dementia. *Neurology*, *73*(18), 1443–1450. https://doi.org/10.1212/WNL.0b013e3181bf9945

Josephs, K. A., Whitwell, J. L., Parisi, J. E., Petersen, R. C., Boeve, B. F., Jack, C. R., & Dickson, D. W. (2010). Caudate atrophy on MRI is a characteristic feature of FTLD-FUS. *European Journal of Neurology*, *17*(7), 969–975. https://doi.org/10.1111/j.1468-1331.2010.02975.x

Katuwal, G. J., Baum, S. A., Cahill, N. D., Dougherty, C. C., Evans, E., Evans, D. W., … Michael, A. M. (2016). Inter-method discrepancies in brain volume estimation may drive inconsistent findings in autism. *Frontiers in Neuroscience*, *10*(439), 1–16. https://doi.org/10.3389/fnins.2016.00439

Kim, E.-J., Park, Y.-E., Kim, D.-S., Ahn, B.-Y., Kim, H.-S., Chang, Y. H., … Kim, S. (2011). Inclusion body myopathy with Paget disease of bone and frontotemporal dementia linked to VCP p.Arg155Cys in a Korean family. *Archives of Neurology*, *68*(6), 787–796. https://doi.org/10.1001/archneurol.2010.376

Knopman, D. S., Jack, 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. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2690986&tool=pmcentrez&rendertype=abstract

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. https://doi.org/10.1093/brain/awn234

Kril, J. J., & Halliday, G. M. (2004). Clinicopathological staging of

frontotemporal dementia severity: Correlation with regional atrophy. In *Dementia and Geriatric Cognitive Disorders* (Vol. 17, pp. 311–315). https://doi.org/10.1159/000077161

Krueger, C. E., Dean, D. L., Rosen, H. J., Halabi, C., Weiner, M. W., Miller, B. L., & Kramer, J. H. (2010). Longitudinal rates of lobar atrophy in frontotemporal dementia, semantic dementia, and Alzheimer's disease. *Alzheimer Disease and Associated Disorders*, *24*(1), 43–48. https://doi.org/10.1097/WAD.0b013e3181a6f101.LONGITUDINAL

Kurth, F., Luders, E., & Gaser, C. (2015). Voxel-Based Morphometry. *Brain Mapping: An Encyclopedic Reference*, *1*, 345–349.

Lam, B. Y. K., Halliday, G. M., Irish, M., Hodges, J. R., & Piguet, O. (2014). Longitudinal white matter changes in frontotemporal dementia subtypes. *Human Brain Mapping*, *35*(7), 3547–3557. https://doi.org/10.1002/hbm.22420

Landin-Romero, R., Kumfor, F., Leyton, C. E., Irish, M., Hodges, J. R., & Piguet, O. (2017). Disease-specific patterns of cortical and subcortical degeneration in a longitudinal study of Alzheimer's disease and behavioural-variant frontotemporal dementia. *NeuroImage*, *151*, 72–80. https://doi.org/10.1016/j.neuroimage.2016.03.032

Landin-Romero, R., & Piguet, O. (2017). Recent advances in longitudinal structural neuroimaging of younger-onset dementias. *Neurodegenerative Disease Management*, *7*(6), 349–352. https://doi.org/10.2217/nmt-2017-0057

Landin-Romero, R., Tan, R., Hodges, J. R., & Kumfor, F. (2016). An update on semantic dementia: Genetics, imaging, and pathology. *Alzheimer's Research and Therapy*, *8*(1), 1–9. https://doi.org/10.1186/s13195-016-0219-5

Lashley, T., Rohrer, J. D., Mead, S., & Revesz, T. (2015). Review: An update on clinical, genetic and pathological aspects of frontotemporal lobar degenerations. *Neuropathology and Applied Neurobiology*, *72*, n/a-n/a. https://doi.org/10.1111/nan.12250

Ledig, C., Heckemann, R. A., Hammers, A., Lopez, J. C., Newcombe, V. F. J., Makropoulos, A., … Rueckert, D. (2015). Robust whole-brain

segmentation: Application to traumatic brain injury. *Medical Image Analysis*, *21*(1), 40–58. https://doi.org/10.1016/j.media.2014.12.003

Lee, S. E., Khazenzon, A. M., Trujillo, A. J., Guo, C. C., Yokoyama, J. S., Sha, S. J., … Seeley, W. W. (2014). Altered network connectivity in frontotemporal dementia with C9orf72 hexanucleotide repeat expansion. *Brain : A Journal of Neurology*, *137*(Pt 11), 3047–3060. https://doi.org/10.1093/brain/awu248

Lee, S. E., Sias, A. C., Mandelli, M. L., Brown, J. A., Brown, A. B., Khazenzon, A. M., … Seeley, W. W. (2017). Network degeneration and dysfunction in presymptomatic C9ORF72 expansion carriers. *NeuroImage: Clinical*, *14*, 286–297. https://doi.org/10.1016/j.nicl.2016.12.006

Lehmann, M., Douiri, A., Kim, L. G., Modat, M., Chan, D., Ourselin, S., … Fox, N. C. (2010). Atrophy patterns in Alzheimer's disease and semantic dementia: A comparison of FreeSurfer and manual volumetric measurements. *NeuroImage*, *49*(3), 2264–2274. https://doi.org/10.1016/j.neuroimage.2009.10.056

Leung, K. K., Barnes, J., Modat, M., Ridgway, G. R., Bartlett, J. W., Fox, N. C., … Neuroimaging Initiative, A. D. (2011). Brain MAPS: An automated, accurate and robust brain extraction technique using a template library. *NeuroImage*, (55), 1091–1108. https://doi.org/10.1016/j.neuroimage.2010.12.067

Leung, K. K., Barnes, J., Ridgway, G. R., Bartlett, J. W., Clarkson, M. J., Macdonald, K., … Neuroimaging Initiative, A. D. (2010). Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. *NeuroImage*, *51*(4), 1345–1359. https://doi.org/10.1016/j.neuroimage.2010.03.018

Leung, K. K., Ridgway, G. R., Ourselin, S., & Fox, N. C. (2012). Consistent multi-time-point brain atrophy estimation from the boundary shift integral. *NeuroImage*, *59*(4), 3995–4005. https://doi.org/10.1016/j.neuroimage.2011.10.068

Long, Z., Irish, M., Piguet, O., Kiernan, M. C., Hodges, J. R., & Burrell, J. R. (2019). Clinical and neuroimaging investigations of language

disturbance in frontotemporal dementia–motor neuron disease patients. *Journal of Neurology*, *266*(4), 921–933. https://doi.org/10.1007/s00415-019-09216-0

Looi, J. C. L., Lindberg, O., Zandbelt, B. B., Östberg, P., Andersen, C., Botes, L., … Wahlund, L. O. (2008). Caudate nucleus volumes in frontotemporal lobar degeneration: Differential atrophy in subtypes. *American Journal of Neuroradiology*, *29*(8), 1537–1543. https://doi.org/10.3174/ajnr.A1168

Macfarlane, M. D., Jakabek, D., Walterfang, M., Vestberg, S., Velakoulis, D., Wilkes, F. A., … Santillo, A. F. (2015). Striatal Atrophy in the Behavioural Variant of Frontotemporal Dementia: Correlation with Diagnosis, Negative Symptoms and Disease Severity. *PLoS ONE*, *10*(6), 1–14. https://doi.org/10.1371/journal.pone.0129692

Mackenzie, I. R. A., Baker, M. C., Pickering-Brown, S. M., Hsiung, G.-Y. R., Lindholm, C., Dwosh, E., … Feldman, H. H. (2006). The neuropathology of frontotemporal lobar degeneration caused by mutations in the progranulin gene. *Brain*, *129*(Pt 11), 3081–3090. https://doi.org/10.1093/brain/awl271

Mackenzie, I. R. A., Neumann, M., Baborie, A., Sampathu, D. M., Du Plessis, D., Jaros, E., … Lee, V. M. Y. (2011). A harmonized classification system for FTLD-TDP pathology. *Acta Neuropathologica*, *122*(1), 111–113. https://doi.org/10.1007/s00401-011-0845-8

Mackenzie, I. R. A., Neumann, M., Bigio, E. H., Cairns, N. J., Alafuzoff, I., Kril, J., … Mann, D. M. A. (2010). Nomenclature and nosology for neuropathologic subtypes of frontotemporal lobar degeneration: an update. *Acta Neuropathologica*, *119*(1), 1–4. https://doi.org/10.1007/s00401-009-0612-2

Mahoney, C. J., Beck, J., Rohrer, J. D., Lashley, T., Mok, K., Shakespeare, T., … Warren, J. D. (2012). Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: clinical, neuroanatomical and neuropathological features. *Brain*, *135*(Pt 3), 736–750. https://doi.org/10.1093/brain/awr361

Mahoney, C. J., Downey, L. E., Ridgway, G. R., Beck, J., Clegg, S., Blair,

M., … Warren, J. D. (2012). Longitudinal neuroimaging and neuropsychological profiles of frontotemporal dementia with C9ORF72 expansions. *Alzheimer's Research & Therapy*, *4*(5), 41–51. https://doi.org/10.1186/alzrt144

Malone, I. B., Cash, D., Ridgway, G. R., MacManus, D. G., Ourselin, S., Fox, N. C., & Schott, J. M. (2013). MIRIAD-Public release of a multiple time point Alzheimer's MR imaging dataset. *NeuroImage*, *70*, 33–36. https://doi.org/10.1016/j.neuroimage.2012.12.044

Malone, I. B., Leung, K. K., Clegg, S., Barnes, J., Whitwell, J. L., Ashburner, J., … Ridgway, G. R. (2015). Accurate automatic estimation of total intracranial volume: A nuisance variable with less nuisance. *NeuroImage*, *104*(October), 366–372. https://doi.org/10.1016/j.neuroimage.2014.09.034

Marshall, C. R., Hardy, C. J. D., Volkmer, A., Russell, L. L., Bond, R. L., Fletcher, P. D., … Warren, J. D. (2018). Primary progressive aphasia: a clinical approach. *Journal of Neurology*, *265*(6), 1474–1490. https://doi.org/10.1007/s00415-018-8762-6

Massey, L. a., Jager, H. R., Paviour, D. C., O'Sullivan, S. S., Ling, H., Williams, D. R., … Micallef, C. (2013). The midbrain to pons ratio: A simple and specific MRI sign of progressive supranuclear palsy. *Neurology*, *80*(20), 1856–1861. https://doi.org/10.1212/WNL.0b013e318292a2d2

Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., & Lancaster, J. (1995). A probabilistic atlas of the human brain: Theory and rationale for its development. *NeuroImage*, *2*(2), 89–101. https://doi.org/10.1006/nimg.1995.1012

Meeter, L. H., Kaat, L. D., Rohrer, J. D., & Van Swieten, J. C. (2017). Imaging and fluid biomarkers in frontotemporal dementia. *Nature Reviews Neurology*, *13*(7), 406–419. https://doi.org/10.1038/nrneurol.2017.75

Mendez, M. F., & Shapira, J. S. (2011). Loss of emotional insight in behavioral variant frontotemporal dementia or " frontal anosodiaphoria." *Consciousness and Cognition*, *20*(4), 1690–1696.

https://doi.org/10.1016/j.concog.2011.09.005

Mesulam, M. M. (2001). Primary progressive aphasia. *Annals of Neurology*, *49*, 425–432. https://doi.org/10.14412/2074-2711-2019-1-4-11

Mesulam, M. M., Weintraub, S., Rogalski, E. J., Wieneke, C., Geula, C., & Bigio, E. H. (2014). Asymmetry and heterogeneity of Alzheimer's and frontotemporal pathology in primary progressive aphasia. *Brain*, *137*(4), 1176–1192. https://doi.org/10.1093/brain/awu024

Mesulam, M., Wicklund, A., Johnson, N., Rogalski, E., Léger, G. C., Rademaker, A., … Bigio, E. H. (2008). Alzheimer and frontotemporal pathology in subsets of primary progressive aphasia. *Annals of Neurology*, *63*(6), 709–719. https://doi.org/10.1002/ana.21388

Mesulam, M., Wieneke, C., Rogalski, E., Cobia, D., Thompson, C., & Weintraub, S. (2009). Quantitative template for subtyping primary progressive aphasia. *Archives of Neurology*, *66*(12), 1545–1551. https://doi.org/10.1001/archneurol.2009.288

Mioshi, E., Hsieh, S., Savage, S., Hornberger, M., & Hodges, J. R. (2010). Clinical staging and disease progression in frontotemporal dementia. *Neurology*, *74*(20), 1591–1597. https://doi.org/10.1212/WNL.0b013e3181e04070

Montembeault, M., Brambati, S. M., Gorno-Tempini, M. L., & Migliaccio, R. (2018, August 21). Clinical, anatomical, and pathological features in the three variants of primary progressive aphasia: A review. *Frontiers in Neurology*. Frontiers Media S.A. https://doi.org/10.3389/fneur.2018.00692

Moore, K. M., Nicholas, J., Grossman, M., McMillan, C. T., Irwin, D. J., Massimo, L., … Geschwind, D. (2020). Age at symptom onset and death and disease duration in genetic frontotemporal dementia: an international retrospective cohort study. *The Lancet Neurology*, *19*(2), 145–156. https://doi.org/10.1016/S1474-4422(19)30394-1

Morris, J. C., Aisen, P. S., Bateman, R. J., Benzinger, T. L., Cairns, N. J., Fagan, A. M., … Buckles, V. D. (2012). Developing an international network for Alzheimer's research: the Dominantly Inherited Alzheimer Network. *Clinical Investigation*, *2*(10), 975–984.

https://doi.org/10.4155/cli.12.93

Norise, C., Ungrady, M., Halpin, A., Jester, C., McMillan, C. T., Irwin, D. J., … Grossman, M. (2019). Clinical correlates of Alzheimer's disease cerebrospinal fluid analytes in primary progressive aphasia. *Frontiers in Neurology*, *10*(MAY), 1–8. https://doi.org/10.3389/fneur.2019.00485

Olm, C. A., McMillan, C. T., Irwin, D. J., Van Deerlin, V. M., Cook, P. A., Gee, J. C., & Grossman, M. (2018). Longitudinal structural gray matter and white matter MRI changes in presymptomatic progranulin mutation carriers. *NeuroImage: Clinical*, *19*, 497–506. https://doi.org/10.1016/j.nicl.2018.05.017

Onyike, C. U., & Diehl-Schmid, J. (2013). The epidemiology of frontotemporal dementia. *International Review of Psychiatry (Abingdon, England)*, *25*(2), 130–137. https://doi.org/10.3109/09540261.2013.776523.The

Pan, P. L., Song, W., Yang, J., Huang, R., Chen, K., Gong, Q. Y., … Sang, H. F. (2012). Gray matter atrophy in behavioral variant frontotemporal dementia: a meta-analysis of voxel-based morphometry studies. *Dementia and Geriatric Cognitive Disorders*, *33*, 141–148. https://doi.org/10.1159/000

Pankov, A., Binney, R. J., Staffaroni, A. M., Kornak, J., Attygalle, S., Schuff, N., … Rosen, H. J. (2016). Data-driven regions of interest for longitudinal change in frontotemporal lobar degeneration. *NeuroImage: Clinical*, *12*, 332–340. https://doi.org/10.1016/j.nicl.2015.08.002

Panman, J. L., Jiskoot, L. C., Bouts, M. J. R. J., Meeter, L. H. H., van der Ende, E. L., Poos, J. M., … Papma, J. M. (2019). Gray and white matter changes in presymptomatic genetic frontotemporal dementia: a longitudinal MRI study. *Neurobiology of Aging*, *76*, 115–124. https://doi.org/10.1016/j.neurobiolaging.2018.12.017

Papma, J. M., Jiskoot, L. C., Panman, J. L., Dopper, E. G., Den Heijer, T., Donker Kaat, L., … Van Swieten, J. C. (2017). Cognition and gray and white matter characteristics of presymptomatic C9orf72 repeat expansion. *Neurology*, *89*(12), 1256–1264. https://doi.org/10.1212/WNL.0000000000004393

Patenaude, B., Smith, S. M., Kennedy, D. N., & Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage*, *56*(3), 907–922. https://doi.org/10.1016/j.neuroimage.2011.02.046

Perini, G., Cotta Ramusino, M., Sinforiani, E., Franciotta, D., Trifirò, G., Ceroni, M., & Costa, A. (2018). Role of Cerebrospinal Fluid Biomarkers and (18)F-florbetapir PET Imaging in the Diagnosis of Primary Progressive Aphasia: A Retrospective Analysis. *Alzheimer Disease and Associated Disorders*, *33*(3), 282–284. https://doi.org/10.1097/WAD.0000000000000289

Popuri, K., Dowds, E., Beg, M. F., Balachandar, R., Bhalla, M., Jacova, C., … Hsiung, G. Y. R. (2018). Gray matter changes in asymptomatic C9orf72 and GRN mutation carriers. *NeuroImage: Clinical*, *18*(January), 591–598. https://doi.org/10.1016/j.nicl.2018.02.017

Prados, F., Cardoso, M. J., Leung, K. K., Cash, D. M., Modat, M., Fox, N. C., … Ourselin, S. (2015). Measuring brain atrophy with a generalized formulation of the boundary shift integral. *Neurobiology of Aging*, *36*(S1), S81–S90. https://doi.org/10.1016/j.neurobiolaging.2014.04.035

Premi, E., Garibotto, V., Gazzina, S., Formenti, A., Archetti, S., Gasparotti, R., … Borroni, B. (2014). Subcortical and Deep Cortical Atrophy in Frontotemporal Dementia due to Granulin Mutations. *Dementia and Geriatric Cognitive Disorders Extra*, *4*(1), 95–102. https://doi.org/10.1159/000355428

Rascovsky, K., Hodges, J. R., Knopman, D., Mendez, M. F., Kramer, J. H., Neuhaus, J., … Miller, B. L. (2011). Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain: A Journal of Neurology*, *134*(Pt 9), 2456–2477. https://doi.org/10.1093/brain/awr179

Rogalski, E. J., Cobia, D., Harrison, T. M., Wieneke, C., Weintraub, S., & Mesulam, M. M. (2011). Progression of language decline and cortical atrophy in subtypes of primary progressive aphasia. *Neurology*, *76*(21), 1804–1810. https://doi.org/10.1212/WNL.0b013e31821ccd3c

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. *Journal of Alzheimer's Disease : JAD*, *30*(2), 407–411. https://doi.org/10.3233/JAD-2012-111556

Rohrer, J. D., Crutch, S. J., Warrington, E. K., & Warren, J. D. (2010). Progranulin-associated primary progressive aphasia: A distinct phenotype? *Neuropsychologia*, *48*(1), 288–297. https://doi.org/10.1016/j.neuropsychologia.2009.09.017

Rohrer, J. D., Geser, F., Zhou, J., Gennatas, E. D., Sidhu, M., Trojanowski, J. Q., … Seeley, W. W. (2010). TDP-43 subtypes are associated with distinct atrophy patterns in frontotemporal dementia. *Neurology*, *75*(24), 2204–2211. Retrieved from http://discovery.ucl.ac.uk/1315712/

Rohrer, J. D., Guerreiro, R., Vandrovcova, J., Uphill, J., Reiman, D., Beck, J., … Rossor, M. N. (2009). The heritability and genetics of frontotemporal lobar degeneration. *Neurology*, *73*(18), 1451–1456. Retrieved from http://discovery.ucl.ac.uk/179269/

Rohrer, J. D., Lashley, T., Schott, J. M., Warren, J. E., Mead, S., Isaacs, A. M., … Warren, J. D. (2011). Clinical and neuroanatomical signatures of tissue pathology in frontotemporal lobar degeneration. *Brain: A Journal of Neurology*, *134*(9), 2565–2581. https://doi.org/10.1093/brain/awr198

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

Rohrer, J. D., Nicholas, J. M., Cash, D. M., van Swieten, J., Dopper, E., Jiskoot, L., … Rossor, M. N. (2015). Presymptomatic cognitive and neuroanatomical changes in genetic frontotemporal dementia in the Genetic Frontotemporal dementia Initiative (GENFI) study: A cross-sectional analysis. *The Lancet Neurology*, *14*(3), 253–262. https://doi.org/10.1016/S1474-4422(14)70324-2

Rohrer, J. D., Ridgway, G. R., Crutch, S. J., Hailstone, J., Goll, J. C., Clarkson, M. J., … Warren, J. D. (2010). Progressive

logopenic/phonological aphasia: Erosion of the language network. *NeuroImage*, *49*(1), 984–993. https://doi.org/10.1016/j.neuroimage.2009.08.002

Rohrer, J. D., Ridgway, G. R., Modat, M., Ourselin, S., Mead, S., Fox, N. C., … Warren, J. D. (2010). Distinct profiles of brain atrophy in frontotemporal lobar degeneration caused by progranulin and tau mutations. *NeuroImage*, *53*(3–3), 1070–1076. https://doi.org/10.1016/j.neuroimage.2009.12.088

Rohrer, J. D., Rossor, M. N., & Warren, J. D. (2010). Syndromes of nonfluent primary progressive aphasia: a clinical and neurolinguistic analysis. *Neurology*, *75*(7), 603–610. https://doi.org/10.1212/WNL.0b013e3181ed9c6b

Rohrer, J. D., & Warren, J. D. (2011). Phenotypic signatures of genetic frontotemporal dementia. *Current Opinion in Neurology*, *24*(6), 542–549. https://doi.org/10.1097/WCO.0b013e32834cd442

Rohrer, J. D., Warren, J. D., Modat, M., Ridgway, G. R., Douiri, A., Rossor, M. N., … Fox, N. C. (2009). Patterns of cortical thinning in the language variants of frontotemporal lobar degeneration. *Neurology*, *72*(18), 1562–1569. Retrieved from http://discovery.ucl.ac.uk/152379/

Schott, J. M., Bartlett, J. W., Barnes, J., Leung, K. K., Ourselin, S., & Fox, N. C. (2010). Reduced sample sizes for atrophy outcomes in Alzheimer's disease trials: baseline adjustment. *Neurobiology of Aging*, *31*(8), 1452–1462, 1462.e1-2. https://doi.org/10.1016/j.neurobiolaging.2010.04.011

Schott, J. M., Frost, C., Whitwell, J. L., MacManus, D. G., Boyes, R. G., Rossor, M. N., & Fox, N. C. (2006). Combining short interval MRI in Alzheimer's disease: Implications for therapeutic trials. *Journal of Neurology*, *253*(9), 1147–1153. https://doi.org/10.1007/s00415-006-0173-4

Schroeter, M. L., Laird, A. R., Chwiesko, C., Deuschl, C., Schneider, E., Bzdok, D., … Neumann, J. (2014). Conceptualizing neuropsychiatric diseases with multimodal data-driven meta-analyses - The case of behavioral variant frontotemporal dementia. *Cortex*, *57*, 22–37.

https://doi.org/10.1016/j.cortex.2014.02.022

Schroeter, M. L., Raczka, K., Neumann, J., & Yves von Cramon, D. (2007). Towards a nosology for frontotemporal lobar degenerations-A meta-analysis involving 267 subjects. *NeuroImage*, *36*(3), 497–510. https://doi.org/10.1016/j.neuroimage.2007.03.024

Seelaar, H., Klijnsma, K. Y., de Koning, I., van der Lugt, A., Chiu, W. Z., Azmani, A., … van Swieten, J. C. (2010). Frequency of ubiquitin and FUS-positive, TDP-43-negative frontotemporal lobar degeneration. *Journal of Neurology*, *257*(5), 747–753. https://doi.org/10.1007/s00415-009-5404-z

Seelaar, H., Rohrer, J. D., Pijnenburg, Y. a L., Fox, N. C., & van Swieten, J. C. (2011). Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review. *Journal of Neurology, Neurosurgery, and Psychiatry*, *82*(5), 476–486. https://doi.org/10.1136/jnnp.2010.212225

Seeley, W. W. (2019). Behavioral Variant Frontotemporal Dementia. *CONTINUUM: Lifelong Learning in Neurology*, *25*(1), 76–100. https://doi.org/10.1212/CON.0000000000000698

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. *Archives of Neurology*, *65*(2), 249–255. https://doi.org/10.1001/archneurol.2007.38

Sha, S. J., Miller, Z. A., Min, S. won, Zhou, Y., Brown, J., Mitic, L. L., … Boxer, A. L. (2017). An 8-week, open-label, dose-finding study of nimodipine for the treatment of progranulin insufficiency from GRN gene mutations. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, *3*(4), 507–512. https://doi.org/10.1016/j.trci.2017.08.002

Sha, S. J., Takada, L. T., Rankin, K. P., Yokoyama, J. S., Rutherford, N. J., Fong, J. C., … Boxer, A. L. (2012). Frontotemporal dementia due to C9ORF72 mutations: Clinical and imaging features. *Neurology*, *79*(10), 1002–1011. https://doi.org/10.1212/WNL.0b013e318268452e

Shaikh, T. A., & Ali, R. (2019). Automated atrophy assessment for Alzheimer's disease diagnosis from brain MRI images. *Magnetic Resonance Imaging*, *62*, 167–173. https://doi.org/10.1016/j.mri.2019.06.019

Sheelakumari, R., Bineesh, C., Varghese, T., Kesavadas, C., Verghese, J., & Mathuranath, P. S. (2019). Neuroanatomical correlates of apathy and disinhibition in behavioural variant frontotemporal dementia. *Brain Imaging and Behavior*, *4*, 1–8. https://doi.org/10.1007/s11682-019-00150-3

Smith, S. M., Rao, A., De Stefano, N., Jenkinson, M., Schott, J. M., Matthews, P. M., & Fox, N. C. (2007). Longitudinal and cross-sectional analysis of atrophy in Alzheimer's disease: Cross-validation of BSI, SIENA and SIENAX. *NeuroImage*, *36*(4), 1200–1206. https://doi.org/10.1016/j.neuroimage.2007.04.035

Smith, S. M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P. M., Federico, A., & De Stefano, N. (2002). Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage*, *17*(1), 479–489. https://doi.org/10.1006/nimg.2002.1040

Spina, S., Farlow, M. R., Unverzagt, F. W., Kareken, D. A., Murrell, J. R., Fraser, G., … Ghetti, B. (2008). The tauopathy associated with mutation +3 in intron 10 of Tau: characterization of the MSTD family. *Brain*, *131*(1), 72–89. https://doi.org/10.1093/brain/awm280

Spinelli, E. G., Mandelli, M. L., Miller, Z. A., Santos-Santos, M. A., Wilson, S. M., Agosta, F., … Gorno-Tempini, M. L. (2017). Typical and atypical pathology in primary progressive aphasia variants. *Annals of Neurology*, *81*(3), 430–443. https://doi.org/10.1002/ana.24885

Staffaroni, A. M., Ljubenkov, P. A., Kornak, J., Cobigo, Y., Datta, S., Marx, G., … Rosen, H. J. (2019). Longitudinal multimodal imaging and clinical endpoints for frontotemporal dementia clinical trials. *Brain*, *142*(2), 443–459. https://doi.org/10.1093/brain/awy319

Stojkovic, T., Hammouda, E. H., Richard, P., López de Munain, A., Ruiz-Martinez, J., Camaño, P., … Eymard, B. (2009). Clinical outcome in 19 French and Spanish patients with valosin-containing protein myopathy

associated with Paget's disease of bone and frontotemporal dementia. *Neuromuscular Disorders : NMD*, *19*(5), 316–323. https://doi.org/10.1016/j.nmd.2009.02.012

Surampalli, A., Gold, B. T., Smith, C., Castellani, R. J., Khare, M., Yu, H., … Kimonis, V. (2015). A case report comparing clinical, imaging and neuropsychological assessment findings in twins discordant for the VCP p.R155C mutation. *Neuromuscular Disorders : NMD*, *25*(2), 177–183. https://doi.org/10.1016/j.nmd.2014.10.003

Tabrizi, S. J., Langbehn, D. R., Leavitt, B. R., Roos, R. A., Durr, A., Craufurd, D., … Stout, J. C. (2009). Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *The Lancet Neurology*, *8*(9), 791–801. https://doi.org/10.1016/S1474-4422(09)70170-X

Tabrizi, S. J., Leavitt, B. R., Landwehrmeyer, G. B., Wild, E. J., Saft, C., Barker, R. A., … Lane, R. M. (2019). Targeting huntingtin expression in patients with Huntington's disease. *New England Journal of Medicine*, *380*(24), 2307–2316. https://doi.org/10.1056/NEJMoa1900907

Teichmann, M., Kas, A., Boutet, C., Ferrieux, S., Nogues, M., Samri, D., … Migliaccio, R. (2013). Deciphering logopenic primary progressive aphasia: A clinical, imaging and biomarker investigation. *Brain*, *136*(11), 3474–3488. https://doi.org/10.1093/brain/awt266

Tustison, N. J., Cook, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction Nicholas. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908.N4ITK

Valente, E. S., Caramelli, P., Gambogi, L. B., Mariano, L. I., Guimarães, H. C., Teixeira, A. L., & De Souza, L. C. (2019). Phenocopy syndrome of behavioral variant frontotemporal dementia: A systematic review. *Alzheimer's Research and Therapy*, *11*(1), 30–45. https://doi.org/10.1186/s13195-019-0483-2

van De Pol, L., Hensel, A., Van Der Flier, W. M., Visser, P. J., Pijnenburg, Y. A. L., Barkhof, F., … Scheltens, P. (2006). Hippocampal atrophy on MRI in frontotemporal lobar degeneration and Alzheimer's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, *77*(4), 439–442.

https://doi.org/10.1136/jnnp.2005.075341

Van Vliet, D., De Vugt, M. E., Bakker, C., Pijnenburg, Y. A. L., Vernooij-Dassen, M. J. F. J., Koopmans, R. T. C. M., & Verhey, F. R. J. (2013). Time to diagnosis in young-onset dementia as compared with late-onset dementia. *Psychological Medicine*, *43*(2), 423–432. https://doi.org/10.1017/S0033291712001122

Walhout, R., Schmidt, R., Westeneng, H. J., Verstraete, E., Seelen, M., Van Rheenen, W., … Van Den Berg, L. H. (2015). Brain morphologic changes in asymptomatic C9orf72 repeat expansion carriers. *Neurology*, *85*(20), 1780–1788. https://doi.org/10.1212/WNL.0000000000002135

Wang, H., Suh, J. W., Das, S. R., Pluta, J. B., Craige, C., & Yushkevich, P. A. (2013). Multi-atlas segmentation with joint label fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(3), 611–623. https://doi.org/10.1109/TPAMI.2012.143

Warfield, S. K., Zou, K. H., & Wells, W. M. (2004). Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, *23*(7), 903–921. https://doi.org/10.1109/TMI.2004.828354

Warren, J. D., Rohrer, J. D., Schott, J. M., Fox, N. C., Hardy, J., & Rossor, M. N. (2013). Molecular nexopathies: A new paradigm of neurodegenerative disease. *Trends in Neurosciences*, *36*(10), 561–569. https://doi.org/10.1016/j.tins.2013.06.007

Whitwell, J. L., Boeve, B. F., Weigand, S. D., Senjem, M. L., Gunter, J. L., Baker, M. C., … Josephs, K. a. (2015). Brain atrophy over time in genetic and sporadic frontotemporal dementia: a study of 198 serial magnetic resonance images. *European Journal of Neurology*, *22*, 745–752. https://doi.org/10.1111/ene.12675

Whitwell, J. L., Jack, C. R., Boeve, B. F., Parisi, J. E., Ahlskog, J. E., Drubach, D. a., … Josephs, K. a. (2010). Imaging correlates of pathology in corticobasal syndrome. *Neurology*, *75*(21), 1879–1887. https://doi.org/10.1212/WNL.0b013e3181feb2e8

Whitwell, J. L., Jack, C. R., Parisi, J. E., Senjem, M. L., Knopman, D. S.,

Boeve, B. F., … Josephs, K. a. (2010). Does TDP-43 type confer a distinct pattern of atrophy in frontotemporal lobar degeneration? *Neurology*, *75*(24), 2212–2220. https://doi.org/10.1212/WNL.0b013e31820203c2

Whitwell, J. L., Josephs, K. a., Avula, R., Tosakulwong, N., Weigand, S. D., Senjem, M. L., … Jack, C. R. (2011). Altered functional connectivity in asymptomatic MAPT subjects. A comparison to bvFTD. *Neurology*, *77*(9), 866–874. https://doi.org/10.1212/WNL.0b013e31822c61f2

Whitwell, J, Jack, C. R., Pankratz, V. S., Parisi, J. E., Knopman, D. S., Boeve, B. F., … Josephs, K. a. (2008). Rates of brain atrophy over time in autopsy-proven frontotemporal dementia and Alzheimer disease. *NeuroImage*, *39*(3), 1034–1040. https://doi.org/10.1016/j.neuroimage.2007.10.001

Whitwell, J L, Weigand, S. D., Gunter, J. L., Boeve, B. F., Rademakers, R., Baker, M., … Josephs, K. a. (2011). Trajectories of brain and hippocampal atrophy in FTD with mutations in MAPT or GRN. *Neurology*, *77*(4), 393–398. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3140800&tool=pmcentrez&rendertype=abstract

Whitwell, Jennifer, Duffy, E., Strand, M., Machulda, M., Senjem, C., Schwarz, R., … Josephs. (2015). Clinical and neuroimaging biomarkers of amyloid-negative logopenic primary progressive aphasia. *Brain and Language*, *142*, 45–53. https://doi.org/10.1016/j.bandl.2015.01.009

Whitwell, Jennifer, Duffy, J. R., Strand, E. a., Machulda, M. M., Tosakulwong, N., Weigand, S. D., … Josephs, K. a. (2015). Sample size calculations for clinical trials targeting tauopathies: a new potential disease target. *Journal of Neurology*, *262*(9), 2064–2072. https://doi.org/10.1007/s00415-015-7821-5

Whitwell, Jennifer L., Sampson, E. L., Watt, H. C., Harvey, R. J., Rossor, M. N., & Fox, N. C. (2005). A volumetric magnetic resonance imaging study of the amygdala in frontotemporal lobar degeneration and Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *20*(4), 238–244. https://doi.org/10.1159/000087343

Whitwell, Jennifer L, Josephs, K. A., Rossor, M. N., Stevens, J. M., Revesz, T., Holton, J. L., … Warren, J. D. (2005). Magnetic resonance imaging signatures of tissue pathology in frontotemporal dementia. *Archives of Neurology*, *62*(9), 1402–1408. https://doi.org/10.1001/archneur.62.9.1402

Whitwell, Jennifer L, Przybelski, S. a, Weigand, S. D., Ivnik, R. J., Vemuri, P., Gunter, J. L., … Josephs, K. a. (2009). Distinct anatomical subtypes of the behavioural variant of frontotemporal dementia: a cluster analysis study. *Brain: A Journal of Neurology*, *132*(Pt 11), 2932–2946. https://doi.org/10.1093/brain/awp232

Whitwell, Jennifer L, Weigand, S. D., Boeve, B. F., Senjem, M. L., Gunter, J. L., Dejesus-Hernandez, M., … Josephs, K. a. (2012). Neuroimaging signatures of frontotemporal dementia genetics: C9ORF72, tau, progranulin and sporadics. *Brain: A Journal of Neurology*, *135*(Pt 3), 794–806. https://doi.org/10.1093/brain/aws001

Whitwell, Jennifer, Weigand, S. D., Boeve, B. F., Senjem, M. L., Gunter, J. L., DeJesus-Hernandez, M., … Josephs, K. A. (2012). Neuroimaging signatures of frontotemporal dementia genetics: C9ORF72, tau, progranulin and sporadics. *Brain*, *135*(3), 794–806. https://doi.org/10.1093/brain/aws001

Wilson, S. M., Galantucci, S., Tartaglia, M. C., & Gorno-Tempini, M. L. (2012). The neural basis of syntactic deficits in primary progressive aphasia. *Brain and Language*, *122*(3), 190–198. https://doi.org/10.1016/j.bandl.2012.04.005

Woollacott, I. (2016). Clinical spectrum of sporadic and familial FTD. *Journal of Neurochemistry*, 6–31.

Woollacott, I., & Rohrer, J. D. (2016). The clinical spectrum of sporadic and familial forms of frontotemporal dementia. *Journal of Neurochemistry*, *138*, 6–31. https://doi.org/10.1111/jnc.13654

Zeun, P., Scahill, R. I., Tabrizi, S. J., & Wild, E. J. (2019). Fluid and imaging biomarkers for Huntington's disease. *Molecular and Cellular Neuroscience*, *97*(October 2018), 67–80. https://doi.org/10.1016/j.mcn.2019.02.004

Zhang, Y; Brady, M; Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the Expectation-Maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57.

9 Appendices

Appendix 1.1 GIF Cortical GM parcellation labels, region acronyms and lobar colour coding (see Appendix 1.2)

Cortical GM labels
Right (AOrG) - anterior orbital gyrus
Left (AOrG) - anterior orbital gyrus
Right (CO) - central operculum
Left (CO) - central operculum
Right (FO) - frontal operculum
Left (FO) - frontal operculum
Right (FRP) - frontal pole
Left (FRP) - frontal pole
Right (Gre) - gyrus rectus
Left (Gre) - gyrus rectus
Right (LOrG) - lateral orbital gyrus
Left (LOrG) - lateral orbital gyrus
Right (MFC) - medial frontal cortex
Left (MFC) - medial frontal cortex
Right (MFG) - middle frontal gyrus
Left (MFG) - middle frontal gyrus
Right (MOrG) - medial orbital gyrus
Left (MOrG) - medial orbital gyrus
Right (MPrG) - precentral gyrus medial segment
Left (MPrG) - precentral gyrus medial segment
Right (MSFG) - superior frontal gyrus medial segment
Left (MSFG) - superior frontal gyrus medial segment
Right (OpIFG) - opercular part of the inferior frontal gyrus
Left (OpIFG) - opercular part of the inferior frontal gyrus
Right (OrIFG) - orbital part of the inferior frontal gyrus
Left (OrIFG) - orbital part of the inferior frontal gyrus
Right (POrG) - posterior orbital gyrus

Left (POrG) - posterior orbital gyrus
Right (PrG) - precentral gyrus
Left (PrG) - precentral gyrus
Right (SCA) - subcallosal area
Left (SCA) - subcallosal area
Right (SFG) - superior frontal gyrus
Left (SFG) - superior frontal gyrus
Right (SMC) - supplementary motor cortex
Left (SMC) - supplementary motor cortex
Right (TrIFG) - triangular part of the inferior frontal gyrus
Left (TrIFG) - triangular part of the inferior frontal gyrus
Right (Ent) - entorhinal area
Left (Ent) - entorhinal area
Right (FuG) - fusiform gyrus
Left (FuG) - fusiform gyrus
Right (ITG) - inferior temporal gyrus
Left (ITG) - inferior temporal gyrus
Right (MTG) - middle temporal gyrus
Left (MTG) - middle temporal gyrus
Right (PHG) - parahippocampal gyrus
Left (PHG) - parahippocampal gyrus
Right (PP) - planum polare
Left (PP) - planum polare
Right (PT) - planum temporale
Left (PT) - planum temporale
Right (STG) - superior temporal gyrus
Left (STG) - superior temporal gyrus
Right (TMP) - temporal pole
Left (TMP) - temporal pole
Right (TTG) - transverse temporal gyrus
Left (TTG) - transverse temporal gyrus
Right (AnG) - angular gyrus
Left (AnG) - angular gyrus

Right (MPoG) - postcentral gyrus medial segment

Left (MPoG) - postcentral gyrus medial segment

Right (Pcu) - precuneus

Left (Pcu) - precuneus

Right (PO) - parietal operculum

Left (PO) - parietal operculum

Right (PoG) - postcentral gyrus

Left (PoG) - postcentral gyrus

Right (SMG) - supramarginal gyrus

Left (SMG) - supramarginal gyrus

Right (SPL) - superior parietal lobule

Left (SPL) - superior parietal lobule

Right (Calc) - calcarine cortex

Left (Calc) - calcarine cortex

Right (Cun) - cuneus

Left (Cun) - cuneus

Right (IOG) - inferior occipital gyrus

Left (IOG) - inferior occipital gyrus

Right (LiG) - lingual gyrus

Left (LiG) - lingual gyrus

Right (MOG) - middle occipital gyrus

Left (MOG) - middle occipital gyrus

Right (OCP) - occipital pole

Left (OCP) - occipital pole

Right (OFuG) - occipital fusiform gyrus

Left (OFuG) - occipital fusiform gyrus

Right (SOG) - superior occipital gyrus

Left (SOG) - superior occipital gyrus

Right (ACgG) - anterior cingulate gyrus

Left (ACgG) - anterior cingulate gyrus

Right (MCgG) - middle cingulate gyrus

Left (MCgG) - middle cingulate gyrus

Right (PCgG) - posterior cingulate gyrus

Left (PCgG) - posterior cingulate gyrus
Right (AIns) - anterior insula
Left (AIns) - anterior insula
Right (PIns) - posterior insula
Left (PIns) - posterior insula

Appendix 1.2. Labels inclusion for each summed lobar region volume (produced for the left and right separately)

FRONTAL Lobe regions	TEMPORAL Lobe regions	PARIETAL Lobe regions	OCCIPITAL Lobe regions	CINGULATE regions	INSULA regions
ORBITOFRONTAL	**MEDIAL TEMP**	**MEDIAL PAR**	**MEDIAL OCC**	**ANTERIOR**	**ANTERIOR**
AOrG	Ent	Pcu	Cun	ACgG	AIns
MOrG	FuG	LATERAL PAR	Calc	**MIDDLE**	**POSTERIOR**
LOrG	PHG	PO	LiG	MCgG	PIns
POrG	**LATERAL TEMP**	SMG	OFuG	**POSTERIOR**	
DLPFC	ITG	SPL	LATERAL OCC	PCgG	
SFG	MTG	AnG	SOG		
MFG	STG	SENSORY	IOG		
OpIFG	**TEMPORAL POLE**	PoG	MOG		
OrIFG	TMP	MPoG	OCP		
TrIFG	**SUPRATEMP**				
VMPFC	PP				
GRe	PT				
MFC	TTG				
SCA					
MSFG					
MOTOR					
PrG					
MPrG					
SMC					
OPERCULAR					
FO					
CO					

Appendix 1.3. Additional GIF tissue labels including subcortical, WM, cerebellum and non-brain regions. The lobar WM segmentations used in conjunction with the combined lobar cortical GM regions in Appendix 1.1 for Chapter 4 lobar analysis are colour coded accordingly. Labels highlighted light blue were combined for the infratentorial GIF volume-difference measure in Chapter 4.

Additional GIF labels	
Non-Brain Outer Tissue - Background	Brain Stem
Non-Brain Low	Right Caudate
Non-Brain Mid	Left Caudate
Non-Brain High	Right Hippocampus
Non-ventricular CSF	Left Hippocampus
3rd Ventricle	Right Inferior Lateral Ventricle
4th Ventricle	Left Inferior Lateral Ventricle
5th Ventricle	Right Lateral Ventricle
Right Frontal White Matter	Left Lateral Ventricle
Right Temporal White Matter	Right Lesion
Right Parietal White Matter	Left Lesion
Right Occipital White Matter	Right Pallidum
Right Cingulate White Matter	Left Pallidum
Right Insula White Matter	Right Putamen
Left Frontal White Matter	Left Putamen
Left Temporal White Matter	Right Thalamus Proper
Left Parietal White Matter	Left Thalamus Proper
Left Occipital White Matter	Right Ventral DC
Left Cingulate White Matter	Left Ventral DC
Left Insula White Matter	Right vessel
Corpus Callosum	Left vessel
Right Cerebellum Exterior	Right Ventricular Lining
Left Cerebellum Exterior	Left Ventricular Lining
Right Cerebellum White Matter	Optic Chiasm
Left Cerebellum White Matter	Right Cerebral Exterior
Cerebellar Vermal Lobules I-V	Left Cerebral Exterior
Cerebellar Vermal Lobules VI-VII	3rd Ventricle (Posterior part)
Cerebellar Vermal Lobules VIII-X	Left Basal Forebrain
Right Accumbens Area	Right Basal Forebrain
Left Accumbens Area	Pons
Right Amygdala	Right Claustrum
Left Amygdala	Left Claustrum