University of Padua

School of Psychology

Padua (Italy)

and

Max Planck Institute for Human Cognitive and Brain Sciences

Department of Neurology

Cognitive Neuropsychiatry

Leipzig (Germany)

Exploring the Diagnosis of Frontotemporal Dementia by Analyzing Neuropsychological Data With K-Means Clustering

Master Thesis

Marie Söntgerath

Prof. Dr. Dr. Matthias Schroeter

Prof. Dr. Simone Cutini

Leipzig, 15th September 2022

Abstract

Background: Differential diagnosis of dementia syndromes is difficult. Treatments are available for Alzheimer's dementia (AD), but effects do not translate to other dementia syndromes. Therefore, early differential diagnosis is necessary to administer appropriate interventions and to boost drug development and research of therapeutic strategies that may lower patient and care-giver burden. The focus of the current study is to disentangle the clinical picture of frontotemporal dementia (FTD) from AD spectrum disorders. Findings could support an early, cheap, and more accurate diagnosis.

Methods: K-means clustering, an unsupervised machine learning algorithm, was used on neuropsychological data from neurologic patients of the FTLD consortium databank. The analysis was performed twice, once including only neuropsychological test scores and a second time combining the neuropsychological variables with questionnaire scores assessing behavioral changes. In total n = 484 and n = 469 participants were included in the analysis with and without questionnaires, respectively. Participants included were either healthy controls with no family relation to patients in the dataset, or patients diagnosed with one of the following dementia disorders: AD, a behavioral variant of FTD (bvFTD), or one of three possible primary progressive aphasia (PPA) syndromes - a semantic variant (svPPA), a non-fluent variant (nfvPPA) or a logopenic variant (lvPPA).

Results: Agreement of results from the various analyses performed was relatively high. Homogeneous clusters of diagnostic groups emerged. Homogeneity seemed higher for bvFTD and svPPA than for the other patient groups. NfvPPA and lvPPA patients were particularly likely to cluster together. Exploring neuropsychological patterns of cluster results demonstrated high variability between patients of the same diagnostic groups, which could partly be explained by differences in disease severity. Tests that might prove particularly relevant to distinguish diagnostic subgroups are the FTLD-CDR sub-scores, the repeat and point task as well as questionnaires assessing apathy.

Conclusion: K-means clustering proved to be a useful technique to explore various diagnostic syndromes that show overlapping clinical pictures. This study helped to formulate specific hypotheses based on the observation of patterns in multidimensional data. Disease severity showed to impact k-means clustering results considerably and should therefore be accounted for in future studies. Future studies will need to test the formulated hypotheses and inspect the meaning of impure clusters.

Keywords: k-means clustering, frontotemporal dementia, primary progressive aphasia, Alzheimer's dementia, differential diagnosis, neuropsychological tests

Table of Content

Abbreviations

AAT	Aachen Aphasia Test
AD	Alzheimer's Dementia/ Alzheimer's Disease
AES	Apathy Evaluation Scale
B.ADL	Bayer Activities of Daily Living
BNT	Boston Naming Test
BvFTD	Behavioral variant Frontotemporal Dementia
CBS	Corticobasal Degeneration
CDR	Clinical Dementia Rating Scale
CERAD	Consortium to Establish a Registry for Alzheimer's Disease
CSF	Cerebrospinal Fluid
CT	Computed Tomography
DTI	Diffusion Tensor Imaging
EEG	Electroencephalography
FrSBe	Frontal Systems Behavior Scale
FTD	Frontotemporal Dementia
FTLD	Frontotemporal Lobar Degeneration
FTLD-CDR	Frontotemporal Lobar Degeneration Clinical Dementia Rating Scale
GDS	Geriatric Depression Scale
H5PT	Hamasch Five Point Test
LvPPA	Logopenic variant Primary Progressive Aphasia
MCAR	Missing Completely at Random
MICE	Multiple Imputation by Chained Equations
ML	Machine Learning
MMSE	Mini-Mental State Examination
MRI	Magnetic Resonance Imaging
NfL	Neurofilament Light Chain
NfvPPA	Non-fluent variant Primary Progressive Aphasia
PAC	Proportion of Ambiguous Clustering
PET	Positron emission tomography
PPA	Primary Progressive Aphasia
PSP	Progressive Supranuclear Palsy
RMET	Reading the Mind in the Eyes Test
SEA	Social Cognition and motional Assessment

SPECT	Single-Photon Emission Computed Tomography
SuStaIn	Subtype and Stage Inference
SVM	Support vector machine
SvPPA	Semantic Variant Primary Progressive Aphasia
TCT	Three Clap Test
TMT	Trail Making Test
TPJ	Temporo-Parietal Junction
VBM	Voxel-Based Morphometry
WMS-R	Wechsler Memory Scale Revised

1 Introduction

Based on the WHO, dementia affects around 55 million people worldwide and with an aging population this number is predicted to double in the upcoming 30 years (*Dementia*, 2021). Dementia is a syndrome characterized by cognitive decline that can be very heterogeneous depending on the underlying cause, the most common being Alzheimer's Disease (AD). Frontotemporal Dementia (FTD) is thought to be the second most common early-onset dementia after AD and the third most common dementia following AD and Lewy Body Dementia (Bang et al., 2015; Hogan et al., 2016) but exact estimates of its prevalence vary across studies (J. J. Young et al., 2018). Despite its clinical and pathological heterogeneity, it is now recognized to contain three main variants: behavioral variant FTD (bvFTD; in older studies sometimes referred to as FTD) and two primary progressive aphasias (PPA), semantic variant PPA (svPPA) and non-fluent variant PPA (nfvPPA) (*for a review see e.g.* Bang et al., 2015; Olney et al., 2017). Related FTD syndromes include FTD with motor neuron disease, progressive supranuclear palsy (PSP) and corticobasal degeneration (CBS) (Elahi & Miller, 2017).

The first description of a patient with FTD was made by Arnold Pick in 1892 (Arnold Pick, 1892). The patient showed left anterior temporal atrophy and would today be considered for a diagnosis with svPPA (Olney et al., 2017). Mesulam in 1982 described two subtypes of a slowly progressive aphasia, a fluent and a non-fluent variant and sought to differentiate them from AD pathology. Later he grouped his observations under the term PPA (M. M. Mesulam, 2001). The first clinical diagnostic criteria were established in 1994 (Neary et al., 1994) and lead to the formulation of consensus criteria in 1998 (Neary et al., 1998). The current diagnostic guidelines in place for bvFTD and the PPAs, were formulated by Rascovsky et al. (2011) and Gorno-Tempini et al. (2011), respectively. Both guidelines keep a similar structure allowing the clinician to diagnose according to the evidence at hand. Thus, a diagnosis can be expressed hierarchically, depending on certainty levels. In that way, for bvFTD based on observation of cognitive and behavioral symptoms alone a diagnosis of "possible bvFTD" may be given. Further converging evidence from neuroimaging allows for a diagnosis of "probable bvFTD". Lastly, based on additional histopathologic or genetic evidence, a clinician can give the diagnosis of "bvFTD with definite pathology". Similarly, a diagnosis of one of the PPA variants can be given as "clinical", "imaging-supported" or, as diagnosis "with definite pathology". This distinction is in accordance with the gold standard for diagnosis of specific dementia subtypes depending on pathology examined via autopsy postmortem (or rarely via biopsy) (Elahi & Miller, 2017). Alternatively, a definite diagnosis may be given based on genetic testing, as

known gene mutations associated with FTD pathology are thought to have complete penetrance (Lanata & Miller, 2016; Onyike & Diehl-Schmid, 2013).

The neurodegenerative process underlying FTD is frontotemporal lobar degeneration (FTLD) and leads to neuronal dysfunction and later atrophy in frontal and anterior temporal lobes with heterogeneity across patients (Piguet et al., 2011). Histopathologically, neurodegeneration in most FTD cases can be attributed to intracellular tau, TDP-43 or FUS aggregates (Bürger et al., 2017; Elahi & Miller, 2017; van der Ende & van Swieten, 2021) but differs markedly also between FTD variants. BvFTD shows similar prevalence of tau and TDP aggregates, nfvPPA is associated mainly with tau and svPPA instead with TDP aggregates (Piguet et al., 2011). However, a significant proportion of nfvPPA patients has also shown AD pathology (Grossman, 2010). Several gene mutations have been associated to a FTD diagnosis later in life (*for a review see* Deleon & Miller, 2018). Family history of a FTD or other neurodegenerative disease is found in 20-50% of FTD patients (Deleon & Miller, 2018; Elahi & Miller, 2017; Rosso et al., 2003). Yet, most cases are thought to be sporadic with autosomal dominant heritability accounting for approximately 10% of cases with higher heritability of bvFTD and lower heritability of svPPA (Goldman et al., 2005).

Generally, FTD is considered an early-onset dementia meaning that most commonly symptoms start before age 65 (Ratnavalli et al., 2002). Particularly bvFTD shows an early onset of the disease (Coyle-Gilchrist et al., 2016). BvFTD is also the most common accounting for 60% of FTD cases (Hogan et al., 2016; Johnson et al., 2005; Onyike & Diehl-Schmid, 2013).

1.1 FTD and PPA Variants

The syndromes associated with the three core variants of FTD are heterogeneous (Beeldman et al., 2018) and the clinical phenotype depends on type and location of underlying pathology. Initially, bvFTD is associated with behavioral and the PPAs with core language impairments. With increasing disease duration, the neurodegenerative process progresses to include more areas, deficits becoming less specific and often overlapping between FTD variants and with other neurodegenerative diseases, particularly with AD and logopenic variant PPA (lvPPA), considered an atypical AD variant (Bürger et al., 2017; Coyle-Gilchrist et al., 2016).

1.1.1 *BvFTD*

Core changes associated with bvFTD are related to behavior and personality involving most commonly apathy and disinhibition but also mental rigidity and loss of empathy (Bang et al., 2015; Rabinovici & Miller, 2010; Rascovsky et al., 2011; William W. Seeley et al., 2008). Disinhibition is observed as socially inappropriate such as aggressive or even criminal behavior. Apathy is associated to significant impairment in basic and instrumental activities of daily

living (Peet et al., 2021; Piguet et al., 2011). With progression of the disease, dietary changes, most commonly overeating, and decreased hygiene are observed (Bang et al., 2015). Missing insight, particularly prevalent in bvFTD puts the safety of individuals at risk (DeLozier & Davalos, 2016; Rabinovici & Miller, 2010). The described clinical picture of bvFTD makes clear why caregiver burden is high and increased also in comparison to AD (Ljubenkov & Boxer, 2021; Piguet et al., 2011; Riedijk et al., 2006). Uncertainty in diagnosis and prognosis of the disease may further add to the caregiver burden.

Especially in the early stages of the disease neuropsychological assessments may not show any impairments (Devineni & Onyike, 2015; Rabinovici & Miller, 2010) and behavioral observations may be more informative than formal cognitive testing (Kertesz et al., 2003; Warren et al., 2013). Current diagnostic criteria do not require impaired cognitive functioning for a clinical diagnosis (Rascovsky et al., 2011). Instead, the neuropsychological profile described in the diagnostic criteria is a supportive but not necessary feature for diagnosis. In comparison to previous guidelines by Neary et al. (1998) current guidelines have been shown to be more sensitive especially to early stages of the disease. This is probably due to a greater flexibility in how criteria for a diagnosis can be fulfilled (Rascovsky et al., 2011; Rascovsky & Grossman, 2013), thus better reflecting the clinical heterogeneity of bvFTD.

The neuropsychological profile including executive dysfunction in combination with relatively intact episodic memory and visuo-spatial skills has been criticized and calls for revision of these guidelines have been expressed (Michael Hornberger & Piguet, 2012; Piguet et al., 2011). Piguet et al. (2011) criticizes the focus on executive dysfunction for diagnosis as well as the expectation of relatively intact episodic memory function. Disturbances in executive function are indeed associated with (Deuschl et al., 2016; Kramer et al., 2003; Schroeter et al., 2014; Walker et al., 2005) but not thought to be specific for bvFTD (Foran et al., 2021; Overbeek et al., 2020; Reul et al., 2017; Schroeter et al., 2018). Inconsistent findings are common. This could be due on the one hand to the inherent complexity of executive functioning involving a variety of cognitive processes. On the other hand specific tests for executive functioning may differ in sensitivity and ecological validity (Rascovsky & Grossman, 2013). In AD executive functions were found to be at least as impaired as in bvFTD (Foran et al., 2021; Reul et al., 2017). To accommodate for these findings executive functioning may be assessed combining different measures (Rascovsky & Grossman, 2013). Additionally, it has been suggested that the number of errors such as perseverations, intrusions or rule violations, could be more informative than the total scores on neuropsychological tests of executive function (Kamath et al., 2019; Kramer et al., 2003).

More recent literature has shifted from a focus of executive function deficits to the study of social and emotional functions in bvFTD for its delineation from other disorders. This may be performed using informant-based ratings or using newer formal testing (Kamath et al., 2019; Rankin, 2021). Indeed, a meta-analysis on the cognitive profile of bvFTD found the largest effect sizes for social cognition impairment (Beeldman et al., 2018). Additionally, one study found the Ekman 60 faces test to be the only one out of a group of tests assessing a variety of cognitive functions, able to differentiate bvFTD from both psychiatric and neurodegenerative patients (Gossink et al., 2018). Using the same test, another study found impairment in most of bvFTD patients, but difference with AD was insignificant (Reul et al., 2017). These contrasting findings may be the result of grouping very different neurodegenerative disorders in the first study. Using a modified version of the Reading the Mind in the Eyes Tests (RMET), thought to assess mentalizing ability, Schroeter et al. (2018) showed that social cognition tests may be more predictive of bvFTD than executive function tests and suggest inclusion of social cognition deficits in the diagnostic criteria of bvFTD. Additionally, this study showed the advantage of using informant-based ratings of patient's behavior compared to neuropsychological testing in bvFTD. A review suggests usefulness of social cognition assessments to differentiate bvFTD from AD and the other FTLD variants (Rankin, 2021).

In contrast to the neuropsychological profile described in the current guidelines for bvFTD diagnosis claiming relatively spared memory function, pooled evidence presented in a review finds memory deficits in a significant proportion of patients (Michael Hornberger & Piguet, 2012). Compared to controls, one third to half of bvFTD patients may show impairment in memory function (Reul et al., 2017) and in some the extent of memory deficits may be comparable to the severity observed in AD patients (Ahmed et al., 2021; Graham, 2005; M. Hornberger et al., 2010). Memory deficits are usually secondary to other deficits such as behavioral changes or executive dysfunctions (Michael Hornberger & Piguet, 2012) and in most bvFTD patients they are less pervasive than in AD which could explain why it has been overlooked for a long time. Further, clinical heterogeneity and possible inclusion of non-neurodegenerative cases (i.e. phenocopy) could be a reason for inconsistencies across studies (Michael Hornberger & Piguet, 2012).

Two main hypotheses have been proposed to explain the memory dysfunctions in bvFTD. Some suggest that memory deficits may be the result of executive dysfunctions leading to retrieval problems. Accordingly, in a meta-analysis Kamath et al. (2019) finds impairment to be more pronounced for delayed than immediate recall and recognition tasks. Additionally, impairments in autobiographical memory seem to correlate with impairments in executive

function in bvFTD (Michael Hornberger & Piguet, 2012). In AD instead memory problems are thought to be hippocampal-dependent and more related to encoding rather than retrieval. The second hypothesis suggests an involvement of hippocampal pathology in bvFTD. This is supported by neuroimaging studies finding pathology to extend to hippocampal regions in some patients (Beeldman et al., 2018; Gordon et al., 2016; Reul et al., 2017). Both explanations are not mutually exclusive but may instead provide insight into the breadth of the clinical spectrum and underlying pathology. Concluding from the presented findings, episodic memory should not be used as an exclusion criterion for bvFTD diagnosis and might not be useful in a differential diagnosis with AD (Diehl et al., 2005; Michael Hornberger & Piguet, 2012).

With progression of the disease language deficits may arise in bvFTD compared to healthy controls. Coyle-Gilchrist et al. (2016) for example found impaired language function in up to 70% of patients. Most consistently impairments are found on tests of semantic or phonologic fluency (Bürger et al., 2017; Sitek et al., 2015), which could be the result of neurodegenerative progression into language areas, or instead result from impaired executive functioning. Additionally, impaired naming has been reported (Bang et al., 2015; Kamath et al., 2020). However, in general it is important to keep in mind that the usefulness of neuropsychological tests especially in early disease stages is limited and behavioral assessments such as informant-based questionnaires or newer developed performance tests may be more informative (Schroeter et al., 2018).

BvFTD is sometimes referred to as frontal variant of FTD in accordance with the neurodegenerative process being most prominent in bilateral frontal and prefrontal regions extending to the anterior temporal regions (Gordon et al., 2016; Pan et al., 2012; Rabinovici & Miller, 2010; Rascovsky et al., 2011). Impaired functioning of the salience network (Moguilner et al., 2021) including frontal lobe, anterior cingulate cortex, insula, amygdala, medial thalamus and ventral striatum, has been found in bvFTD and was related to behavioral symptoms (Rankin, 2021) and clinical deficits in social cognition such as mentalizing and emotion recognition (Gordon et al., 2016). Pooling results from magnetic resonance imaging (MRI) and positron emission tomography (PET) studies, a meta-analysis on the three variants of FTD, suggests a triple dissociation in areas underlying the clinical syndromes. For bvFTD, seven clusters including frontomedian areas, thalamus, left superior frontal sulcus and right anterior insula showed significant changes (i.e. atrophy and hypometabolism) compared to controls (Schroeter et al., 2007). Similar results were found in a second, later meta-analysis (Schroeter et al., 2014). Drawing from functional neuroimaging studies in healthy controls, the authors state that impaired functioning in these areas might explain the pattern of symptoms observed

in bvFTD, namely changes in executive and social functioning, apathy, disinhibition, and a loss of empathy. Findings did not however support the involvement of areas related to mentalizing abilities (Schroeter et al., 2014). On the neural level, the neurodegenerative process in bvFTD may be particularly focused on von Economo neurons found in anterior cingulate, insular and orbitofrontal regions (William W. Seeley et al., 2008). Interestingly, these neurons were previously related to social cognition, lending further support to the conception of bvFTD symptoms relating to behavioral and socioemotional changes (E.-J. Kim et al., 2012; William W. Seeley et al., 2006).

1.1.2 PPA

The PPAs are characterized by main changes in language function (Deuschl et al., 2016; Elahi & Miller, 2017; Neary et al., 1998). A diagnosis of PPA requires language difficulties to be progressive over time and changes in language function to be the most prominent cause for impairments in activities of daily living (Gorno-Tempini et al., 2011). After a generic diagnosis of PPA, further evidence may support a specification as one of three possible variants of PPA: nfvPPA and svPPA pertaining to the FTD spectrum or lvPPA considered as atypical form of AD (M.-M. Mesulam et al., 2021). Some patients may not be easily classified in one of the three variants (Gil-Navarro et al., 2013; Grossman, 2010; Harris et al., 2013; Leyton et al., 2014; Utianski et al., 2019) and literature has suggested existence of more variants and mixed pathology (M. M. Mesulam, 2001).

1.1.2.1 NfvPPA. Core symptoms of nfvPPA are effortful speech or agrammatism with some patients showing or with progression developing both deficits (Gil-Navarro et al., 2013; Leyton et al., 2014). Characteristic speech apraxia leads to slowed speech production rate and omission of words is described as telegraphic style. Apraxia may extend also to other, non-speech movements of mouth and face in some patients (Marshall et al., 2018). With progression, patients may develop complete mutism (Grossman, 2010). Compared to bvFTD, svPPA and AD, patients with nfvPPA are only rarely affected by loss of insight (DeLozier & Davalos, 2016) and thus suffer from their difficulties, commonly resulting in frustration (Marshall et al., 2018) and mild depression (Sitek et al., 2015). Over time, decreased motivation and apathy may emerge (Grossman, 2010).

The most useful neuropsychological tests involve reading and repetition tasks as well as confrontation naming. Phonologic or articulatory errors, omissions and simplifications are commonly observed and are more pronounced with increasing stimulus complexity and length (Macoir et al., 2021). For the assessment of grammar, spontaneous speech or picture description tasks may be useful (Gorno-Tempini et al., 2011; M. Henry & Grasso, 2018). While in early

stages of the syndrome, patients may rely on preserved written abilities, dysgraphia commonly develops over time (Bürger et al., 2017; Marshall et al., 2018). Comprehension and semantic knowledge are usually spared.

Neuroimaging findings suggest primary degeneration in left fronto-insular and temporal areas including inferior frontal and superior temporal gyrus (Dave et al., 2020; Gordon et al., 2016; Gorno-Tempini et al., 2011; Ruksenaite et al., 2021). A meta-analysis found all abnormal clusters from MRI and PET imaging compared to controls to be located in the left hemisphere (Schroeter et al., 2007). Degeneration in this peri-sylvian network may be visible on structural neuroimaging as enlargement of the left Sylvian fissure (Grossman, 2010; Marshall et al., 2018). An impaired link between language and motor networks may disturb generation of motor output and explain the symptoms observed (Ruksenaite et al., 2021; William W. Seeley et al., 2009). With progression, neurodegeneration spreads to more anterior and pre-frontal regions as well as to parietal regions including the basal ganglia (Gordon et al., 2016; Schroeter et al., 2007).

1.1.2.2 SvPPA. In svPPA loss of semantic knowledge results in core problems in confrontation naming (i.e. anomia) and word comprehension (Gil-Navarro et al., 2013; Gorno-Tempini et al., 2011; Kramer et al., 2003; M. M. Mesulam, 2001). Deficits are more pronounced for less common words (M. Henry & Grasso, 2018). Speech is fluid but becomes progressively empty in content, reducing to platitudes. During reading, patients may make regularization errors in pronunciation due to loss of associated meaning (Macoir et al., 2021). With disease progression, loss of semantic knowledge extends to non-verbal domains and may develop to object agnosia and prosopagnosia (Bang et al., 2015; Rabinovici & Miller, 2010; Ruksenaite et al., 2021). Prosopagnosia may be particularly common in a right-hemisphere dominant variant of svPPA (Sitek et al., 2015). Non-language and behavioral symptoms of svPPA may be similar to those observed in bvFTD (Rabinovici & Miller, 2010; Ruksenaite et al., 2021) and insight may be reduced (DeLozier & Davalos, 2016).

Neurodegeneration in svPPA usually develops in one hemisphere, most commonly the dominant one, and spreads to the contralateral one over time. SvPPA is also referred to as temporal variant with pathology most pronounced in anterior temporal regions including rhinal, hippocampal and temporal pole regions that are part of the semantic network (Gordon et al., 2016; Ruksenaite et al., 2021). Involvement of the amygdala and subcallosal area may explain deficits in socioemotional processing (Schroeter et al., 2007). Over time, degeneration progresses to more anterior and posterior regions and to the contralateral hemisphere.

1.1.2.3 LvPPA. Diagnostic criteria for lvPPA require deficits in word retrieval during

spontaneous speech and impaired sentence repetition (Gorno-Tempini et al., 2011). Disturbed phonemic processing influences speech production. This is reflected in phonological errors in both spontaneous speech and structured tasks of repetition or naming. Difficulties in repetition may be intensified by phonological short-term memory deficits, with greater impairments for longer sentences (Gil-Navarro et al., 2013). Phonological dyslexia and dysgraphia are present for both words and nonwords (M. Henry & Grasso, 2018; Macoir et al., 2021). Generally, comprehension and semantic knowledge are spared but impaired phonemic processing may impede parsing of speech sounds into meaning leading to deficits in understanding degraded speech (Ruksenaite et al., 2021).

Originally only two variants of PPA were recognized: semantic and non-fluent variants referred to as semantic dementia and progressive non-fluent aphasia, respectively (Mesulam, 2001). LvPPA is clinically most similar to nfvPPA but differs in underlying pathology. While nfvPPA is most associated with tau-aggregates, lvPPA shows AD pathology, i.e. tau and amyloid deposition (Deuschl et al., 2016). PET scans to detect amyloid pathology are usually positive and analysis of cerebrospinal fluid (CSF) shows heightened tau and reduced amyloid levels (Grossman, 2010; Henry & Gorno-Tempini, 2010). Neurodegeneration in lvPPA shows a more posterior profile with involvement of the parietal lobes. This is in accordance with major language deficits and other non-verbal deficits observed such as limb apraxia and dyscalculia (M. L. Henry & Gorno-Tempini, 2010). A key area affected by lvPPA pathology may be the temporo-parietal junction (TPJ) (Leyton et al., 2014; Rohrer et al., 2013) involved in auditory phonemic transformations and phonological short-term memory (Gil-Navarro et al., 2013). Over time, pathology spreads to the contralateral side but rate of atrophy remains higher for the left hemisphere compared to the right one, intensifying asymmetry with progression (Rohrer et al., 2013).

1.2 Distinction of the Three PPA Variants with Neuropsychological Assessments

In contrast to bvFTD, in PPA cognitive abnormalities usually predominate over behavioral ones facilitating a clinical diagnosis using common neuropsychological tests (Devineni & Onyike, 2015). The primary deficit being language, the use of various language assessments is well established and required for differentiation of the three syndromes. Especially the assessment of confrontation naming, using for example the Boston naming test, may be useful in differentiating svPPA from the other two variants although anomia is common in all PPA variants (Gil-Navarro et al., 2013). A combination of a lexical-semantic and a syntactical test may further differentiate the three variants with nfvPPA showing primary impairment in the test of syntax in accordance with agrammatism and svPPA showing impaired

lexical-semantic processing. Performance in patients with lvPPA was relatively spared in both tasks (M. Mesulam et al., 2009). A repeat and point task showed particularly informative for the distinction of svPPA and the two other variants (Seckin et al., 2022). Based solely on clinical features, the distinction between nfvPPA and lvPPA may be the most challenging of the three (Leyton et al., 2014).

Although, not the primary deficit, memory is commonly impaired in the PPAs. Interestingly, findings of memory dysfunction were already observed in part of the earliest cases described by Pick (A. Pick, 1901, 1904; Arnold Pick, 1892) and a systematic clinicopathological study by Constantinidis et al. (1974 *as cited in* Michael Hornberger & Piguet, 2012) found a majority of patients with Pick's disease to present with memory deficits. Those patients showed atrophy in the medial temporal lobe including the hippocampus and a significant correlation between extent of atrophy and clinical impairment. Another group showed more prefrontal atrophy which was related to memory deficits particularly in the correct ordering of past events. Further, behavioral deficits similar to those observed in bvFTD may arise, most commonly in svPPA but also in the other variants (Bang et al., 2015; Coyle-Gilchrist et al., 2016). Thus, although main deficits are language-related, the syndromes are not restricted to language symptoms but instead encompass a variety of symptoms and current diagnostic criteria may not capture their diversity (Ruksenaite et al., 2021).

1.3 Correct Diagnosis is Difficult

Several factors make correct and early diagnosis of FTD and its variants complicated and years may pass by until correct diagnosis is reached (Beber & Chaves, 2013; Coyle-Gilchrist et al., 2016; van Vliet et al., 2013). Neurodegenerative dementia syndromes are characterized by an insidious onset, making changes hard to recognize for close others (Convery et al., 2019; Warren et al., 2013). Additionally, early onset dementias may pose a particular diagnostic challenge due to their heterogeneity, scarcity and low familiarity in the society (Devineni & Onyike, 2015; Rossor et al., 2010).

FTD syndromes and even more so PPA variants are relatively rare (Hogan et al., 2016; Ratnavalli et al., 2002) leading several authors to suggest possible underdiagnosis (Bertoux et al., 2012; Knopman & Roberts, 2011; Marshall et al., 2018; Rascovsky et al., 2011; J. J. Young et al., 2018). However, a risk for high number of false positive diagnoses has also been reported for bvFTD (Shinagawa et al., 2014) and the doubt for possible overdiagnosis of PPA in recent years has been expressed (M.-M. Mesulam et al., 2021). While late diagnosis may prevent patients from receiving early treatment or therapeutical intervention, false positive diagnoses cause harm to the individual and their surrounding and need to be avoided (van Vliet et al.,

2013).

The PPA variants compared to bvFTD may be easier to diagnose as deficits are primarily cognitive rather than behavioral (Devineni & Onyike, 2015). However, significant heterogeneity within each FTD variant (Beber & Chaves, 2013; Beeldman et al., 2018; Gordon et al., 2016; Harris et al., 2013; Piguet et al., 2011; A. L. Young et al., 2018) and considerable overlap between variants (*e.g.* Bang et al., 2015) complicate correct differentiation. In part responsible for the heterogeneity could be the rarity of studies with pathologically confirmed cases (e.g. Foran et al., 2021; Muñoz-Neira et al., 2019; van't Hooft et al., 2021) making patient groups less homogeneous and possibly contaminated with misdiagnoses of other neurodegenerative or non-neurodegenerative diseases (M. Hornberger et al., 2010; Valente et al., 2019). With progressive neurodegeneration greater overlap in the clinical pictures of different syndromes is observed which further complicates accurate differentiation (Dave et al., 2020; H. J. Rosen et al., 2000).

Within FTD syndromes, overlap between svPPA and bvFTD is particularly prominent (Coyle-Gilchrist et al., 2016; Kamath et al., 2019). Specifically, right-sided dominant svPPA has been associated with behavioral symptoms typical of bvFTD (Marshall et al., 2018; W. W. Seeley et al., 2005). Additionally, existing diagnoses may not be sufficient to explain pathologic and clinical characteristics observed in all patients. This may be particularly true for the PPAs (Grossman, 2010; Kamath et al., 2020). Instead, mixed pathologies and comorbidities make a clear delineation between disorders difficult (M.-M. Mesulam et al., 2021; W. W. Seeley et al., 2005). Newer suggestions include the description of individuals on a multidimensional spectrum rather than within a specified diagnostic category. This has been proposed for FTD syndromes (Murley et al., 2020) and for PPA variants (Ingram et al., 2020) and may have benefits for therapeutical interventions (Ingram et al., 2020; Murley et al., 2020).

The most common misdiagnosis of FTD syndromes is AD or an atypical AD variant such as lvPPA (Ahmed et al., 2021; Beber & Chaves, 2013; Bürger et al., 2017; Seo et al., 2018). Risk for misdiagnosis is particularly high for late-onset FTD while early onset AD may be at risk for misdiagnosis as FTD. Thus, although the separation of early and late onset dementia relies on an arbitrary cut-off at 65 years (Devineni & Onyike, 2015; Rossor et al., 2010), the recognition of bvFTD and AD as early and late-onset dementias respectively may influence diagnostic tendencies. Additionally early onset AD may more commonly show atypical, non-amnestic symptoms including executive and language deficits (Tellechea et al., 2018) and personality changes (Josh D. Woolley et al., 2011). The challenge of accurate differential diagnosis is that finding significant group differences is not sufficient. Instead, large

effect sizes and low percent overlap are required for a test to be sensitive and specific for one diagnosis over another (Hutchinson & Mathias, 2007). In a meta-analytic study by Hutchinson & Mathias (2007) all cognitive measures that showed a significant difference between AD and FTD groups, showed overlap ranging from 32-48%. Overlap in cognitive deficits with AD was also reported separately for the PPAs (Kramer et al., 2003; Lecerf et al., 2020) and for bvFTD (Baborie et al., 2012; Buhl et al., 2013; Musa et al., 2020; Peet et al., 2021; Reul et al., 2017; Walker et al., 2005). Additionally, nfvPPA and lvPPA are at a particularly high risk of being confounded due to similarity in clinical phenotypes (Bürger et al., 2017; M.-M. Mesulam et al., 2021).

Overlap has also been reported for pathology underlying the different syndromes. Rabinovici & Miller (2010) state that 10-30% of patients with an FTD diagnosis may show AD pathology at autopsy. With increasing age the presence of AD pathology alongside FTD pathology becomes more likely and patients with amyloid-positive PET scans were found in all three FTD variants (Gordon et al., 2016). Additionally, amyloid-negative lvPPA patients were reported and have been linked to the presence of TDP-43 aggregates typical of FTLD pathology (Matias-Guiu et al., 2019).

1.4 Misdiagnosis With Primary Psychiatric Disorders is Common

A considerable proportion of patients with neurodegenerative diseases is first diagnosed with a primary psychiatric disorder. This is particularly true for FTD (Josh D. Woolley et al., 2011) and has the adverse effect of delaying duration until correct diagnosis is reached (Rosness et al., 2008). Differential diagnosis is difficult due to patients commonly showing psychiatric symptoms. In fact, psychiatric symptoms are now considered a hallmark of dementia (Collins et al., 2020; Desmarais et al., 2020; Gambogi et al., 2019; Mulder-Heijstra et al., 2021). In a systematic review on the prevalence of apathy, depression and anxiety, Collins et al. (2020) found apathy to be most common in bvFTD in accordance with diagnostic criteria (Rascovsky et al., 2011). Instead, more than 70% of patients with lvPPA and svPPA showed anxiety and depression, respectively. Prevalence of psychiatric symptoms in nfvPPA was more variable but up to half of the patients showed depression.

People with primary psychiatric disorders on their side are not preserved from cognitive changes (Overbeek et al., 2020). Generally, cognitive deficits are greater in dementia syndromes than in primary psychiatric disorders (Gambogi et al., 2019). However, during mood episodes patients with bipolar disorder showed greater deficits compared to patients with bvFTD (Simjanoski et al., 2021). Bipolar disorder, schizoaffective disorder and schizophrenia have all been related to impaired verbal memory, verbal fluency and executive function of

different severities (Baez et al., 2019; Ducharme et al., 2020; Gambogi et al., 2019).

A complex relationship between dementias and primary psychiatric disorders has been reported in the literature (Desmarais et al., 2020) with an increased risk for dementia following a diagnosis of psychiatric diseases. For example patients with bipolar disorder showed an increased risk for a dementia syndrome later in life (Roman Meller et al., 2021; Simjanoski et al., 2021), a specific link was found for depression and FTD (Kuring et al., 2018) and patients with PTSD were associated with a greater prevalence of svPPA (Bonanni et al., 2018). One hypothesis for the observed relationship suggests that psychiatric diseases may pose an endogenous or reactive risk for the development of a dementia by producing changes in inflammatory processes or lifestyle, respectively (Gambogi et al., 2019; Kuring et al., 2018). Another hypothesis suggests that psychiatric symptoms may represent a prodromal stage of a neurodegenerative disease (Caixeta & Caixeta, 2011; Roman Meller et al., 2021; Josh D. Woolley et al., 2011). Further, some authors have hypothesized that cognitive decline, similar to the known dementias, may represent a late stage of psychiatric disorders (Gambogi et al., 2019). Possibly, psychiatric disorders may thus be associated with disorder-specific dementia syndromes. These hypotheses are not mutually exclusive (Josh D. Woolley et al., 2011). One reason for the complex interplay between FTD syndromes and psychiatric disorders, could be common genetic risk factors involved in both as was reported for example for psychotic and bipolar syndromes (*for bvFTD:* Ducharme et al., 2020; Lanata & Miller, 2016; *for FTD*: Roman Meller et al., 2021).

Symptomatic overlap with primary psychiatric disorders is particularly prominent in bvFTD (Peet et al., 2021). In the early stages of bvFTD patients may not show cognitive symptoms and neuropsychiatric symptoms such as apathy, disinhibition and compulsions are not just common (Ducharme et al., 2020) but also part of the diagnostic criteria for bvFTD (Rascovsky et al., 2011). In combination with a younger age of symptom onset these behavioral changes make a referral to a psychiatrist rather than a neurologist common (Lanata & Miller, 2016). BvFTD being a relatively rare disease compared to psychiatric disorders, symptoms of bvFTD may then be misinterpreted as pertaining to a psychiatric disorder, most commonly major depressive disorder, bipolar disorder or schizophrenia (Gambogi et al., 2019; Lanata & Miller, 2016; Josh D. Woolley et al., 2011; Joshua D. Woolley et al., 2007). In a retrospective study by Josh D. Woolley et al. (2011) for example, half of the patients with a diagnosis of bvFTD had first received a psychiatric diagnosis. Further, bvFTD patients with a specific gene mutation (C9orf72 gene expansion) are reported to show high prevalence of psychotic symptoms and are at risk of being misdiagnosed with schizophrenia (Lanata & Miller, 2016).

Psychiatric symptoms are common also in PPAs and were already noted in the case report by Mesulam (1982). Misdiagnosis with a primary psychiatric disorder may be especially prevalent in (right-sided dominant) svPPA (W. W. Seeley et al., 2005). Single cases have been reported that may suggest that PPA has a prodromal stage of psychiatric symptoms (Caixeta & Caixeta, 2011). However, psychiatric symptoms in PPAs are more common with progression rather than at onset of the disease and most studies therefore focus on the emergence of psychiatric symptoms following diagnosis of PPA (Mulder-Heijstra et al., 2021).

Studies investigating the overlap between psychiatric disorders and FTD are difficult and often rely on small cohorts (Modirrousta et al., 2013). Further consortium studies may be necessary to elucidate their link and facilitate clear diagnosis and appropriate treatment. A differential diagnosis may be helped by longitudinal observation with a progressive decline in cognitive function suggesting a neurodegenerative disease. The analysis of neurofilament light chain (NfL) levels in blood serum or CSF shows promising differentiation between FTD and primary psychiatric disorders (Ducharme et al., 2020; Vijverberg et al., 2017). A battery of social and emotional cognition assessments was suggested to perform well in differentiating bvFTD from major depressive disorder (Bertoux et al., 2012). Another study found use of verbal function tests, particularly those including semantic components, to be helpful for differentiation of bvFTD with primary psychiatric disorders (Overbeek et al., 2020).

1.5 Combining Multimodal Evidence

Differential diagnosis may be aided with the use of neuroimaging techniques, the analysis of biofluids and genetic testing in cases of known family history. Multimodality may be key for an accurate diagnosis. In clinical practice the use of neuropsychological assessments, patient- and informant-based history taking in combination with volumetric MRI are most common to diagnose a patient with FTD (Dev et al., 2021; Piguet et al., 2011). The diagnostic criteria include evidence from volumetric imaging, acquired by MRI or computed tomography (CT), or functional imaging, by PET or single-photon emission CT (SPECT) for higher confidence in the diagnosis (Gorno-Tempini et al., 2011; Rascovsky et al., 2011).

1.5.1 *Neuroimaging*

Neuroimaging may aid to exclude non-neurodegenerative causes of dementia (Warren et al., 2013) or differentiate FTD from other neurodegenerative syndromes, most commonly forms of AD. Compared to MRI, the use of PET may aid earlier diagnosis as metabolic changes may closely correspond to but have the advantage of preceding atrophy (Schroeter et al., 2007; Yeo et al., 2013). Specifically, the use of FDG-PET has shown good discriminability between FTD and AD (Dev et al., 2021; Foster et al., 2007) and might be particularly informative in

bvFTD as neuropsychological assessments may be of limited help (Guillén et al., 2020). In the search of "metabolic signatures" associated with different dementia syndromes, Dave et al. (2020) also finds large overlap between syndromes complicating their delineation. If ambiguity persists between a diagnosis of AD and FTD syndromes, amyloid PET scans may help shed a light on the underlying pathology. As seen previously, this is of particular interest for differentiating bvFTD from AD and nfvPPA from lvPPA due to the difficulty of distinguishing them clinically (Gordon et al., 2016). Newer developments in research also highlight the potential use of tau-PET (Hall et al., 2017) but its clinical utility still needs to be shown (Ruksenaite et al., 2021). Electroencephalography (EEG), a cheap imaging technique that could be easily applicable in clinics, has so far not proven useful for FTD (Livinț Popa et al., 2021; Micanovic & Pal, 2014). Other research has investigated the added value of diffusion tensor imaging (DTI). This may be interesting for early diagnosis as white matter changes are thought to precede grey matter changes (Gordon et al., 2016). To sum up, it is important to keep in mind the clinical context of these applications. While use of FDG-PET in clinics is increasing, it is not the primary imaging method also due to costs and invasiveness. More sophisticated methods such as (resting state) functional MRI, arterial spin labelling and DTI remain completely reserved to the use in the research setting (Dev et al., 2021; Elahi & Miller, 2017; Vernooij et al., 2019).

1.5.2 *Fluid Biomarker*

For AD the analysis of CSF to determine amyloid-beta, phosphorylated-tau and total tau levels is an established biomarker predictive of the disease (Deuschl et al., 2016). Its use for differential diagnosis of FTD and AD however may be limited as it cannot confirm or rule out FTD and amyloid pathology is also involved in healthy aging. So far, no FTD-specific fluid-biomarker exists in clinical practice (Denk et al., 2018; van der Ende & van Swieten, 2021). One promising candidate is NfL, that may be detected in the CSF or, less invasively in the blood and shows possible use for differentiation of FTD from primary psychiatric and other neurodegenerative diseases (Bridel et al., 2019; van der Ende & van Swieten, 2021). One study suggests that serum NfL levels could indicate disease intensity, meaning that it is associated with faster disease progression and correlates significantly with rate of frontal lobe atrophy (Rohrer et al., 2016). In two studies Steinacker et al. (2017, 2018) investigates the use of serum NfL as a marker of disease progression in the PPA variants and bvFTD, respectively. In bvFTD, NfL serum concentrations correlated significantly with brain atrophy and functional impairment, as assessed by the clinical dementia rating scale (CDR), FTLD-CDR and mini-mental state examination (MMSE). Similar results were found for the PPAs but correlation with

functional impairment were limited to the CDR only. However, serum NfL concentrations were able to distinguish lvPPA from nfvPPA and svPPA with a similar performance as other CSF marker.

Despite the difficulties faced in the diagnosis of dementia syndromes it is important to keep in mind the importance of continuing to search for more accurate criteria and possible biomarkers. An early and accurate diagnosis is crucial for appropriate and early interventions to the patient and caregivers (Romero & Wenz, 2002), powers clinical trials and allows for prediction about disease progression (Gordon et al., 2016; M. L. Henry & Gorno-Tempini, 2010). So far, no disease-modifying treatment is available for FTD. Pharmacological treatments aimed at reducing cognitive and behavioral symptoms are only rarely tested specifically for FTD and instead may be based on evidence from other diseases such as AD. Positive effects from drugs observed in AD may however not be reflected when used by FTD patients and in some cases even worsen the course of FTD (J. J. Young et al., 2018). A review on available pharmaceutical treatments, development of new treatments as well as the hurdles in the research of treatments for FTD highlights the potential of targeting tau pathology (Panza et al., 2020).

To summarize, no in vivo gold standard for the (differential) diagnosis of FTD exists. Diverse techniques, neuropsychological tests or interviews, structural or functional neuroimaging as well as the analysis of biofluids can inform clinical decisions. No single method outperforms the others while the use of multiple and sophisticated methods may not be feasible in clinical practice due to time and cost constraints but also due to invasiveness. Considering patient history, symptomatology and, if available, informant reports, a specific clinical question may be formulated that may help deciding which technique could be most appropriate. Combining the evidence from multiple modalities is necessary to ensure a sufficiently confident diagnosis but clinical feasibility forces research to find time- and cost-efficient answers. One large potential comes from the use of machine learning (ML) techniques allowing for computer-aided diagnoses or prognoses. Especially, in the case of FTD, where heterogeneity between cases of the same syndrome and overlap with other disorders is a big problem, ML could aid in finding patterns within data rather than looking for single predictive tests.

1.6 Applications of ML for Medical Questions

ML, one of the major branches of artificial intelligence, is concerned with learning specific patterns in data with a major goal of making predictions on unseen data (Kononenko, 2001; Shailaja et al., 2018; Wiens & Shenoy, 2018). Being inherently multivariate, ML tries to find complex relationships between many, sometimes hundred or thousand covariates (Wiens

& Shenoy, 2018). Ideally, by describing the data or allowing for accurate predictions, it uncovers new relationships and knowledge (Yanase & Triantaphyllou, 2019). Common problems investigated include computer vision and language processing in written or spoken form (Qayyum et al., 2021; Wiens & Shenoy, 2018). Possible applications of ML are diverse and include the financing and security sector but also the healthcare sector (Shailaja et al., 2018).

Most interesting for the current study are the applications of ML on the lowest level of the healthcare sector, that is supporting healthcare providers, usually doctors, in diagnostic and treatment decisions (Shahid et al., 2019). The tasks in which ML can support clinicians, include prognosis, diagnosis, treatment and in the clinical workflow (Qayyum et al., 2021). For example, prognosis includes predictions on whether a patient is at risk for a specific disorder or disease (Callahan & Shah, 2017; Deo, 2015; Koutsouleris et al., 2009). Such screening may allow for prevention and earlier treatment (Wiens & Shenoy, 2018).

ML techniques can be categorized in supervised, unsupervised, semi-supervised and reinforcement learning methods with the first two being most common for applications in brain disorders. Common supervised learning algorithms applied to clinical problems include decision trees, support vector machines (SVM) and artificial neural networks (Christodoulou et al., 2019). They are trained on labelled data, whether the label is a group or a continuous score (Qayyum et al., 2021). Thus, in supervised learning the outcome to be predicted is known (Callahan & Shah, 2017; Deo, 2015; Shailaja et al., 2018). Instead in unsupervised learning, the data does not contain labels that need to be matched by predictions but algorithms group data by similarity defined by a specific criterion. Common unsupervised learning techniques are clustering methods (e.g., hierarchical clustering or k-means clustering) or dimensionality reduction techniques such as principal component analysis. Unsupervised learning studies are used to detect outliers or subgroups (Bhardwaj et al., 2017; Qayyum et al., 2021). It is a way to investigate disease heterogeneity using a data-driven approach (Habes et al., 2020) and compared to supervised learning models they are less hypothesis-driven as they do not assume specific classes in the data (Feczko et al., 2019).

1.6.1 *Advantages of ML Applications in Healthcare*

With the emergence of electronic health records and continuous data acquisition via wearable devices, medical research has access to large amounts of data from various modalities (Callahan & Shah, 2017; Qayyum et al., 2021). Further, an increasing number of consortia and multi-centric studies allows for acquisition of data from larger cohorts and often over a long term (*e.g.,* Lei et al., 2019; Otto et al., 2011; Postema et al., 2019). This offers new possibilities

especially for the investigation of rare conditions in which sample sizes for classic case-control studies are rather small. However, despite the great advantage of increasing data-availability, this data may be unstructured (Bhardwaj et al., 2017) and noisy (Kononenko, 2001) and with its high dimensionality not suitable for classic statistical analysis. Exemplifying this, the analysis of neuroimaging data is often performed by voxel-wise comparison using voxel-based morphometry (VBM) which assumes independence of voxels (Khvostikov et al., 2018). Instead, most brain disorders are thought to be network-level disorders. With its multivariate approach, ML techniques can handle large, complex data sets and provide new solutions for the analysis of such data to understand the heterogeneity underlying many disorders (Koutsouleris et al., 2009; Orrù et al., 2012; Yanase & Triantaphyllou, 2019). Additionally, algorithms can be continuously updated when new data is acquired (Bhardwaj et al., 2017).

Medical decisions are made on the level of the individual. Group-level statistics may not be appropriate for such decisions. A study by Scarpazza et al. (2013) demonstrates that comparing single patients to a control group is not appropriate as it is prone to false positives due to individual heterogeneity, even when strict criteria are applied. Such studies assume the single case to represent a "typical" patient which is rarely true. Instead, ML techniques make predictions on the level of the individual (Lei et al., 2019; Orrù et al., 2012) and may offer new opportunities towards the goal of a patient-centered medical care (Bhardwaj et al., 2017; Deo, 2015; Shahid et al., 2019). This in turn may allow for earlier and better treatment as well as for the development of new treatments targeting specific subgroups of patients (Deo, 2015).

Further advantages of ML applications are reduced costs and diffusion of expert knowledge. Reduced costs are ensured by more efficient administration of time and resources (Shailaja et al., 2018). This could be especially useful for low- and middle-income countries (Musa et al., 2020). Diffusion of expertise can be enacted by allowing the use of these expert systems to health care professionals that may be less trained or less aware of possible differential diagnoses and treatment options. Additionally, they may help in working conditions in which fatigue, stress and distraction may otherwise lead to error prone decisions by health professionals (Yanase & Triantaphyllou, 2019). However, it is necessary to stress that these decision-support systems should never be used blindly. The goal of clinical ML applications is to support decisions rather than to establish full automation (Deo, 2015; Qayyum et al., 2021; Topol, 2019; Yanase & Triantaphyllou, 2019). An increasing reliance on ML comes with increasing risks for security (Qayyum et al., 2021) and ethical problems, such as whom to be held accountable in the case of a mistake (Grote & Berens, 2020; Vayena et al., 2018) and possible biases with algorithms performing better for some societal groups than others (I. Y.

Chen et al., 2021). To avoid these concerns, it is important that clinical experts are involved in the development of decision-support systems and that decisions made are maximally transparent meaning that the reasons underlying predictions can be retraced (Ahmad et al., 2018; Kononenko, 2001; Wiens & Shenoy, 2018). This so called explainability may generally decrease with increasing model complexity and is one reason for simple models to be favored.

1.6.2 *Points of Caution*

While one strength of ML models is that they are able to find complex relationships between great number of variables, maximizing simplicity of the models is a main goal (Deo, 2015; Kononenko, 2001). Additional to ensuring explainability, simplicity also reduces the time and costs needed for acquisition of the data. Some types of data such as genetic, functional imaging or pathologic data may be used in research but are not readily available in the clinical field due to costs, limited time or invasiveness (Dev et al., 2021). Thus, a trade-off between precision and simplicity exists to ensure usefulness in practice. To be suitable for clinical applications a decision-support system needs to offer added value for diagnostic precision, exceeding current diagnostic accuracies, while at the same time maximizing simplicity. This may be helped by feature selection finding the best subset of variables (Deo, 2015; Yanase & Triantaphyllou, 2019) and regularization algorithms which penalize more complex models (P.-H. C. Chen et al., 2019; Wiens & Shenoy, 2018).

One problem that may be faced when establishing decision support-systems using supervised learning techniques, is that the quality of their predictions is highly dependent on label quality (P.-H. C. Chen et al., 2019; Qayyum et al., 2021; Wiens & Shenoy, 2018). Taking the example of FTD, in which accurate diagnosis is difficult, there may be uncertainty in the diagnoses of patients within a dataset. Predictions made by a ML algorithm trained on this dataset can then be only as good as the diagnostic labels in the dataset it trained on.

1.6.3 *Use of ML Techniques for Research About FTD*

The last decade has seen a great interest in the application of ML techniques for FTD patient cohorts. Most commonly these studies use supervised learning algorithms such as SVM (*e.g.,* Bron et al., 2014; Kloppel et al., 2008; Meyer et al., 2017). Some studies use a variety of ML algorithms to compare their ability to correctly predict the class that participants belong to (*e.g.,* Ficiarà et al., 2021; Garcia-Gutierrez et al., 2022; Lage et al., 2020). Most studies make use of volumetric data from MRI (*e.g.,* Chagué et al., 2020; Harper et al., 2016; Ma et al., 2021) while some use other neuroimaging tools (*e.g., Arterial Spin Labelling:* Bron et al., 2014, 2017; *resting-state fMRI:* Bouts et al., 2018; Donnelly-Kehoe et al., 2019; Feis et al., 2019; Moguilner et al., 2021; Premi et al., 2016; *QEEG:* Garn et al., 2017; *MEG:* Shaw et al., 2021; *PET or*

SPECT: Abdi et al., 2012; Horn et al., 2009; Xia et al., 2014), cognitive assessments (*e.g.,* Cope et al., 2017; Garcia-Gutierrez et al., 2022; Zimmerer et al., 2020), data from biofluid samples (*e.g.,* Ficiarà et al., 2021; Lin et al., 2020), eye tracking data *(e.g.,* Lage et al., 2020; Primativo et al., 2017) or transcranial magnetic stimulation (*e.g.,* Benussi et al., 2020). Some studies also investigate the added value of multimodal classifiers combining neuroimaging with neuropsychological data (*e.g.*, Dottori et al., 2017; Wang et al., 2016; Zhutovsky et al., 2019). Multi-centric studies are common and are able to demonstrate the robustness of the developed algorithms for classification (*e.g.,* Bachli et al., 2020; Benussi et al., 2020; Young et al., 2018). To evaluate the performance of the developed classifiers, most studies report performance metrics without any direct comparison. However, some studies compare the classification accuracy of the ML algorithm with the performance of radiologists (*e.g.,* Chagué et al., 2020; Harper et al., 2016; Horn et al., 2009). A group of studies focuses on cohorts with a known genetic risk for developing FTD. For example two longitudinal studies have investigated the prognostic value of MRI for predicting symptom onset in participants carrying a mutation associated with FTD (Feis, Bouts, de Vos, et al., 2019; Jiskoot et al., 2019).

Studies using unsupervised algorithms commonly use clustering to investigate natural grouping of patients and its correspondence with the diagnostic classes. For example based on FDG-PET data, patients with a diagnosis of PPA grouped in five clusters (Matias-Guiu et al., 2019). In a different study, four clusters emerged from a cohort of lvPPA and nfvPPA, based on a language assessment and the presence of amyloid pathology on a PET scan (Leyton et al., 2014). Clustering of bvFTD patients also indicated the possibility of several subtypes (Bruun et al., 2019; Ranasinghe et al., 2016; Whitwell et al., 2009). As multimodal studies are relatively rare the possible correspondence of the clusters found across different studies remains to be investigated (Habes et al., 2020).

Limitations of the existing studies applying ML techniques on FTD patient cohorts. While most studies suggest an implementation of these classifiers in clinical practice, some of the methods used are not readily available to clinicians. Additionally, only few studies evaluate multiclass predictions *(e.g.,* Kim et al., 2019; Klöppel et al., 2018; Torso et al., 2020; Zimmerer et al., 2020). Multiclass predictions are more similar to decisions made by clinicians than binary predictions and are thus more relevant when trying to maximize usability in the clinical context. Further, many studies do not differentiate between FTD subgroups. Instead they group bvFTD, svPPA and nfvPPA together under the umbrella term FTD in order to separate them from AD, dementia with Lewy bodies, Parkinson's disease dementia or Vascular Dementia (*e.g.,* Bruun et al., 2018; Garn et al., 2017; Lin et al., 2020).

1.7 The Current Study

The current study makes use of k-means clustering, an unsupervised learning algorithm, in order to explore data of five patient groups belonging to the spectrum of FTD (*i.e.,* bvFTD, svPPA, nfvPPA) or AD and the related disorder lvPPA. The analysis includes behavioral and cognitive data only. This means variables analyzed stem either from results of neuropsychological assessments or from questionnaires filled-in by the patient or a relevant informant. The main goal of the current study is reaching a better understanding of the characteristic cognitive and behavioral symptoms of the different patient groups. Such an exploratory analysis may inform differential diagnosis between these groups which persists to be difficult still today. Possibly a better understanding of the differences may allow for a refinement of diagnostic criteria. Diagnostic precision is relevant to psychological and social well-being of both patients and caregivers (Musa et al., 2020; Weder et al., 2007). Additionally, it is necessary for precision of research investigating therapeutic strategies and possible drug treatments. Using only neuropsychological data ensures a high clinical applicability as it is comparatively neither cost nor time intense. Likely, the combination of neuropsychological with neuroimaging and pathological data would outperform diagnostic decisions based on neuropsychological data alone. However, we believe that in a first step it is necessary to investigate only neuropsychological data as its potential contribution is not yet well understood. As presented earlier, rather than focusing on group differences, which may have small effect sizes and do not always reflect diagnosis on the individual level, ML may allow for a better understanding of the heterogeneity within patient groups and possibly unveil patterns in the data not previously known. Application of clustering algorithms for dementia syndromes is still relatively new but has been increasing in recent years. The most common clustering algorithm used in these studies is hierarchical clustering (*review on clustering applied to AD see* Alashwal et al., 2019; *e.g.,* Machulda et al., 2013; Matias-Guiu et al., 2018). Analysis of PPA in these studies commonly focuses on data from language assessments (*e.g.,* Fan et al., 2020; Knibb et al., 2006; Leyton et al., 2014) while for bvFTD data from various neuroimaging techniques is used (*e.g., MRI:* Cerami et al., 2016; *FDG-PET:* Josephs et al., 2009). Thus, an analysis of a breadth of neuropsychological data using k-means clustering may yield new insights for each one of the patient groups as well as for their distinction. K-means clustering groups data points in a way to reduce total Euclidean distance (Hennig et al., 2015).

Exploration of the data in this study follows two major goals: In a first step, we inspect whether and how clusters correspond to diagnostic groups in the data. Subsequently, we inspect clusters more closely to understand their differences on neuropsychological test scores, called

cluster centers. Why do participants of different diagnostic groups cluster apart or together and how does this compare to current diagnostic standards? Based on previous research using clustering methods on patients from the FTD spectrum, one may hypothesize that svPPA is more likely to cluster separately than nfvPPA and lvPPA. Additionally, both for PPAs and bvFTD more than a single cluster may be needed per diagnostic group. To our knowledge, no single study has investigated clustering approaches with all five diagnostic groups included in the current study. For interpretation of the results a particular focus will be put on the exploration of the following comparisons as they are thought to be the most difficult to differentially diagnose: 1. bvFTD and AD, 2. The three PPA variants and 3. svPPA and bvFTD.

2 Methods

2.1 Patient Data

Patient data was provided by the German Consortium for frontotemporal lobar degeneration (FTLD consortium study) (Otto et al., 2011) collected from 13 clinics or research centers. From the n = 1036 participants originally in the dataset, n = 254 were excluded for not having a diagnosis (n = 2), having a diagnosis that was not relevant for the current study (n = 66 *CBS*, n = 91 *PSP*, n = 1 *FTLD phenocopy*), being assigned to an unspecific group (n = 9 *other neurological disorder*, n = 7 diagnosed with bvFTD without specification, n = 22 diagnosed with bvFTD and specification *Neurodegenerative Disease*, n = 18 *other healthy controls*), carrying gene alterations related to FTLD but being asymptomatic at the moment of assessment (n = 3), being related to diagnosed patients in the dataset (n = 20 *family member of patient with positive genetic test*, n = 14 *family member of patient without genetic test*), not having information on their gender (n = 1). In total n = 782 patients remained for further analysis (n = 69 healthy controls, n = 132 AD, n = 292 bvFTD, n = 97 svPPA, n = 118 nfvPPA, n = 74 lvPPA). Next, patients with more than 20% of missing variables were removed. The analysis was performed twice, once including and once excluding questionnaire scores. A total of n = 469 and n = 462 participants remained for the analysis with and without questionnaires, respectively. The size of each diagnostic group varied highly with bvFTD patients forming the largest group (n = 171 and n = 173 in the analysis with and without questionnaires respectively) and lvPPA patients forming the smallest group (n = 47 and n = 48 in the analysis with and without questionnaires respectively). Tables 1 and 2 summarize demographic scores of the six participant groups for the cohort in the analysis excluding and including questionnaires, respectively. BvFTD and PPA patients met standard diagnostic criteria as stated by Rascovsky et al. (2011) and Gorno-Tempini et al. (2011), respectively. Participants provided written informed consent. Participants were not compensated for their participation in the study. The

study was approved by the ethics committees of all contributing Universities and was in accordance with the latest version of the Declaration of Helsinki (ethics committee Leipzig ID 137-11-18042011).

Table 1

Demographics Table Excluding Questionnaires

						% Gender		Age		Years of Education		FTLD-CDR Score		Age at Symptom Onset		Disease Duration	
Group	N	% of Total	% Clinical Diagnosis	% Imaging-Supported Diagnosis	% Diagnosis With Definite Pathology	Male	Female	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Healthy Controls	49	10.4	-	-	-	44.9	55.1	64.5	9.3	15	3.1	0.1	0.3	-	-	-	-
Alzheimer	74	15.8	-	-	-	54.1	45.9	66.1	9.7	13.5	3.3	5.8	3.4	63.1	10	2.8	2.8
bvFTD	171	36.5	34.5	57.3	8.2	62.6	37.4	62.3	9.3	13.5	3	6.8	3.9	58.8	10.5	3.4	3.8
svPPA	52	11.1	17.3	78.8	3.8	46.2	53.8	62.9	7.9	14.7	3.2	4.9	2.4	60.3	7.9	2.6	1.9
nfvPPA	76	16.2	43.4	55.3	1.3	48.7	51.3	68.9	8.3	13.1	3.3	4.2	2.4	67	8.6	1.9	1.2
lvPPA	47	10	38.3	48.9	12.8	48.9	51.1	68.7	6.1	13.3	3.5	4.7	2.8	64.9	6.5	4	4
Total	469	100	34.4	59	6.7	53.9	46.1	64.9	9.2	13.7	3.2	5.1	3.7	61.9	9.9	3	3.2

Table 2

Demographics Table Including Questionnaires

						% Gender		Age		Years of Education		FTLD-CDR Score		Age at Symptom Onset		Disease Duration	
Group	N	% of Total	% Clinical Diagnosis	% Imaging-Supported Diagnosis	% Diagnosis With Definite Pathology	Male	Female	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Healthy Controls	49	10.6	-	-	-	44.9	55.1	64.5	9.3	15.0	3.1	0.1	0.3	-	-	-	-
Alzheimer	70	15.2	-	-	-	52.9	47.1	66.4	9.7	13.4	3.4	5.7	3.3	63.5	9.9	2.8	2.8
bvFTD	173	37.4	36.4	55.5	8.1	62.4	37.6	62.3	9.5	13.4	2.9	6.9	4.0	58.8	10.4	3.4	3.5
svPPA	50	10.8	20.0	78.0	2.0	46.0	54.0	63.2	7.9	14.6	3.3	4.8	2.4	60.5	7.9	2.6	1.9
nfvPPA	72	15.6	45.8	52.8	1.4	47.2	52.8	68.7	8.6	13.1	3.4	4.2	2.5	66.6	8.9	2.0	1.3
lvPPA	48	10.4	37.5	50.0	12.5	47.9	52.1	68.4	6.2	13.3	3.4	4.7	2.8	64.7	6.6	4.0	3.9
Total	462	100	36.2	57.4	6.4	53.5	46.5	64.9	9.2	13.6	3.2	5.1	3.7	61.8	9.8	3.0	3.1

2.2 Neuropsychological Test Scores

Demographic variables included for further analysis were participants' age and years of education. The neuropsychological test scores used can broadly be attributed to the six neurocognitive domains defined in the DSM-5 (American Psychiatric Association, 2013). They will be described in the following and are summarized in Tables 3 and 4. Further, Tables 6 and 7 summarize the mean scores of participant groups on all variables included in the analysis.

2.2.1 Clinical Dementia Rating (CDR) (Morris, 1993) and FTLD-CDR Scales (Knopman et al., 2008)

The CDR consists of six dimensions that are rated by a clinician based on a semi-structured interview with the patient and an informant. It is a measure of dementia severity. Originally, it was developed for the assessment in patients with AD and contains the dimensions *Memory*, *Orientation*, *Judgment and Problem Solving*, *Community Affairs*, *Home and Hobbies* and *Personal Care*. For the assessment in FTD the adaptation by Knopman et al. (2008) added the two dimensions *Language* and *Behavior, comportment & personality* resulting in a total of eight dimensions. Following a semi-structured interview, the clinician rates each dimension on a five point scale ranging across 0, 0.5, 1, 2, 3 corresponding to no, questionable, mild, moderate or severe impairment, respectively. A global score of the CDR is derived from the subscores of a patient. *Memory* is used as primary category which scores are adjusted depending on all other secondary categories following the rules described by Morris (1993) to yield the global score.

Additionally, the sum of boxes created by simply summing across all subscores is commonly used as a measure of severity. For the sum of boxes, no difference is made between domains and can thus vary between 0 and 18 or 0 and 24 for the CDR and FTLD-CDR, respectively. The current study made use of all subscores, and both sum of boxes. The global score was not assessed.

2.2.2 CERAD-NAB Plus Battery (Fillenbaum et al., 2008; Moms et al., 1989)

The CERAD battery is a screening tool originally constructed for AD and consists of a variety of neuropsychological subtests.

Two tests of verbal fluency assess semantic (Isaacs & Kennie, 1973) and phonological (Thurstone, 1973) fluency. In the semantic fluency test, participants are required to name as many words as possible falling into a specified category. In the current study, the category was "animals". For the phonological fluency assessment, participants need to name as many words as possible starting with a specified letter. In the current study, the s-version was used. Proper nouns are not allowed. Both tests are limited in time to one minute. For this study, only the total number of correct words of each test was used for analysis. The semantic and phonological

fluency tests form the first and last tests of the CERAD battery, respectively.

The second test included in the battery is the modified Boston Naming Test (BNT) (Kaplan et al., 1978). It consists of 15 drawings of objects that need to be named by participants. The pictures were selected so that five pictures correspond to each, high, medium and low frequency occurrence in the English language. Only the total score of correctly identified pictures was used for the analysis.

The next part of the CERAD battery is the MMSE (Folstein et al., 1975). This test is the most commonly used screening tool for dementia both in research and the clinical setting (Arevalo-Rodriguez et al., 2021). It consists of short test items assessing orientation, learning and memory, language, attention, and calculation and requires less than 30 min for completion. It has been criticized to overestimate cognitive impairment in case memory or language impairment are present and to not be sensitive to mild cognitive impairment (Tombaugh & McIntyre, 1992).

Next, a list consisting of ten words (Atkinson & Shiffrin, 1971; Mohs et al., 1986) is read by the participant and immediate recall is assessed. This is repeated three times with the same words presented in varying order to assess whether the participant can increase the number of words memorized with repetition. After a delay period, recall of the words is assessed. Additionally, recognition is tested by presenting 20 words including the ten memorized ones. The participant's task is to determine the words read previously and reject the new ones. For analysis total scores of first, second, and third immediate recall as well as total immediate and delayed recall are used. Additionally, a measure of savings is calculated by dividing the number of words recalled after the delay period by the number of the third immediate recall. Intrusions are counted. Discriminability is calculated as measure of recognition performance by summing the number of correctly identified and the number of correctly rejected words and dividing it by the maximum score of 20.

A measure of constructional praxis (W. G. Rosen et al., 1984) assesses visuospatial construction and memory performance. The participant is sequentially presented with four figures and asked to copy them as good as possible. The figures selected are in increasing difficulty, consisting of a circle, a rhombus, two overlapping rectangles and finally a cube. Visuospatial memory is assessed by letting participants draw the figures from memory following a delay period. Both copy and recall are assessed on a list of criteria with maximum summed points of the figures corresponding to eleven. For recall ability, savings is assessed by calculating the ratio of summed score for the recall divided by the total score from the copy task.

Finally, the Trail Making Test (TMT) (Reitan, 1958) consists of two versions, A and B. Version A is thought to measure attention and processing speed. Version B additionally is used as a measure of cognitive flexibility meaning the ability to maintain two tasks in mind and flexibly switch between them. Switching ability is part of executive functioning. In version A participants are presented with a sheet of paper on which the numbers from one to 25 are distributed across the page. Participants are required to draw a line connecting all numbers starting from one and continuing sequentially with the next higher number to end at 25. In version B both numbers ranging from one to 13 and letters ranging from A to L are arranged on a sheet of paper. Participants need to draw a line combining all numbers and letters while constantly switching between both sequences, starting with the number one and ending with 13. In both versions time is assessed, and participants are told to perform the task as quickly and as correctly as possible. The test is stopped after a maximum of 180 seconds and 300 seconds for versions A and B, respectively. Additional measures are the sum of errors committed and the ratio of time needed for version B and A. The ratio is thought to indicate the switching cost.

2.2.3 Modified Stroop Task (Stroop, 1935)

The Stroop task consists of three subtests. The first two subtests are thought to measure processing speed and attention. The first subtest is the *color naming test* in which participants are presented with lines of red, green and blue color patches. Participants are requested to name the color. In the second subtest, the *word reading test*, participants are presented with color words that they need to read out loud. The third subtest, the *color-word interference test*, is used as a measure of selective attention, considered an executive function. In this test, participants are presented with color words printed in a color that does not correspond to the meaning of the word. Participants are required to name the color in which words are printed. It is based on the Stroop effect describing the observation made by Stroop (1935) that highly trained tasks are performed automatically. It is thought, that reading is highly automatic and that in case color of the ink and color word do not correspond, the automatic reading creates an interference. For participants to give the correct response without an extensive delay, they must inhibit the automatic reading process. In all subtests, the participant is given 45 seconds to name or read as many items as possible. For analysis the total score of each subtest as well as the total number of errors made is assessed.

2.2.4 Wechsler Memory Scale Revised (WMS-R) (Wechsler, 1987)

Two tests of the WMS-R were used in the current study: the *digit span* and the *visual memory span*, similar to the block tapping test (*e.g.,* Matias-Guiu et al., 2020). Both tests are

similar in structure. They assess memory and working memory function by testing the maximum number of items that can be recalled. In the digit span task verbal memory is assessed. The clinician reads a sequence of digits. Once finished, the participant is requested to repeat the digits. In the visual memory span task instead, blocks are assembled that the experimenter taps in a predetermined sequence. The observing participant is requested to retrace the blocks tapped by the experimenter. In the forward version of both tasks, participants are required to repeat or retrace the sequences in the same order as presented. In the backward version instead, participants are required to inverse the order, starting from the last number heart or the last block that was tapped by the clinician. The clinician starts with a small span of two or three items (depending on the specific subtest). If at least one of two tasks from the same span length are correctly repeated by the participant, the clinician increases the length by one item. The task is stopped when the participant fails twice to correctly repeat the sequence of a specific length. One point is given for each correctly repeated sequence. The scores for each one of the versions, digit span forward and backward and visual memory span forward and backward, can vary between zero and twelve.

2.2.5 Cookie Theft Task (Goodglass & Kaplan, 1972)

The cookie theft task is part of the Boston Diagnostic Aphasia Examination. The instructions are simple: The participant is shown a drawing depicting a scene and is asked to describe it. The picture shows a mother and her two children in the kitchen. The mother is washing the dishes while behind her the son stands on a stool to grab a cookie from a jar. The stool on which the son is standing is about to fall. His sister is on the ground next to him raising her hand to take a cookie from her brother. The mother is standing in a puddle created by the overflowing sink. She does not seem to notice neither the water she is standing in, nor the children behind her. There is a window from which one can view outside. The task is used as a measure of spontaneous speech and can be analyzed for various language characteristics, such as grammar and fluency (*e.g.,* Ash et al., 2017). In the current study, only a quantitative score assessing the content was used. Number of specified items mentioned in the participants story are counted reaching a possible maximum score of 20.

2.2.6 Repeat and Point Test German Adaptation (Heitkamp et al., 2010)

The repeat and point test is used to assess word comprehension and repetition ability. Ten words pronounced by the clinician need to be repeated by participants. Additionally, they are presented with seven images that are semantically and partly perceptually similar. From these images the one corresponding to the repeated word needs to be identified and pointed at. Each correct repeating or pointing yields one point. Thus, a maximum of ten points on each the

Repeat and *Point* part can be reached. This test was previously shown to separate between the different PPA variants (Hodges et al., 2008; Seckin et al., 2022). Specifically, patients with svPPA were more impaired on the pointing part, while nfvPPA was related to greater impairment on the repetition part of the test. No difference was found between both tasks for participants diagnosed with lvPPA (Seckin et al., 2022).

2.2.7 Aachen Aphasia Test (AAT) (Willmes et al., 1983)

The AAT was developed to assess aphasic disorders in German speaking participants. The two subtests used in the current study were the Token Test and the Written Language Test. The Token Test assesses general language function and can be used to distinguish aphasic from non-aphasic disorders but not to differentiate between different aphasic disorders. Participants are presented with squares and circles of different sizes and colors. The clinician gives an instruction of what needs to be done with an item of a specified shape, color, and size. For each task that participants fulfill correctly at the first trial, they receive two points. If they successfully perform at the second trial, they get one point. The maximum points reachable is 50.

The Written Language test of the AAT consists of three parts with each being tested on ten items of different complexities. Items may be single- or multi-syllable words, compound nouns or whole sentences. The first task is to read written items out loud. The second task is to assemble heard words using displayed letters. The third task is to write a word from dictated letters. Each item is scored on a four-point scale from 0-3. Thus, the total score can vary between zero and 90.

2.2.8 Hamasch Five Point Test (H5PT) (Regard et al., 1982)

The H5PT is used to assess cognitive fluency, considered an executive function. A sheet of paper is filled with squares and five points arranged inside each square in the position of the five on a die. Participants need to draw lines connecting points in a way to create a new pattern in each square. Patterns can be created by connecting minimum two dots. Participants are given three minutes to create as many new patterns as possible. The variables assessed are the total number of squares filled with patterns, the total number of unique patterns and the ratio of the two.

2.2.9 Cognitive Estimation Task (Shallice & Evans, 1978) updated version (MacPherson et al., 2014)

In the cognitive estimation task participants are confronted with questions concerning speed, length, area, number and weight. Participants are requested to estimate the answer and increasing points are given depending on how far participants' estimates deviate from a defined

range around the correct answer. The first version was developed by Shallice & Evans (1978) and a newer version with questions adapted to changed lifestyles was constructed by MacPherson et al. (2014). It was used in the current study as a part of various executive function assessments.

2.2.10 Reading the Mind in the Eyes Test (RMET) Revised Version (Baron-Cohen et al., 2001)

The RMET is thought to assess social cognition. Participants are shown pictures of eyes depicting diverse facial expressions corresponding to specified emotions. Participants are asked to identify the fitting emotion out of a selection of emotion words. In total 20 picture-word pairs are used, 10 taken from female and 10 from male faces. Each correctly selected emotion yields one point. The current study used only the sum of points for analysis. The RMET has previously been used in a cohort of bvFTD patients and was suggested to yield better distinction of bvFTD from healthy controls than executive function assessments (Schroeter et al., 2018).

2.2.11 Three Clap Test (TCT) (Dubois et al., 1995)

The TCT is used to assess the applause sign. The clinician claps three times and the participant is asked to do the same. Clapping three times results in zero points. With higher number of claps higher scores are reached indicating motor perseverations. This was repeated twice, and the total of both trials was used for further analysis. Motor perseverations are particularly common in the related FTD syndromes PSP, CBS and FTD with ALS. The motor sign may however also be observed in patients with bvFTD and PPA (Schönecker et al., 2019). It is not expected to differ between patient groups included in the current study.

Table 3

Screening tools included and respective variables used for analysis

Screening tool	*Variables used for analysis*
(FTLD-)CDR	FTLD-CDR Sum of boxes
	CDR Sum of boxes
	Memory
	Orientation
	Judgment and Problem Solving
	Community Affairs
	Home and Hobbies
	Personal Care
	Behavior, Comportment and Personality
	Language
MMSE	CERAD MMSE total score

Table 4

Tests used to assess neuropsychological function across the six neurocognitive domains defined

by the DSM-5 (American Psychiatric Association, 2013; Sachdev et al., 2014)

Neurocognitive domain	*Neuropsychological variables used for analysis*
Complex attention	CERAD TMT A
	Stroop Task - color naming
	Stroop Task - word reading
Perceptual-motor function	CERAD Visuo-constructional praxis copy
Language	CERAD Boston Naming Test
	Repeat and point test – Point
	Repeat and point test – Repeat
	CERAD semantic fluency
	CERAD phonemic fluency
	Cookie theft task
	AAT Token Test
	AAT Written Language Test
Learning and memory	CERAD word list immediate recall 1
	CERAD word list immediate recall 1
	CERAD word list immediate recall 1
	CERAD word list delayed recall
	CERAD word list discriminability
	WMS-R digit span forward
	WMS-R visual memory span forward
	CERAD savings visuo-constructional praxis
Social cognition	RMET
Executive function	CERAD TMT B
	CERAD TMT error scores
	CERAD TMT ratio B/A
	WMS-R digit span backward
	WMS-R visual memory span backward
	Stroop color-word interference
	H5PT total
	H5PT total correct
	H5PT percent correct
	Cognitive estimation

2.3 Behavioral Questionnaire Scores

Questionnaires assessing participants deviation from normal behavior are described in the following section. Scores included in the analysis are summarized in Table 5.

2.3.1 Apathy Evaluation Scale (AES) (Marin et al., 1991)

The AES consists of 18 items to assess the presence of apathic symptoms. Ratings are provided both via a self-version by the patient and via an informant-version by the clinician or a person close to the patient. Each item is rated on a four-point scale spanning from zero to three. Reachable scores thus vary between zero and 54 with higher scores indicating greater presence of apathic symptoms. Apathy is one of the most common behavioral symptoms in neurodegenerative disorders and was found to be the most common neuropsychiatric symptom both in AD and FTD syndromes in a systematic review (Collins et al., 2020). Apathy is

characterized by low levels of motivation or drive and emotional indifference (Massimo et al., 2018). Examples of items of the AES ask whether the patient is interested in new experiences, has friends, or shows initiative.

2.3.2 Bayer Activities of Daily Living Scale (B.ADL) (Hindmarch et al., 1998)

The B.ADL was developed as an instrument to assess severity of functional impairment in patients with mild cognitive impairment or dementia. It indicates how independent a patient is in the performance of his*her daily activities. It consists of 25 items that are filled in by the patient and an informant of the patient. The items span general screening questions about whether the patient is able to take care of him*herself or his*her daily activities, as well as questions about abilities to perform specific tasks and questions referring to cognitive functions required for appropriate completion of daily tasks. Each item is rated on a ten-point scale with the extremes *never* (0) and *always* (10) as well as the possibility to indicate whether an item is *not applicable,* or the answer is *unknown* to the informant. To evaluate results, the average of all answered items is taken (*i.e.,* not including the *not applicable* or *unknown* items). Final scores thus vary between zero and ten with a higher score indicating a greater impact of the dementia on daily life functioning and greater help required.

2.3.3 Frontal Systems Behavior Scale (FrSBe) (Grace & Malloy, 2001)

The FrSBe assesses behavioral change related to damage of frontal circuits. Three subscales assess levels of apathy, executive dysfunction and disinhibition associated with anterior cingulate cortex, dorsolateral prefrontal cortex and orbitofrontal circuits respectively. The questionnaire consists of 24 items administered to the patient and to an informant. Each item describing a behavior is rated across both frequency and distress caused by the behavior to the patient and to the informant. Items are rated on a five-point scale. On the frequency scale answers range from *almost never* (1) to *almost always* (5). On the distress scale, possible answers range from *not at all distressing* (1) to *extremely distressing or very severe* (5). The apathy subscale contains items similar to the items included in the AES asking for reduced drive and emotional blunting. The executive dysfunction subscale includes items about the patient's ability to monitor the own behavior, distractibility or working memory. Items on the disinhibition subscale ask for inappropriate behaviors such as sexual or aggressive behaviors and impulsiveness. As items on the distress scale do not correspond between the self- and the informant- version referring to distress felt by the patient or by the informant, respectively, this study included only the frequency of observed behaviors. For the analysis sum scores for each one of the three subscales was used and may vary between 24 and 120.

2.3.4 Short Form Geriatric Depression Scale (GDS) (Yesavage et al., 1982)

The GDS is used as a screening tool for depressive symptoms in the aging population. The original version still in use today is a long form containing 30 items. A shorter version consisting of 15 items is commonly used with patients who have limited attention or resilience due to dementia or other physical or cognitive illness. This short version was used in our study. Items are *yes* or *no* questions that are answered by the patient. The number of answers provided that indicate presence of a depressive symptom is counted. Sum scores are evaluated as follows:

0-4: no depression

5-8: mild depression

9-11: moderate depression

12-15: severe depression

It has also been suggested for use to track change of depressive symptoms over time (Ishihara & Terada, 2001).

Table 5

Questionnaire scores used for assessment of behavioral and psychiatric changes

Questionnaire	Variables used for analysis
AES	AES total score
B.ADL	B.ADL total score
FrSBe	Executive dysfunction frequency
	Disinhibition frequency
	Apathy frequency
GDS	GDS total score

Table 6

Group Differences Across all Variables Included (Without Questionnaires)

Variables	***Healthy Controls (N = 49)***	***Alzheimer (N = 74)***	***bvFTD (N = 171)***	***svPPA (N = 52)***	***nfvPPA (N = 76)***	***lvPPA (N = 47)***
Age	64.5 (9.3)	66.1 (9.7)	62.3 (9.3)	62.9 (7.9)	68.9 (8.3)	68.7 (6.1)
Years of Education	15.0 (3.1)	13.5 (3.3)	13.5 (3.0)	14.7 (3.2)	13.1 (3.3)	13.3 (3.5)
CDR Sum of Boxes	0.1 (0.2)	4.8 (2.9)	5.0 (3.1)	2.8 (1.7)	2.2 (1.8)	2.8 (2.3)
FTLD-CDR Sum of Boxes	0.1 (0.3)	5.8 (3.4)	6.8 (3.9)	4.9 (2.4)	4.2 (2.4)	4.7 (2.8)
CDR: Memory	0.1 (0.2)	1.4 (0.7)	0.8 (0.5)	0.8 (0.4)	0.5 (0.4)	0.9 (0.6)
CDR: Orientation	0.0 (0.1)	0.7 (0.8)	0.4 (0.5)	0.1 (0.3)	0.2 (0.4)	0.4 (0.5)
CDR: Judgment, PBS	0.0 (0.0)	1.0 (0.7)	1.1 (0.7)	0.6 (0.6)	0.5 (0.5)	0.6 (0.6)
CDR: Community Affairs	0.0 (0.1)	0.8 (0.6)	1.0 (0.7)	0.6 (0.4)	0.5 (0.4)	0.6 (0.6)
CDR: Home & Hobbies	0.0 (0.0)	0.8 (0.6)	1.0 (0.7)	0.5 (0.4)	0.4 (0.4)	0.5 (0.5)
CDR: Personal Care	0.0 (0.0)	0.2 (0.4)	0.6 (0.7)	0.1 (0.3)	0.1 (0.2)	0.1 (0.2)
FTLD-CDR: Behavior	0.0 (0.1)	0.4 (0.4)	1.3 (0.7)	0.6 (0.6)	0.3 (0.4)	0.2 (0.4)
FTLD-CDR: Language	0.0 (0.0)	0.6 (0.7)	0.5 (0.6)	1.4 (0.6)	1.6 (0.8)	1.7 (0.7)
CERAD: Semantic Fluency	26.5 (5.6)	12.2 (5.7)	12.8 (6.2)	9.0 (4.5)	11.0 (6.5)	10.7 (5.3)
CERAD: BNT	14.9 (0.2)	12.4 (2.7)	13.2 (2.3)	6.5 (3.2)	12.1 (3.2)	10.9 (3.6)
CERAD: MMSE	29.2 (0.9)	22.9 (4.8)	25.4 (3.5)	23.8 (4.5)	24.7 (5.1)	22.3 (6.0)
CERAD: Total Wordlist	22.6 (3.1)	11.4 (4.5)	15.2 (4.6)	14.5 (6.2)	15.1 (5.9)	12.0 (6.5)
CERAD: Wordlist 1	5.8 (1.4)	2.7 (1.6)	3.8 (1.7)	3.4 (1.8)	3.7 (1.9)	2.7 (2.1)
CERAD: Wordlist 2	7.8 (1.3)	4.1 (1.7)	5.3 (1.7)	5.2 (2.4)	5.2 (2.2)	4.2 (2.2)
CERAD: Wordlist 3	9.0 (1.0)	4.6 (1.9)	6.1 (1.9)	5.9 (2.5)	6.1 (2.2)	5.1 (2.7)
CERAD: Wordlist Recall	8.1 (1.8)	2.4 (2.2)	4.4 (2.3)	3.8 (3.1)	5.1 (2.5)	4.1 (2.7)
CERAD: Wordlist Savings	89.8 (17.3)	46.8 (37.9)	69.3 (28.9)	53.3 (36.6)	82.2 (30.0)	76.8 (39.9)
CERAD: Wordlist Recognition	99.2 (2.1)	84.4 (12.8)	88.7 (14.8)	83.4 (13.5)	91.7 (13.4)	91.4 (10.3)

Variables	*Healthy Controls (N = 49)*	*Alzheimer (N = 74)*	*bvFTD (N = 171)*	*svPPA (N = 52)*	*nfvPPA (N = 76)*	*lvPPA (N = 47)*
CERAD: Wordlist Intrusions	0.3 (0.6)	2.8 (4.6)	1.6 (2.1)	1.7 (3.1)	0.9 (1.3)	1.1 (1.8)
CERAD: Figure Copy	10.8 (0.7)	9.3 (2.1)	9.6 (1.8)	10.4 (1.3)	9.8 (1.8)	9.9 (1.3)
CERAD: Figure Recall	9.7 (1.8)	3.5 (3.0)	6.0 (3.6)	6.8 (3.3)	7.6 (3.2)	6.0 (3.3)
CERAD: Figure Savings	89.9 (16.3)	36.2 (29.6)	61.1 (33.7)	65.5 (29.1)	75.8 (28.6)	59.1 (30.5)
CERAD: Phonemic Fluency	17.8 (4.3)	8.5 (4.4)	7.3 (4.5)	7.5 (3.8)	4.5 (3.7)	6.1 (3.4)
CERAD: TMT A	37.6 (12.3)	86.9 (44.6)	73.6 (37.9)	54.6 (24.6)	85.8 (43.7)	78.0 (43.0)
CERAD: TMT B	79.9 (32.3)	215.2 (75.9)	177.5 (81.8)	132.9 (67.1)	211.8 (78.4)	205.7 (78.0)
CERAD: TMT B/A	2.2 (0.6)	3.1 (1.1)	2.9 (1.3)	2.7 (1.4)	2.9 (0.9)	3.4 (1.1)
CERAD: TMT Errors	0.6 (1.0)	1.7 (2.6)	2.2 (2.3)	0.4 (0.8)	2.7 (2.7)	1.3 (1.4)
WMS-R: Digit Span Fw	8.7 (1.7)	5.8 (2.2)	6.2 (1.8)	6.5 (2.2)	4.5 (1.9)	3.8 (2.2)
WMS-R: Digit Span Bw	7.1 (2.2)	4.1 (1.8)	4.1 (1.7)	5.2 (1.7)	3.5 (1.6)	3.3 (1.8)
WMS-R: Visual Span Fw	8.0 (2.0)	5.3 (2.0)	5.9 (2.2)	7.8 (1.9)	6.2 (1.8)	5.6 (1.7)
WMS-R: Visual Span BW	7.3 (1.9)	4.2 (1.8)	5.1 (2.2)	6.9 (1.8)	5.1 (2.3)	4.9 (1.9)
Cookie Theft Task	14.7 (2.8)	10.9 (3.5)	9.8 (3.5)	8.8 (3.7)	9.4 (3.9)	9.7 (3.4)
AAT: Token Test	0.6 (1.1)	6.7 (7.0)	5.8 (6.9)	8.4 (10.4)	10.8 (9.5)	10.6 (9.0)
AAT: Written Language	89.7 (0.7)	84.7 (6.3)	84.3 (10.5)	84.7 (6.0)	78.7 (15.5)	79.2 (13.0)
Repeat	9.9 (0.3)	9.5 (1.2)	9.8 (0.7)	9.5 (0.9)	8.2 (2.5)	8.3 (2.2)
Point	9.9 (0.4)	8.7 (1.5)	8.7 (1.6)	6.6 (2.4)	9.0 (1.2)	8.6 (1.7)
Stroop: Color Naming	72.8 (11.1)	43.7 (15.0)	45.9 (17.1)	50.4 (19.0)	31.6 (14.1)	34.8 (14.5)
Stroop: Word Reading	98.2 (10.2)	59.9 (18.4)	65.3 (21.4)	76.1 (21.8)	46.1 (18.3)	57.3 (18.7)
Stroop: Interference	41.2 (10.2)	17.1 (8.4)	20.5 (11.8)	27.6 (11.9)	16.1 (10.6)	14.3 (11.2)
Stroop: Error	0.4 (1.0)	2.6 (3.6)	3.0 (8.1)	2.7 (7.9)	1.6 (4.1)	2.5 (4.4)
H5PT: Total Correct	32.0 (8.0)	15.6 (6.3)	16.3 (10.0)	20.7 (8.4)	17.4 (8.3)	18.6 (8.2)
H5PT: Total	35.1 (7.9)	20.7 (9.6)	25.1 (13.8)	25.6 (10.2)	23.8 (10.5)	24.9 (13.3)
H5PT: Percent Correct	91.2 (12.6)	79.5 (18.4)	68.4 (27.2)	82.2 (20.6)	77.2 (22.4)	81.1 (19.8)

Variables	*Healthy Controls (N = 49)*	*Alzheimer (N = 74)*	*bvFTD (N = 171)*	*svPPA (N = 52)*	*nfvPPA (N = 76)*	*lvPPA (N = 47)*
Cognitive Estimation	12.7 (2.1)	8.8 (3.1)	9.4 (3.2)	9.2 (3.1)	9.4 (3.1)	9.5 (3.2)
Applause Sign	0.1 (0.3)	0.1 (0.4)	0.3 (0.9)	0.3 (0.7)	0.5 (1.3)	0.1 (0.5)
RMET	16.5 (2.7)	12.5 (3.6)	10.9 (3.7)	10.1 (2.6)	10.6 (2.7)	11.6 (3.8)

Table 7

Group Differences Across all Variables Included (With Questionnaires)

Variables	*Healthy Controls (N = 49)*	*Alzheimer (N = 70)*	*bvFTD (N = 173)*	*svPPA (N = 50)*	*nfvPPA (N = 72)*	*lvPPA (N = 48)*
Age	64.5 (9.3)	66.4 (9.7)	62.3 (9.5)	63.2 (7.9)	68.7 (8.6)	68.4 (6.2)
Years of Education	15.0 (3.1)	13.4 (3.4)	13.4 (2.9)	14.6 (3.3)	13.1 (3.4)	13.3 (3.4)
CDR Sum of Boxes	0.1 (0.2)	4.7 (2.8)	5.1 (3.2)	2.8 (1.7)	2.2 (1.8)	2.8 (2.3)
FTLD-CDR Sum of Boxes	0.1 (0.3)	5.7 (3.3)	6.9 (4.0)	4.8 (2.4)	4.2 (2.5)	4.7 (2.8)
CDR: Memory	0.1 (0.2)	1.4 (0.7)	0.8 (0.6)	0.8 (0.4)	0.5 (0.5)	0.9 (0.6)
CDR: Orientation	0.0 (0.1)	0.7 (0.8)	0.5 (0.5)	0.2 (0.3)	0.2 (0.4)	0.4 (0.5)
CDR: Judgment, PBS	0.0 (0.0)	1.0 (0.6)	1.2 (0.7)	0.6 (0.6)	0.5 (0.5)	0.6 (0.6)
CDR: Community Affairs	0.0 (0.1)	0.8 (0.6)	1.0 (0.7)	0.6 (0.4)	0.5 (0.4)	0.6 (0.6)
CDR: Home & Hobbies	0.0 (0.0)	0.8 (0.6)	1.0 (0.7)	0.5 (0.4)	0.4 (0.4)	0.5 (0.5)
CDR: Personal Care	0.0 (0.0)	0.2 (0.4)	0.6 (0.7)	0.1 (0.3)	0.1 (0.3)	0.1 (0.2)
FTLD-CDR: Behavior	0.0 (0.1)	0.3 (0.4)	1.3 (0.7)	0.6 (0.6)	0.3 (0.4)	0.2 (0.4)
FTLD-CDR: Language	0.0 (0.0)	0.6 (0.7)	0.5 (0.6)	1.4 (0.6)	1.6 (0.7)	1.7 (0.7)
CERAD: Semantic Fluency	26.5 (5.6)	12.2 (5.7)	12.8 (6.2)	8.9 (4.5)	11.4 (6.6)	10.8 (5.3)
CERAD: BNT	14.9 (0.2)	12.3 (2.8)	13.1 (2.4)	6.4 (3.1)	12.1 (3.2)	10.9 (3.5)
CERAD: MMSE	29.2 (0.9)	22.9 (5.1)	25.3 (3.7)	23.7 (4.6)	24.9 (5.2)	22.2 (6.0)
CERAD: Total Wordlist	22.6 (3.1)	11.4 (4.7)	15.2 (4.5)	14.3 (5.8)	15.4 (6.0)	12.0 (6.5)
CERAD: Wordlist 1	5.8 (1.4)	2.7 (1.6)	3.8 (1.6)	3.3 (1.6)	3.8 (1.9)	2.7 (2.0)

Variables	*Healthy Controls (N = 49)*	*Alzheimer (N = 70)*	*bvFTD (N = 173)*	*svPPA (N = 50)*	*nfvPPA (N = 72)*	*lvPPA (N = 48)*
CERAD: Wordlist 2	7.8 (1.3)	4.1 (1.8)	5.3 (1.6)	5.2 (2.3)	5.3 (2.3)	4.2 (2.2)
CERAD: Wordlist 3	9.0 (1.0)	4.7 (1.9)	6.1 (1.9)	5.9 (2.4)	6.2 (2.3)	5.1 (2.7)
CERAD: Wordlist Recall	8.1 (1.8)	2.4 (2.2)	4.4 (2.2)	3.6 (3.0)	5.1 (2.6)	4.1 (2.7)
CERAD: Wordlist Savings	89.8 (17.3)	46.0 (38.2)	70.0 (28.8)	52.5 (36.4)	82.3 (30.8)	76.5 (39.5)
CERAD: Wordlist Recognition	99.2 (2.1)	84.0 (13.0)	88.7 (14.7)	82.3 (14.0)	92.4 (12.1)	91.5 (10.2)
CERAD: Wordlist Intrusions	0.3 (0.6)	2.8 (4.7)	1.6 (2.1)	1.8 (3.1)	0.8 (1.3)	1.1 (1.7)
CERAD: Figure Copy	10.8 (0.7)	9.4 (1.9)	9.5 (1.8)	10.3 (1.3)	9.8 (1.8)	9.8 (1.4)
CERAD: Figure Recall	9.7 (1.8)	3.7 (3.0)	6.0 (3.6)	6.8 (3.3)	7.7 (3.1)	5.9 (3.2)
CERAD: Figure Savings	89.9 (16.3)	37.3 (29.6)	60.6 (34.1)	64.9 (29.5)	77.0 (28.0)	59.4 (30.2)
CERAD: Phonemic Fluency	17.8 (4.3)	8.6 (4.4)	7.3 (4.5)	7.3 (3.7)	4.8 (3.8)	6.1 (3.4)
CERAD: TMT A	37.6 (12.3)	88.8 (46.0)	74.6 (39.5)	52.6 (20.9)	81.3 (40.7)	77.3 (42.8)
CERAD: TMT B	79.9 (32.3)	213.5 (75.7)	177.9 (82.0)	136.8 (72.0)	211.7 (78.4)	208.5 (78.5)
CERAD: TMT B/A	2.2 (0.6)	3.1 (1.1)	3.0 (1.3)	2.8 (1.4)	3.0 (1.0)	3.5 (1.2)
CERAD: TMT Errors	0.6 (1.0)	1.7 (2.6)	2.2 (2.3)	0.4 (0.8)	2.6 (2.5)	1.3 (1.4)
WMS-R: Digit Span Fw	8.7 (1.7)	5.7 (2.3)	6.1 (1.8)	6.5 (2.2)	4.5 (1.9)	3.8 (2.2)
WMS-R: Digit Span Bw	7.1 (2.2)	4.1 (1.8)	4.0 (1.7)	5.1 (1.7)	3.6 (1.6)	3.2 (1.8)
WMS-R: Visual Span Fw	8.0 (2.0)	5.4 (2.0)	5.9 (2.2)	7.8 (1.9)	6.4 (1.6)	5.5 (1.7)
WMS-R: Visual Span BW	7.3 (1.9)	4.2 (1.8)	5.1 (2.3)	6.9 (1.8)	5.1 (2.4)	4.8 (1.9)
Cookie Theft Task	14.7 (2.8)	10.9 (3.5)	9.9 (3.5)	8.7 (3.8)	9.6 (3.8)	9.7 (3.3)
AAT: Token Test	0.6 (1.1)	6.9 (7.7)	6.0 (7.3)	8.9 (11.2)	10.4 (9.0)	11.0 (9.3)
AAT: Written Language	89.7 (0.7)	84.3 (8.5)	83.9 (10.8)	84.5 (6.1)	79.9 (14.5)	79.4 (12.9)
Repeat	9.9 (0.3)	9.4 (1.3)	9.7 (0.9)	9.5 (1.0)	8.4 (2.1)	8.3 (2.2)
Point	9.9 (0.4)	8.7 (1.6)	8.7 (1.6)	6.5 (2.5)	9.1 (1.2)	8.6 (1.7)
Stroop: Color Naming	72.8 (11.1)	44.0 (14.9)	45.8 (17.0)	49.8 (18.5)	31.5 (14.2)	34.5 (14.5)
Stroop: Word Reading	98.2 (10.2)	59.8 (19.0)	65.4 (21.4)	75.2 (22.2)	45.8 (18.2)	57.6 (18.6)

Variables	*Healthy Controls (N = 49)*	*Alzheimer (N = 70)*	*bvFTD (N = 173)*	*svPPA (N = 50)*	*nfvPPA (N = 72)*	*lvPPA (N = 48)*
Stroop: Interference	41.2 (10.2)	17.2 (8.5)	20.6 (11.7)	27.4 (12.1)	16.0 (10.5)	14.0 (11.2)
Stroop: Error	0.4 (1.0)	2.7 (3.7)	3.0 (8.0)	2.9 (8.0)	1.6 (4.1)	2.4 (4.3)
H5PT: Total Correct	32.0 (8.0)	15.6 (6.3)	16.4 (10.2)	20.5 (8.4)	17.9 (8.3)	18.5 (8.1)
H5PT: Total	35.1 (7.9)	20.3 (9.5)	25.2 (13.9)	25.4 (10.3)	23.3 (9.5)	24.8 (13.2)
H5PT: Percent Correct	91.2 (12.6)	80.6 (17.6)	68.0 (26.8)	82.5 (20.9)	79.2 (20.4)	80.8 (19.6)
Cognitive Estimation	12.7 (2.1)	8.9 (3.1)	9.4 (3.2)	9.1 (3.0)	9.4 (3.1)	9.5 (3.2)
Applause Sign	0.1 (0.3)	0.1 (0.4)	0.3 (1.0)	0.3 (0.7)	0.4 (1.1)	0.1 (0.5)
RMET	16.5 (2.7)	12.4 (3.7)	10.9 (3.8)	10.0 (2.7)	10.6 (2.8)	11.7 (3.7)
AES	7.4 (4.8)	20.1 (12.1)	34.6 (9.7)	17.7 (9.0)	16.7 (10.3)	15.9 (10.3)
B-ADL	1.5 (0.5)	4.2 (2.3)	5.1 (2.4)	3.7 (2.1)	3.3 (1.9)	3.6 (2.1)
FrSBe: Executive Dysfunction	13.8 (4.1)	21.5 (8.1)	26.6 (6.7)	18.5 (7.2)	15.0 (6.1)	18.4 (6.8)
FrSBe: Disinhibition	14.3 (3.5)	12.8 (5.0)	17.5 (6.3)	14.2 (5.4)	10.6 (3.5)	12.7 (4.0)
FrSBe: Apathy	11.7 (3.7)	18.0 (8.5)	27.8 (7.2)	17.5 (6.3)	16.7 (6.6)	16.3 (6.9)
GDR	1.6 (1.8)	3.6 (2.5)	3.5 (3.1)	4.9 (3.6)	3.5 (2.6)	4.5 (3.0)

2.4 **Missing Data**

Due to high levels of missing data, several steps were taken to prepare the data for the k-means clustering analysis. Exploration of the pattern of missingness across variables and diagnostic groups indicated that patients had particularly high levels of missingness in the self-rating parts of the three questionnaires (*i.e.,* Apathy Scale, B.ADL, FrSBe) while healthy controls had particularly high levels of missingness in the informant-rating parts of the same questionnaires (Figure 1). The pattern observed in the patient groups may be explained by patients being too affected to comprehend and provide appropriate answers to the questionnaires. For healthy controls instead a high number of missingness in the informant version of the questionnaire scores may be explained by informants not being systematically recruited for control participants. To accommodate for this observation, we decided to minimize missingness by including only the informant-based scores for the patient groups while including only the self-rating scores for healthy controls. This decision was supported by the observation from previous studies, that informant-based questionnaires may be more indicative of behavioral changes of the patients than self-ratings (*e.g.* Schroeter et al., 2018). In the case of control participants, no behavioral abnormalities are expected and thus informant- and self-ratings are not thought to differ meaningfully. We opted for this procedure rather than simply excluding questionnaire scores, as the information contained in the questionnaires might have an added value, especially in the differential diagnosis of bvFTD, whose clinical picture is thought to be characterized primarily by behavioral rather than cognitive changes. However, validity of this method may be questioned. Additionally, missingness observed in the current dataset might reflect difficulties to collect questionnaire data in the clinical setting. For these reasons, we decided to perform the analysis twice: once including and once excluding questionnaire scores to compare results and to estimate the usefulness of questionnaire data for differential diagnosis.

Figure 1

Percentage of Data Present Across all Variables, Split by Participant Group

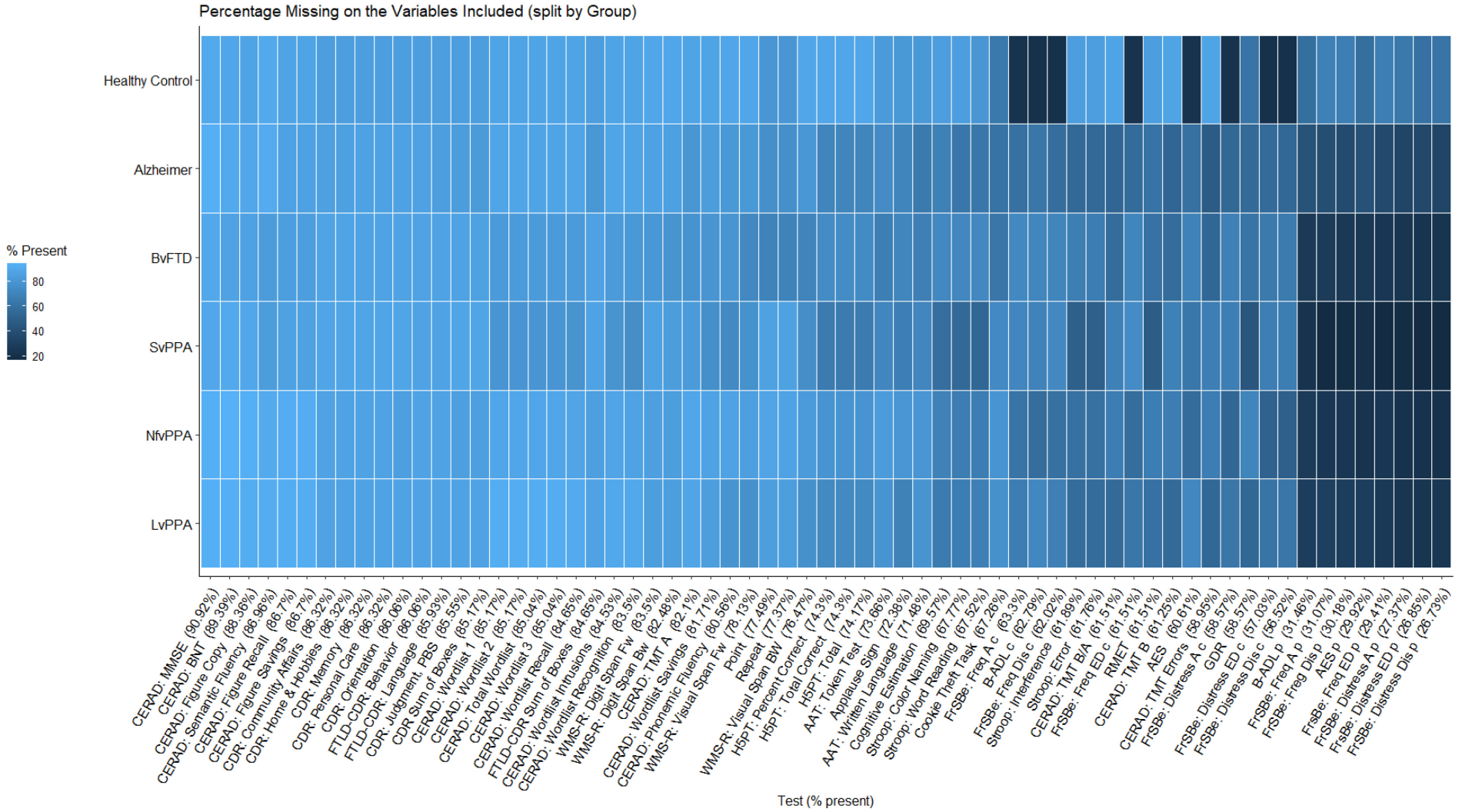


Figure 2

Number of Participants with Specified Percentage of Missingness (top: Count, bottom: Density)

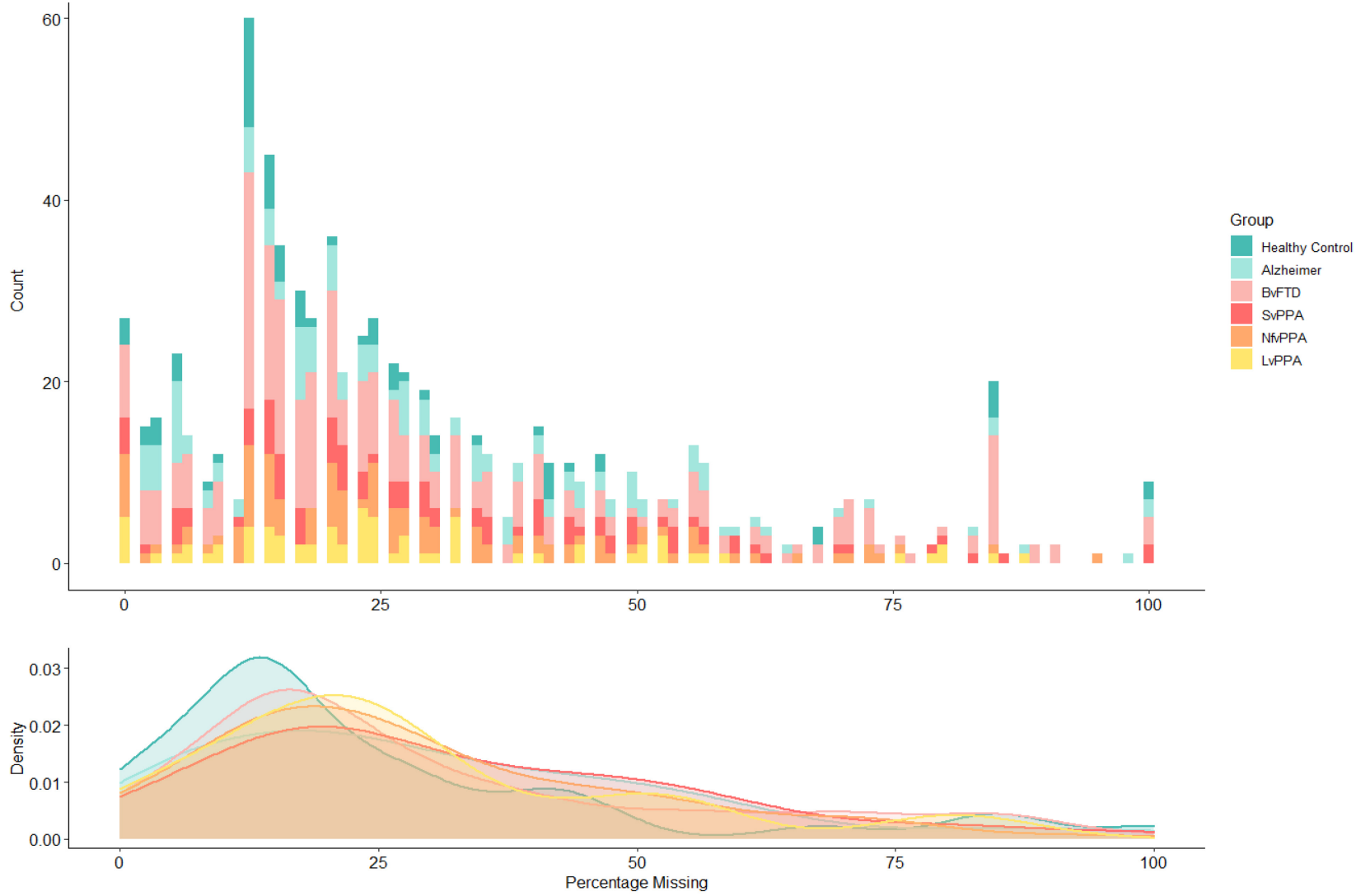


Table 8

Number and Percentage of Missing Values for Each one of the Variables Included

	Excluding Questionnaires			***Including Questionnaires***		
Variable	***Patients***	***Healthy Controls***	***Total***	***Patients***	***Healthy Controls***	***Total***
Participant	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Age	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Gender	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Years of Education	14 (3.33%)	2 (4.08%)	16 (3.41%)	17 (4.12%)	2 (4.08%)	19 (4.11%)
Group	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Diagnostic Specification	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
CDR Sum of Boxes	2 (0.48%)	0 (0%)	2 (0.43%)	10 (2.42%)	0 (0%)	10 (2.16%)
FTLD-CDR Sum of Boxes	5 (1.19%)	0 (0%)	5 (1.07%)	13 (3.15%)	0 (0%)	13 (2.81%)
CDR: Memory	0 (0%)	0 (0%)	0 (0%)	9 (2.18%)	0 (0%)	9 (1.95%)
CDR: Orientation	0 (0%)	0 (0%)	0 (0%)	9 (2.18%)	0 (0%)	9 (1.95%)
CDR: Judgment, PBS	2 (0.48%)	0 (0%)	2 (0.43%)	10 (2.42%)	0 (0%)	10 (2.16%)
CDR: Community Affairs	1 (0.24%)	0 (0%)	1 (0.21%)	9 (2.18%)	0 (0%)	9 (1.95%)
CDR: Home & Hobbies	1 (0.24%)	0 (0%)	1 (0.21%)	9 (2.18%)	0 (0%)	9 (1.95%)
CDR: Personal Care	1 (0.24%)	0 (0%)	1 (0.21%)	9 (2.18%)	0 (0%)	9 (1.95%)
FTLD-CDR: Behavior	2 (0.48%)	0 (0%)	2 (0.43%)	11 (2.66%)	0 (0%)	11 (2.38%)
FTLD-CDR: Language	3 (0.71%)	0 (0%)	3 (0.64%)	12 (2.91%)	0 (0%)	12 (2.6%)
CERAD: Semantic Fluency	5 (1.19%)	0 (0%)	5 (1.07%)	5 (1.21%)	0 (0%)	5 (1.08%)
CERAD: BNT	1 (0.24%)	0 (0%)	1 (0.21%)	1 (0.24%)	0 (0%)	1 (0.22%)
CERAD: MMSE	1 (0.24%)	0 (0%)	1 (0.21%)	1 (0.24%)	0 (0%)	1 (0.22%)
CERAD: Total Wordlist	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
CERAD: Wordlist 1	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
CERAD: Wordlist 2	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

	Excluding Questionnaires			*Including Questionnaires*		
Variable	***Patients***	***Healthy Controls***	***Total***	***Patients***	***Healthy Controls***	***Total***
CERAD: Wordlist 3	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
CERAD: Wordlist Recall	2 (0.48%)	0 (0%)	2 (0.43%)	1 (0.24%)	0 (0%)	1 (0.22%)
CERAD: Wordlist Savings	8 (1.9%)	0 (0%)	8 (1.71%)	7 (1.69%)	0 (0%)	7 (1.52%)
CERAD: Wordlist Recognition	3 (0.71%)	0 (0%)	3 (0.64%)	2 (0.48%)	0 (0%)	2 (0.43%)
CERAD: Wordlist Intrusions	2 (0.48%)	0 (0%)	2 (0.43%)	1 (0.24%)	0 (0%)	1 (0.22%)
CERAD: Figure Copy	1 (0.24%)	0 (0%)	1 (0.21%)	0 (0%)	0 (0%)	0 (0%)
CERAD: Figure Recall	3 (0.71%)	0 (0%)	3 (0.64%)	3 (0.73%)	0 (0%)	3 (0.65%)
CERAD: Figure Savings	3 (0.71%)	0 (0%)	3 (0.64%)	3 (0.73%)	0 (0%)	3 (0.65%)
CERAD: Phonemic Fluency	15 (3.57%)	0 (0%)	15 (3.2%)	13 (3.15%)	0 (0%)	13 (2.81%)
CERAD: TMT A	15 (3.57%)	0 (0%)	15 (3.2%)	11 (2.66%)	0 (0%)	11 (2.38%)
CERAD: TMT B	106 (25.24%)	0 (0%)	106 (22.6%)	98 (23.73%)	0 (0%)	98 (21.21%)
CERAD: TMT B/A	104 (24.76%)	0 (0%)	104 (22.17%)	96 (23.24%)	0 (0%)	96 (20.78%)
CERAD: TMT Errors	119 (28.33%)	0 (0%)	119 (25.37%)	109 (26.39%)	0 (0%)	109 (23.59%)
WMS-R: Digit Span Fw	3 (0.71%)	0 (0%)	3 (0.64%)	3 (0.73%)	0 (0%)	3 (0.65%)
WMS-R: Digit Span Bw	6 (1.43%)	0 (0%)	6 (1.28%)	6 (1.45%)	0 (0%)	6 (1.3%)
WMS-R: Visual Span Fw	17 (4.05%)	0 (0%)	17 (3.62%)	17 (4.12%)	0 (0%)	17 (3.68%)
WMS-R: Visual Span BW	24 (5.71%)	0 (0%)	24 (5.12%)	24 (5.81%)	0 (0%)	24 (5.19%)
Cookie Theft Task	62 (14.76%)	14 (28.57%)	76 (16.2%)	57 (13.8%)	14 (28.57%)	71 (15.37%)
AAT: Token Test	30 (7.14%)	1 (2.04%)	31 (6.61%)	26 (6.3%)	1 (2.04%)	27 (5.84%)
AAT: Written Language	40 (9.52%)	1 (2.04%)	41 (8.74%)	33 (7.99%)	1 (2.04%)	34 (7.36%)
Repeat	20 (4.76%)	2 (4.08%)	22 (4.69%)	16 (3.87%)	2 (4.08%)	18 (3.9%)
Point	19 (4.52%)	2 (4.08%)	21 (4.48%)	16 (3.87%)	2 (4.08%)	18 (3.9%)
Stroop: Color Naming	36 (8.57%)	1 (2.04%)	37 (7.89%)	29 (7.02%)	1 (2.04%)	30 (6.49%)
Stroop: Word Reading	35 (8.33%)	3 (6.12%)	38 (8.1%)	28 (6.78%)	3 (6.12%)	31 (6.71%)
Stroop: Interference	70 (16.67%)	1 (2.04%)	71 (15.14%)	64 (15.5%)	1 (2.04%)	65 (14.07%)

	Excluding Questionnaires			*Including Questionnaires*		
Variable	***Patients***	***Healthy Controls***	***Total***	***Patients***	***Healthy Controls***	***Total***
Stroop: Error	70 (16.67%)	1 (2.04%)	71 (15.14%)	64 (15.5%)	1 (2.04%)	65 (14.07%)
H5PT: Total Correct	14 (3.33%)	0 (0%)	14 (2.99%)	14 (3.39%)	0 (0%)	14 (3.03%)
H5PT: Total	15 (3.57%)	0 (0%)	15 (3.2%)	15 (3.63%)	0 (0%)	15 (3.25%)
H5PT: Percent Correct	14 (3.33%)	0 (0%)	14 (2.99%)	14 (3.39%)	0 (0%)	14 (3.03%)
Cognitive Estimation	34 (8.1%)	0 (0%)	34 (7.25%)	32 (7.75%)	0 (0%)	32 (6.93%)
Applause Sign	34 (8.1%)	0 (0%)	34 (7.25%)	33 (7.99%)	0 (0%)	33 (7.14%)
RMET	66 (15.71%)	1 (2.04%)	67 (14.29%)	58 (14.04%)	1 (2.04%)	59 (12.77%)
AES	-	-	-	91 (22.03%)	13 (26.53%)	104 (22.51%)
B-ADL	-	-	-	88 (21.31%)	13 (26.53%)	101 (21.86%)
FrSBe: Executive Dysfunction	-	-	-	94 (22.76%)	10 (20.41%)	104 (22.51%)
FrSBe: Disinhibition	-	-	-	92 (22.28%)	10 (20.41%)	102 (22.08%)
FrSBe: Apathy	-	-	-	84 (20.34%)	8 (16.33%)	92 (19.91%)
GDR	-	-	-	103 (24.94%)	15 (30.61%)	118 (25.54%)

Next, participants who exceeded the threshold of 20% missingness in the variables relevant for analysis were excluded. In accordance with the current standard in research, missing data in the remaining participants was imputed (Austin et al., 2021; Hayati Rezvan et al., 2015; Madley-Dowd et al., 2019). Imputation was performed using multiple imputation by chained equations (MICE) (Buuren & Groothuis-Oudshoorn, 2011; Raghunathan et al., 2001). In comparison to single imputation, multiple imputation does not replace each missing value with a single but with multiple, commonly five or more, values (van Buuren, 2018). This creates several datasets on which analysis is then performed before results are combined in a last step. Compared to single imputation this procedure avoids underestimating standard errors (Enders, 2010). While single imputation treats imputed and non-imputed data equally, the level of uncertainty stemming from the imputation of missing data can be explored when applying multiple imputation. As analysis is performed on each imputed dataset, one can inspect how much the results from the repeated analysis diverge. Results that vary highly across imputations should be interpreted with care. Especially in the case in which data is not missing completely at random (MCAR) (Rubin, 1976) and proportion of missing data exceeds few percent points (usually a threshold of 5-10%), multiple imputation is considered a powerful way for dealing with missing data (Enders, 2010; Schafer & Graham, 2002; van Buuren, 2018).

2.5 Procedure

2.5.1 *Statistical Analysis*

Statistical analysis was performed entirely using RStudio (RStudio Team, 2020) version 4.2.0. Analysis was restricted to the first visit of participants, in accordance with a goal of a diagnosis at the earliest time point possible. Diagnostic groups were inspected on differences in demographic variables. For the subsequent analysis, only the two numerical demographic variables age and years of education were included. Each neuropsychological variable was checked for implausible values and in total 9 datapoints were detected to fall outside the defined range of the respective variable and thus coded as missing value (*i.e.,* "NA"). This small amount of data points not interpretable within the range of the different variables can most probably be explained by typing errors.

2.5.2 *Multiple Imputation*

Multiple imputation was performed using the mice package (Buuren & Groothuis-Oudshoorn, 2011) for multiple imputation by chained equations. In case both sub- and total scores were present in the data, sub-scores were imputed, and total scores derived from the imputations post hoc. To avoid circularity of the imputation process and subsequent clustering analysis, the diagnostic grouping variable was not included for imputation of other variables.

Imputation was performed m = 10 times and for a maximum of maxit = 100 iterations. To ensure that results of imputation would yield realistic values within the ranges of the variables used, paired mean matching was defined as imputation method. Visual checks of the imputation were performed (see Appendix 1 – Data Preparation). First, convergence was ensured by inspecting mean and standard deviation across datasets and iterations as described by van Buuren (2018). Next density distributions of the original and the imputed data were inspected

2.5.3 *Clustering Analysis*

Prior to the k-means clustering analysis, the following variables were reversed in order for higher scores to consistently indicate better performance and thereby facilitate interpretation of clustering results: CDR score, FTLD-CDR total and sub-scores, number of intrusions when recalling a word list, time needed for the TMT versions A and B as well as their ratio and the number of errors made, results on the Token Test, the applause sign, cognitive estimation test, the number of errors on the Stroop task and all questionnaire scores (*i.e.,* Apathy scale, B.ADL, FrSBe, Depression Scale). In a next step the scores were rescaled to vary between -1 and 1 to avoid range effects across variables (*e.g.,* Brigadoi et al., 2017). K-means clustering was implemented using the kmeans function from the stats package (R Core Team, 2021). To explore the data flexibly, the analysis was performed multiple times with the cluster size varying from k = 2 to k = 9. The groups included in the analysis were a) all groups (*i.e.,* healthy controls, bvFTD, nfvPPA, svPPA, lvPPA, AD), b) only the patient groups, c) bvFTD and PPAs, d) only PPAs and e) bvFTD and AD. The analysis was performed on each one of the 10 imputed datasets. Results from the 10 datasets were inspected on similarity using Cramer's V. Additionally, proportion of ambiguous clustering (PAC) was assessed as suggested by Şenbabaoğlu et al. (2015) to determine stability of the clustering results. PAC was calculated with lower bound of consensus indices set to 0.1 and upper bound set to 0.9. When comparing clustering results across the ten imputations, PAC expresses the proportion of participant pairs for whom less than nine out of the ten analyses agreed on whether the two participants cluster together or apart (Şenbabaoğlu et al., 2015). In a next step the results were combined using majority_voting() from the diceR package (Chiu & Talhouk, 2021). This simple method for consensus clustering evaluates correspondence of cluster labels across different runs of the analysis before assigning to each participant the cluster that it was most often attributed to. Results were evaluated first looking at whether clusters separated between different diagnostic groups and how this changed with increasing number of clusters. Next results were inspected more in detail by looking at neuropsychological test performance to possibly identify the pattern underlying the observed clustering. The complete code and results can be found in the

Appendix.

Based on the exploratory nature of the clustering analysis, results will be summarized in a purely descriptive way as was performed in previous studies (*e.g.,* Maruta et al., 2015; Scheltens et al., 2016). This will help generation of new hypotheses that can be tested in more rigorous future studies. Results will be presented for each one of the patient comparisons separately. Due to the exhaustiveness of analyses performed the focus is on consistent findings observed across cluster sizes and comparisons.

3 Results

3.1 General Quality Checks

In most cases Cramer's V varied between 0.7 and 1 for pairs of datasets (Figure 3). Only in rare cases it dropped below 0.7 for single datasets. Generally, Cramer's V decreased with increasing number of clusters but remained high even for the maximum number of clusters which was set to k = 9 (Table 8). An additional index, PAC, supported the idea that clustering results were relatively stable across imputed datasets (Table 10 and 11, Figure 4). In most cases PAC ranged from 0.06 to 0.26. Analysis c) with PPA and bvFTD patients resulted in particularly bad results on this index ranging from 0.13 to 0.37. With three clusters (k = 3) a PAC of 0.37 indicated that nearly 40% of participants can be considered as ambiguously clustered. Interpretation of the results from this analysis should thus be taken with care. Except for this specific case, we concluded that clustering results across imputations were sufficiently stable for further analysis. Interestingly, in some cases Cramer's V was lowest for intermediate numbers of clusters. Similarly, PAC was usually higher (indicating lower stability) for intermediate cluster sizes. One possible explanation could be that a higher number of clusters may be more fitting to the data.

Table 8

Cramer's V across all analyses performed (Split by Amount of Clusters)

Amount of clusters (k)	***Mean (sd)***
2	0.96 (0.03)
3	0.9 (0.11)
4	0.92 (0.05)
5	0.86 (0.09)
6	0.86 (0.1)
7	0.87 (0.09)
8	0.86 (0.09)
9	0.86 (0.08)

Table 9

Cramer's V across all analyses performed (Split by Groups Compared)

Groups included	***Mean (sd)***
All Groups	0.9 (0.07)
All Patient Groups	0.86 (0.11)
PPA and bvFTD	0.88 (0.09)
PPA	0.89 (0.08)
AD and bvFTD	0.88 (0.1)

Figure 3

Density Distribution Cramer's V

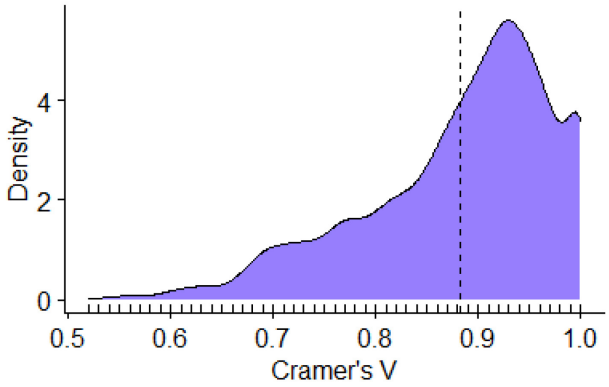


Table 10

PAC across all analyses performed (Split by Amount of Clusters)

Amount of clusters (k)	***Mean (sd)***
2	0.08 (0.04)
3	0.17 (0.1)
4	0.14 (0.05)
5	0.22 (0.08)
6	0.17 (0.07)
7	0.15 (0.05)
8	0.14 (0.04)
9	0.12 (0.03)

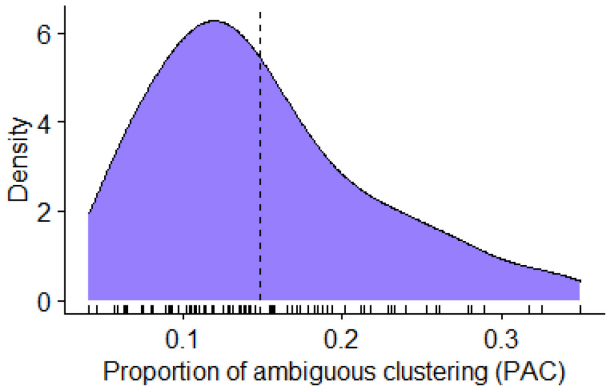
Table 11

PAC across all analyses performed (Split by Groups Compared)

Groups included	***Mean (sd)***
All Groups	0.12 (0.05)
All Patient Groups	0.17 (0.08)
PPA and bvFTD	0.17 (0.09)
PPA	0.13 (0.05)
AD and bvFTD	0.15 (0.07)

Figure 4

Density Distribution PAC



3.2 Consensus Clustering

For clarity, results in the following section will be exemplified with clusters referred to using the nomenclature Cj_k with j indicating the cluster number and the subscript k referring to the clustering parameter k (*i.e.,* amount of clusters defined for the analysis).

3.2.1 All Groups

3.2.1.1 Diagnostic Groups. The first analysis was done including all groups of interest (*i.e.,* healthy controls, AD, bvFTD, nfvPPA, lvPPA, svPPA). When inspecting the diagnosis of patients in each cluster, it becomes apparent that healthy controls cluster together but not separately from the patient groups (Figure 5 and 6). With increasing number of clusters, the homogeneity of the cluster containing the healthy controls increases. However, even at the maximum number of clusters (k = 9) it does not surpass 75% ($C9_9$ consists of n = 46 healthy controls, n = 6 bvFTD, n = 5 nfvPPA, n = 3 svPPA and n = 1 lvPPA patients). With the number of clusters being set to k = 4, a cluster emerges ($C3_4$) that, to a great majority (81%) consists of bvFTD patients, the remaining being AD patients (19%). The participants that make up this cluster continue to form a separate cluster also for the analysis with a greater number of k (*i.e.,* k = 5 – 9). Taking a closer look at the analysis with cluster size set to k = 9, it seems that several clusters show some degree of homogeneity in the diagnostic groups that they include. Apart from a cluster containing mainly healthy controls ($C9_9$), and a cluster with mainly bvFTD patients ($C8_9$) described earlier, an additional cluster contains mainly bvFTD patients ($C5_9$: 92% bvFTD patients) and one cluster contains mainly svPPA patients ($C6_9$: 71% svPPA patients). The cluster grouping most AD patients also contains a great proportion of bvFTD patients ($C1_9$: 60% AD and 31% bvFTD patients). Additionally, lvPPA and nfvPPA seem to cluster together. They do not separate well from other patient groups but most lvPPA and nfvPPA patients are distributed in clusters $C2_9$ and $C3_9$. Although very mixed, it may be noted that one cluster that contains mainly bvFTD patients also contains a considerable proportion of svPPA patients in the dataset ($C4_9$: 22% of svPPA patients). At least three clusters remain very mixed ($C1_9$, $C2_9$, $C7_9$).

Including questionnaire scores (Figure 6) in the analysis does not seem to greatly affect the cluster results. However, one observation is that bvFTD patients seem to cluster more easily apart from other patient groups when questionnaire scores are included. For example, at cluster size set to k = 3, a relatively homogeneous cluster containing mainly bvFTD patients emerges ($C3_3$ with 72% bvFTD patients) which is not observed in the analysis without questionnaire scores. Additionally, at k = 5 two relatively homogeneous clusters emerge compared to one in the analysis without questionnaire scores ($C2_5$ and $C5_5$ with 82% and 80% bvFTD patients, respectively). Interestingly, cluster results seem to be more homogeneous for svPPA in the analysis without compared to the analysis with questionnaire scores.

Figure 5

Visualizing Clusters by their Composition of Patient Groups (Including all Groups, Without Questionnaires)

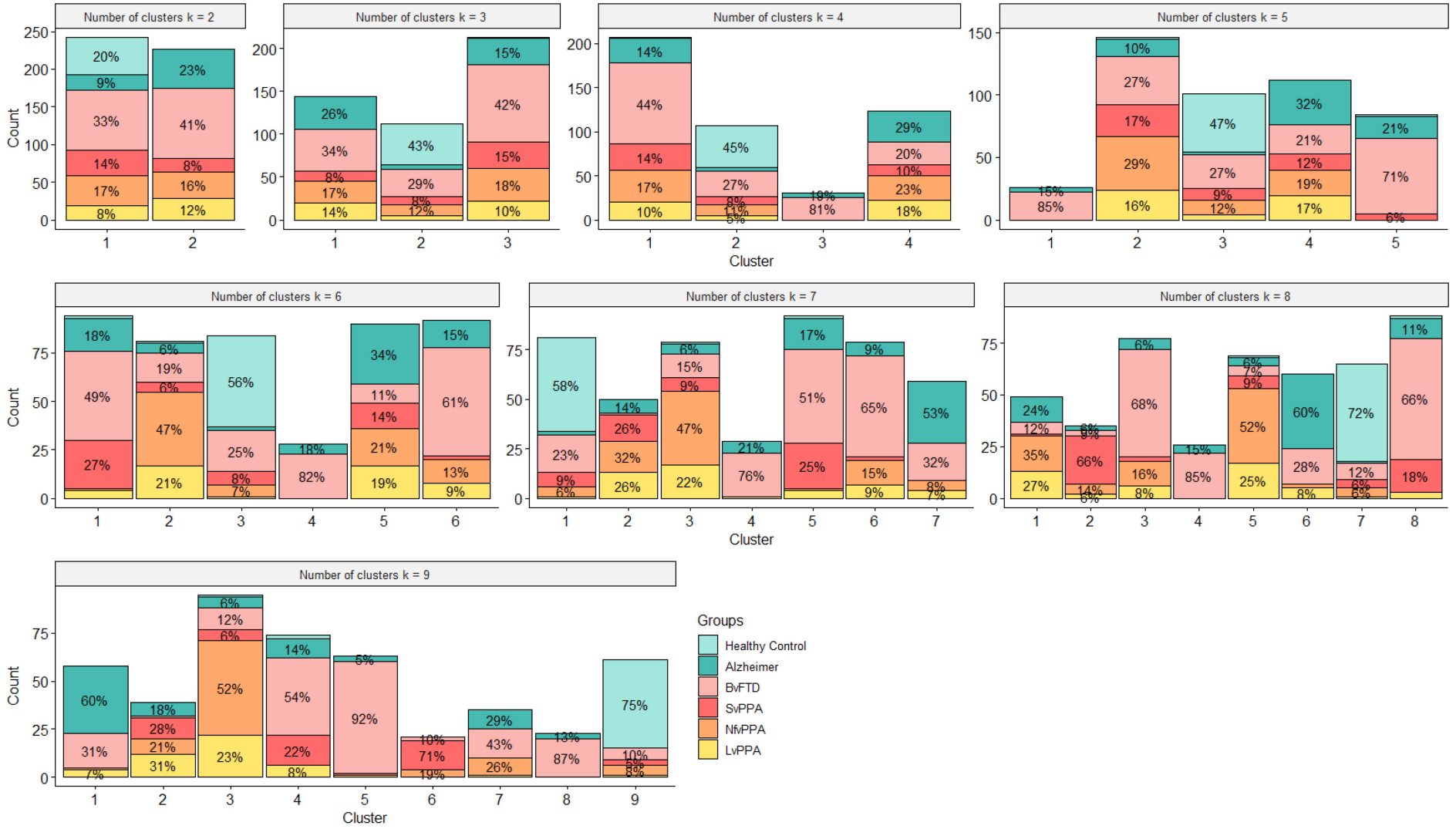
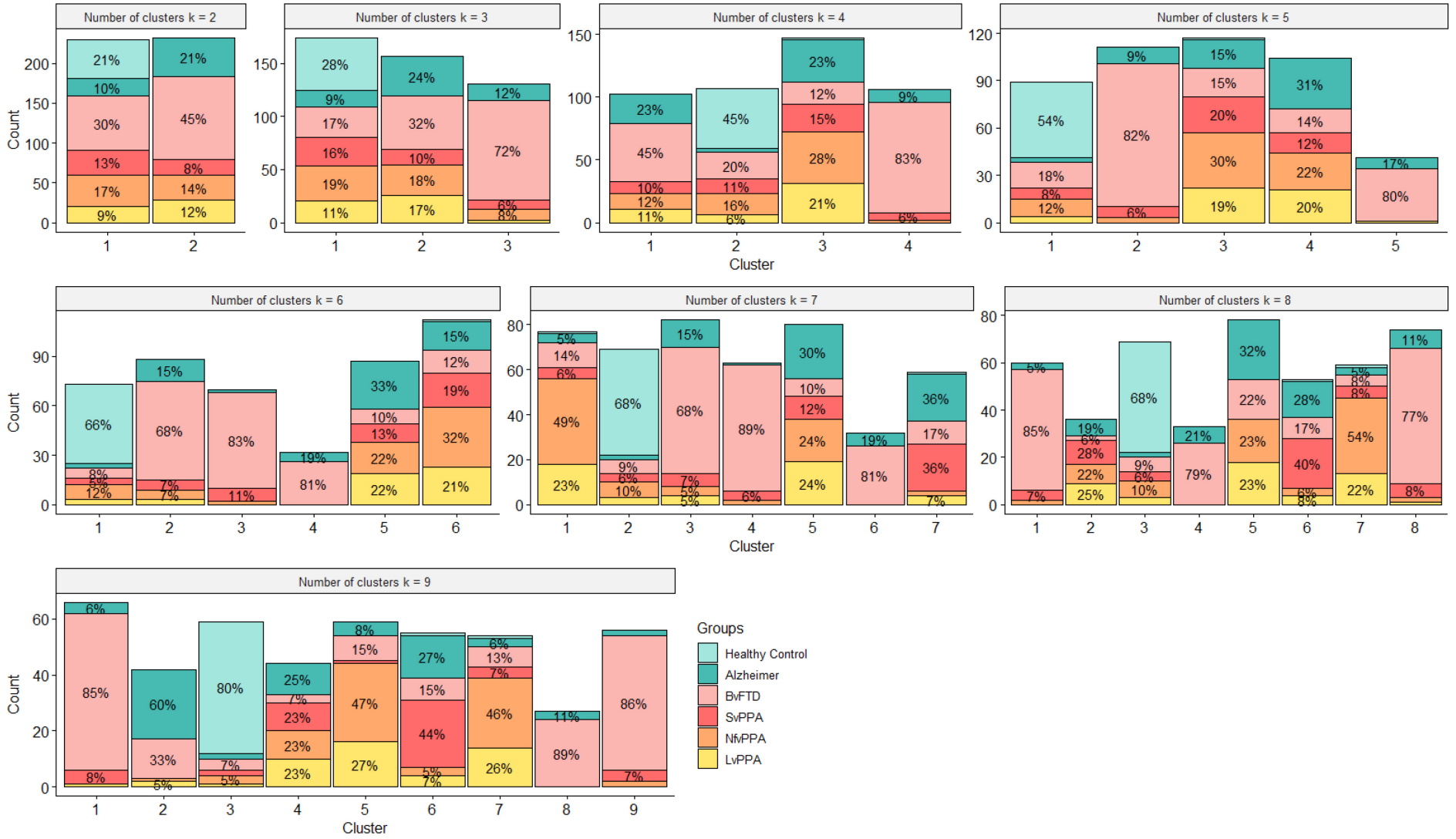


Figure 6

Visualizing Clusters by their Composition of Patient Groups (Including all Groups, With Questionnaires)



3.2.1.2 Neuropsychological Profile. Next, cluster centers were inspected. Results were similar with and without questionnaires. A special focus was put on those clusters that showed a relatively great homogeneity in the diagnostic groups they comprised. With the smallest number of clusters (k = 2 – 3 and k = 2 without and with questionnaire scores, respectively), results suggest clusters to differ mainly with respect to disease severity. This interpretation is based on the observation that clusters show consistently better or worse performance on nearly all scores included. In part clustering results of analyses with a greater amount of clusters could also be explained by differences in severity. For example, consistent differences can be observed between clusters at k = 4 (Figure 7) without questionnaires (*e.g.,* $C2_4$ seems less impaired than $C1_4$ which seems less impaired than $C3_4$). Further, when there was more than one cluster that consisted predominantly of the same patient group, these clusters seemed to differ by severity. For example, in the analysis without questionnaire scores, $C5_9$ and $C8_9$ predominantly consisted of bvFTD patients and mean neuropsychological scores of $C8_9$ are uniformly lower than the ones of $C5_9$ (Figure 8). Despite this general observation that clustering results may in part be explained by differences in severity, neuropsychological patterns could also be observed.

For bvFTD, one cluster that emerged in the analysis without questionnaire scores (*e.g.,* $C3_4$) yielded particularly low scores on all sub-scores of the FTLD-CDR except *Language* compared to the other clusters.

To differentiate the three PPA variants, results from the clustering analysis may be of particular use to separate svPPA patients from the other two PPA variants. Results did not support distinction of nfvPPA and lvPPA as these continuously clustered together even at k = 9. Particularly the repeat and point test may prove relevant to distinguish svPPA from the two other PPA variants. This distinction becomes most apparent when comparing $C3_9$ and $C6_9$ in the analysis with questionnaire scores consisting of mainly lvPPA/ nfvPPA patients and svPPA patients, respectively. While the svPPA cluster ($C6_9$) scores lower on the point task than the nfvPPA/ lvPPA cluster ($C3_9$), both groups seem to perform similarly on the repeat task. Additionally, the svPPA cluster seemed to perform particularly badly on the BNT, the test of verbal but not phonemic fluency and the Cookie Theft Test. Interestingly, while performing better on most other scores, the lvPPA/ nfvPPA cluster shows greater impairment than the svPPA cluster on tests involving number processing, i.e., TMT versions A and B as well as digits forward and backwards tests and the block span assessing visuospatial memory.

Inspecting clusters $C1_9$, $C5_9$ and $C8_9$ to explore possible differences between an AD-dominant and two bvFTD-dominant clusters indicates that both bvFTD clusters score lower on

most FTLD-CDR scores and this is particularly pronounced for the sub-score *Behavior, comportment, & personality* while it is not true for the sub-scores *Memory*, *Orientation* and *Language*. Interestingly, memory tests do not seem to differentiate between the AD cluster ($C1_9$) and the more severely affected bvFTD cluster ($C8_9$). Similarly, executive functions do not seem to differentiate between the two patient groups.

Inspecting results from the analysis including questionnaire scores suggests that questionnaires may aid the separation of a group of bvFTD patients that seem relatively high functioning. They may be characterized by little impairment on most variables with isolated low scores on the Apathy Scale and FrSBe questionnaire (*e.g.*, $C3_3$, $C4_4$, $C2_5$, $C5_5$). Additionally, these clusters present with low scores on the FTLD-CDR sub-score *behavior, comportment, & personality*.

Figure 7

Radar Plot Summarizing Cluster Centers (all Groups, Without Questionnaires, k = 4), Bar Plot Indicating Composition of Clusters

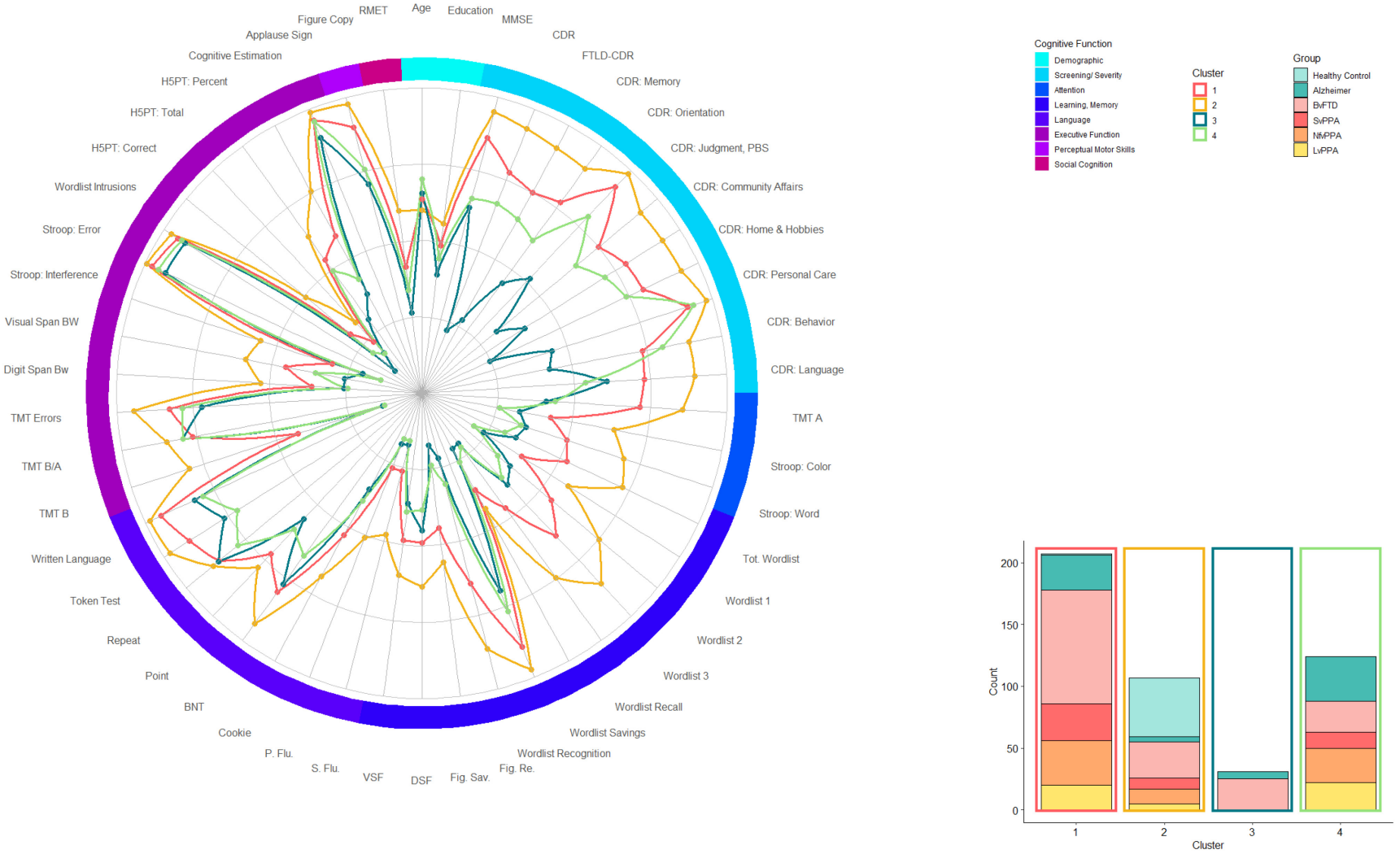
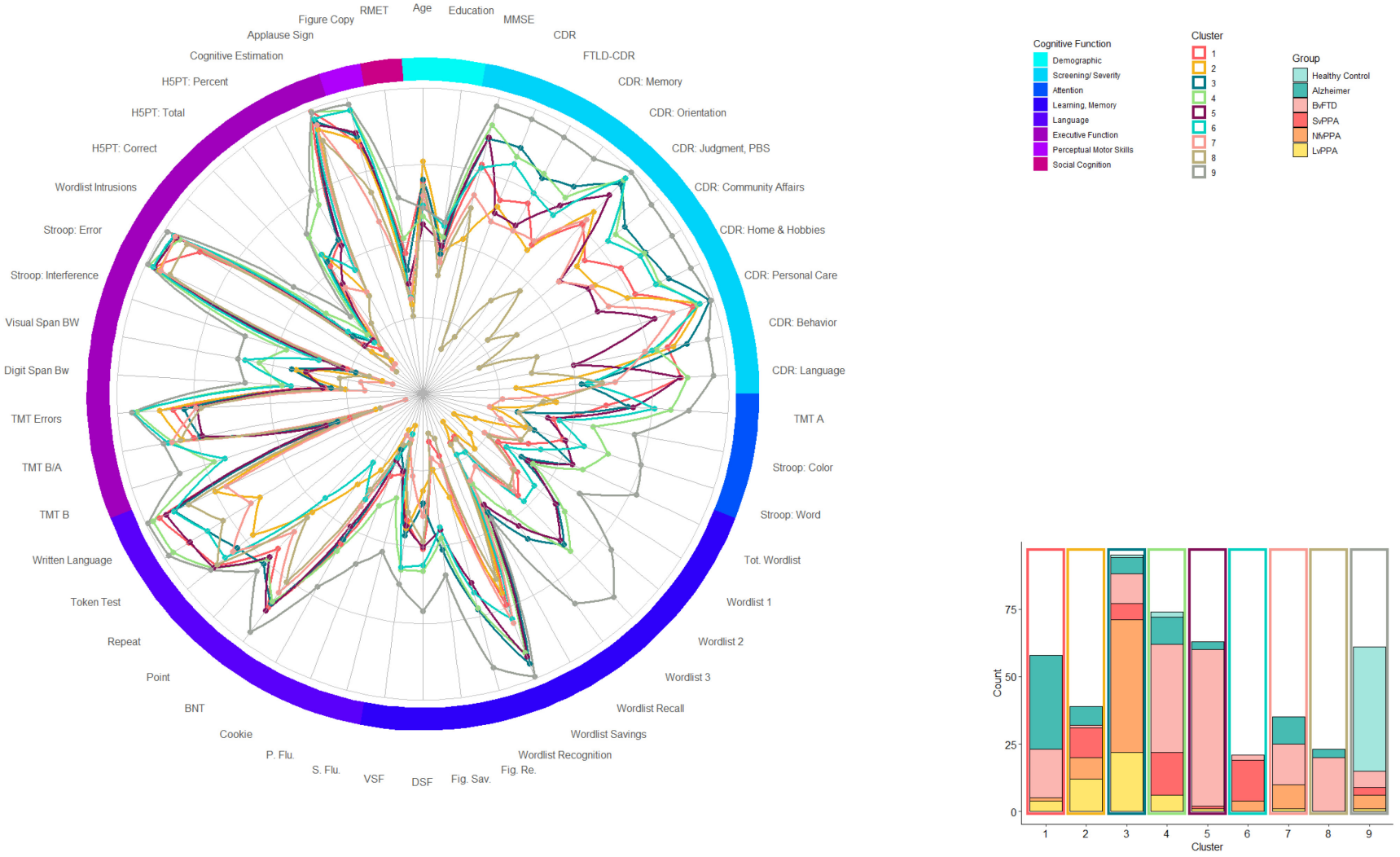


Figure 8

Radar Plot Summarizing Cluster Centers (all Groups, Without Questionnaires, k = 9), Bar Plot Indicating Composition of Clusters



3.2.2 *Only Patient Groups*

3.2.2.1 Diagnostic Groups. Overall, the analysis with all patient groups, excluding healthy controls yields similar results as the analysis with all groups. However, in comparison to the previous analysis, bvFTD patients seem to separate more easily yielding a very homogeneous cluster at k = 7 - 9 (*e.g.,* $C7_7$: 92% bvFTD patients in the analysis without questionnaires). Like the previous analysis, this homogeneous clustering of bvFTD patients seems to be further facilitated when including questionnaire scores (Figure 10). The number of bvFTD patients contained in very homogeneous clusters in the analysis without questionnaires is highest for k = 7 reaching 48% (*i.e.,* 82 out of 171 bvFTD patients in $C1_7$ or $C7_7$). When including questionnaire scores, homogeneous bvFTD clusters emerge at k = 3 and percentage of homogeneously clustered bvFTD patients increases with increasing number of clusters reaching a peak at k = 6 with 83% of bvFTD patients contained in one of the three homogeneous bvFTD clusters (*i.e.,* 143 out of 173 bvFTD patients in $C2_6$, $C4_6$ or $C5_6$). An additional observation is a cluster that emerges in the analysis without questionnaire scores (Figure 9). This cluster ($C1_9$) contains only one bvFTD patient while the other patient groups contribute to this cluster to a similar percentage (between 18% and 30%). In the light of bvFTD forming the largest group, this cluster is surprising.

Figure 9

Visualizing Clusters by their Composition of Patient Groups (Including Only Patient Groups, Without Questionnaires)

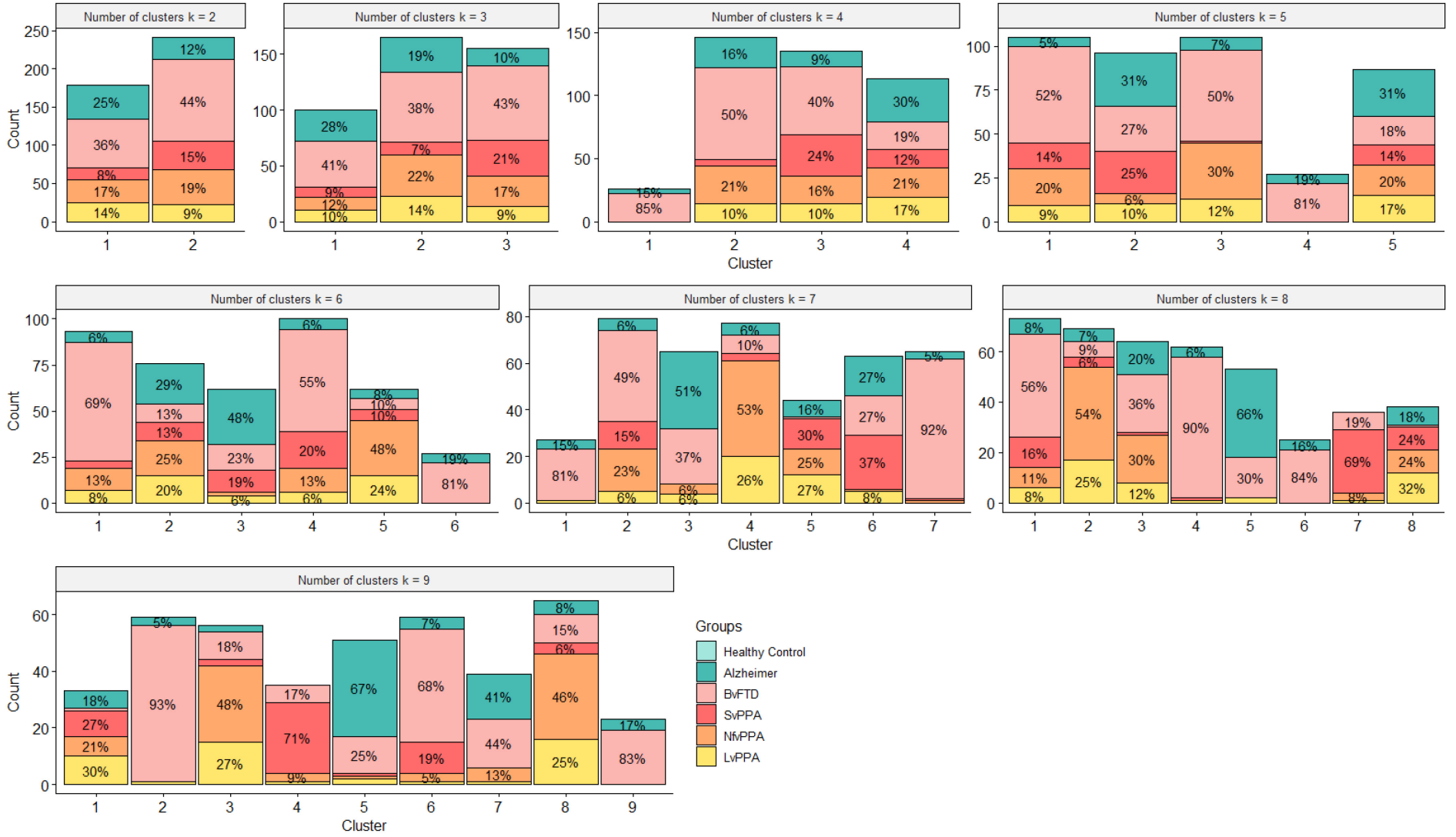
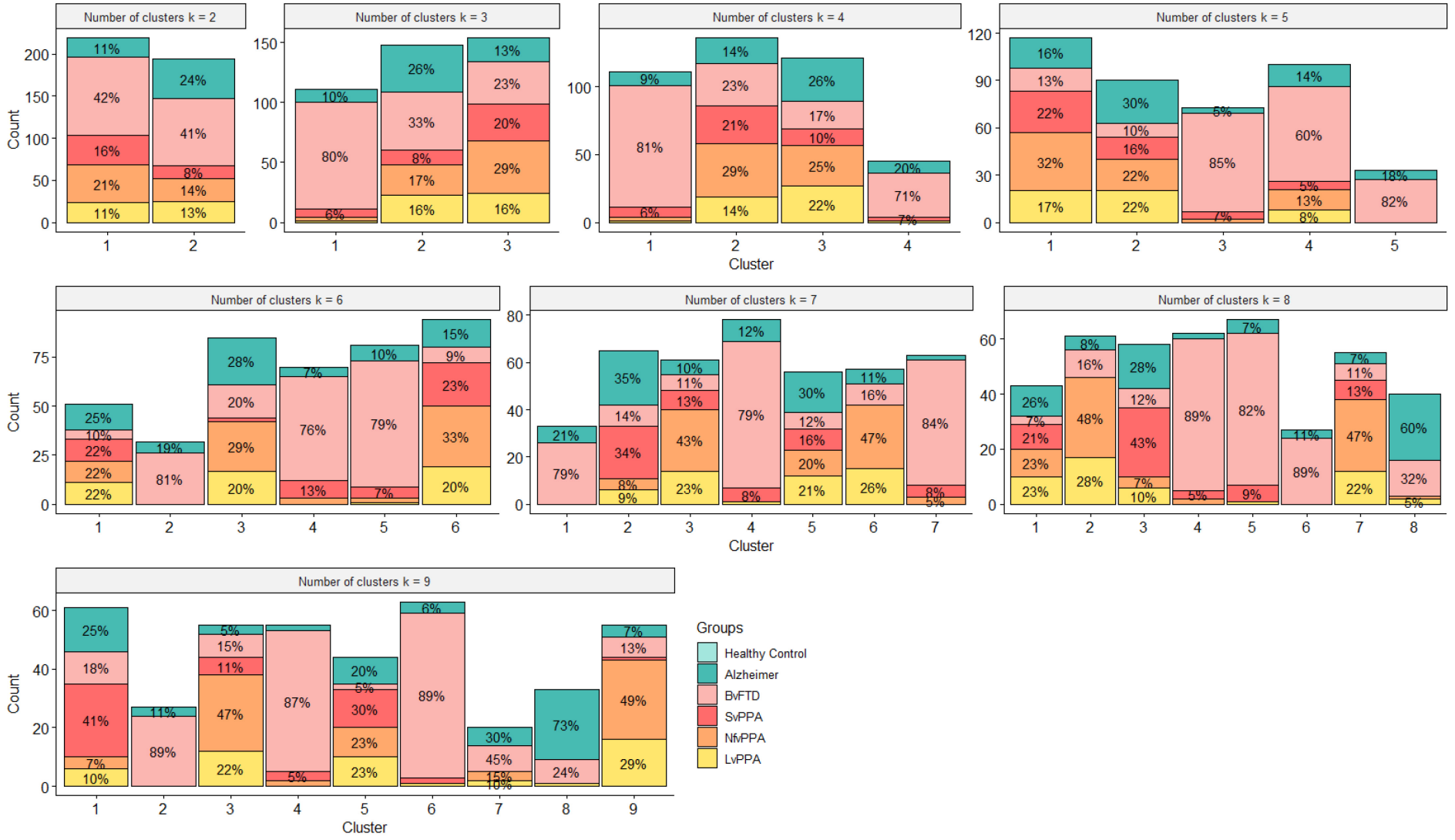


Figure 10

Visualizing Clusters by their Composition of Patient Groups (Including only Patient Groups, With Questionnaires)



3.2.2.2 Neuropsychological Profile. Description of cluster results will be exemplified in this section by clusters from the analysis including questionnaires. Comparing the neuropsychological features of the clusters which mostly contain bvFTD patients, indicates that their main difference lies in disease severity with some clusters scoring consistently worse on nearly all variables. Apparent is that each one of the homogeneous bvFTD clusters scores low on the Apathy Scale and the FrSBe sub-scores, even when highly proficient in all other domains assessed (*i.e.,* $C1_4$, $C3_5$, $C4_6$, $C7_7$, $C4_8$, $C4_9$). Additionally, simple neuropsychological tests such as the BNT and the repeat part of the repeat and point test show similarly high performance across bvFTD clusters and thus do not seem affected by disease severity. For the remaining variables, scores of the different bvFTD clusters remain in parallel. A consistent finding for homogeneous bvFTD clusters is that these clusters reached lower scores on the *Behavior, comportment, & personality* than on the *Language* domain of the FTLD-CDR. The opposite was observed for the remaining clusters. A further observation made in the analysis with and without questionnaires is the emergence of a cluster containing all patient groups but only very few bvFTD patients. This cluster (*i.e.,* $C1_9$ and $C5_9$ in the analysis without and with questionnaires, respectively) is characterized by low scores across nearly all neuropsychological assessments while showing comparatively preserved scores on the FTLD-CDR and the questionnaire scores. BvFTD patients seem to show the opposite pattern, namely low scores on behavioral assessments with relatively preserved neuropsychological scores.

Similarly, comparing neuropsychological scores of the clusters that primarily consist of lvPPA/ nfvPPA patients suggests differences due to disease severity. Parallelism is maintained across variables except for FTLD-CDR *Behavior, comportment, & personality* and for FrSBe *Executive dysfunction* and *Apathy* which remain good independent of severity (*e.g.,* $C3_9$, $C9_9$).

Comparing the lvPPA/ nfvPPA clusters (*i.e.,* $C3_9$, $C9_9$) with the svPPA cluster (*i.e.,* $C1_9$) suggests that svPPA patients score particularly low on the *Point* compared to the *Repeat* part of the repeat and point task. On the FTLD-CDR sub-score *Behavior, comportment, & personality* the svPPA cluster seems to score lower than all lvPPA/ nfvPPA clusters. Additionally, patients in the svPPA cluster seem to perform better than patients in the lvPPA/ nfvPPA clusters on parts of the Stroop and the digits and block tapping tasks.

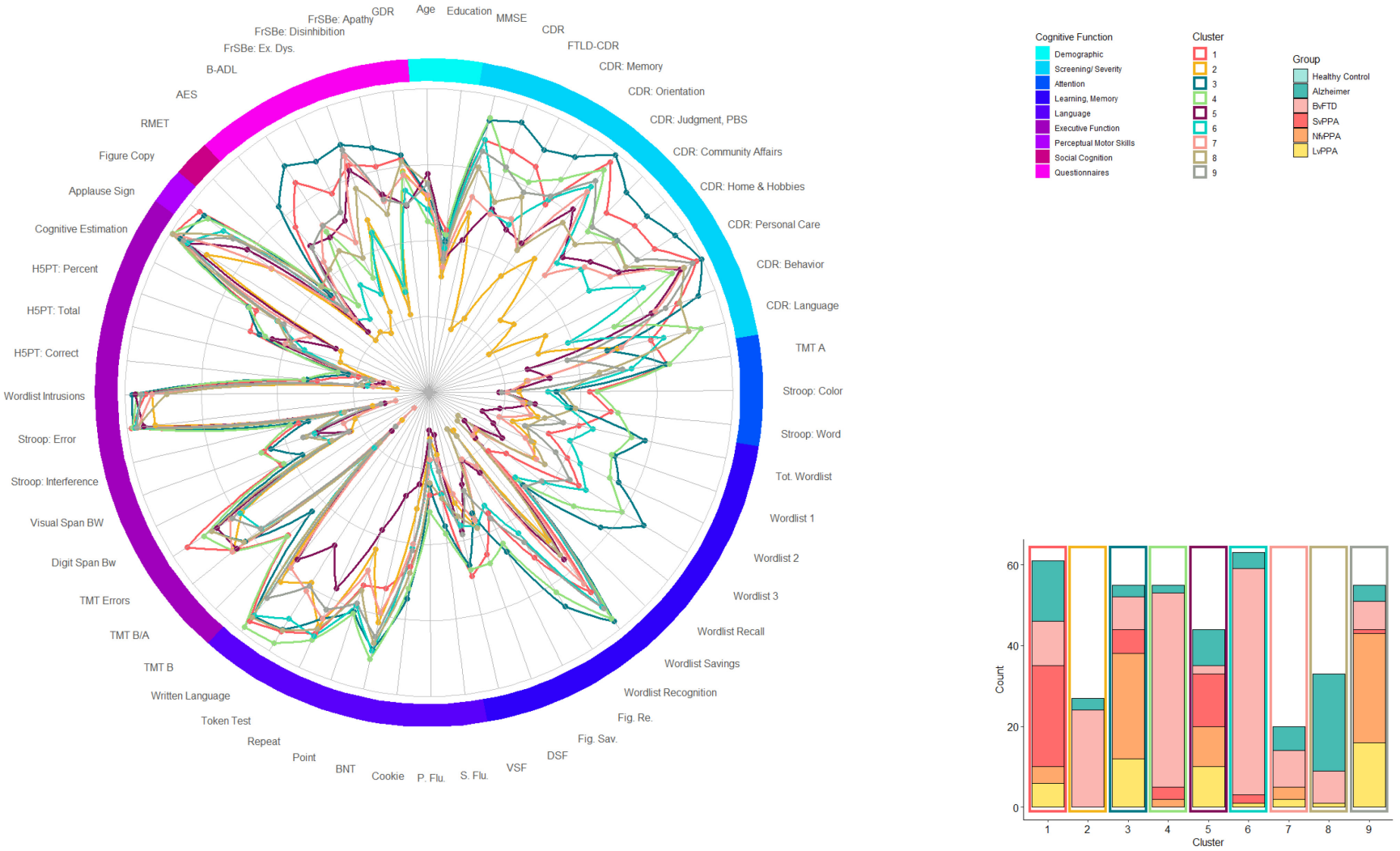
The comparison of the bvFTD clusters (*i.e.,* $C2_9$, $C4_9$, $C6_9$) with the svPPA cluster (*i.e.,* $C1_9$, $C5_9$) suggests differences between the patient groups on the FTLD-CDR. Results from the svPPA cluster are worse on the *Language* domain and better on the *Behavior, comportment, & personality* domain. Except for lower results on the BNT in the svPPA cluster, results on other language assessments were mixed. The two executive function variables block tapping

backward, and TMT error scores were better in the svPPA cluster than the bvFTD clusters. Larger differences were observed on the Apathy Scale and the FrSBe on which the svPPA cluster showed smaller behavioral impairment as reported by a companion compared to all bvFTD clusters.

Evaluation of the AD cluster (*i.e.,* $C8_9$) in comparison to the bvFTD clusters (*i.e.,* $C2_9$, $C4_9$, $C6_9$) suggests patients in the AD cluster to score lower than all bvFTD clusters only on the variables assessing memory of unstructured verbal information after a delay period (CERAD wordlist recall and word savings). The bvFTD clusters instead show lower scores on the Apathy Scale and all sub-scores of the FrSBe as well as on the FTLD-CDR sub-scores *Personal care* and *Behavior, comportment, & personality*.

Figure 11

Radar Plot Summarizing Cluster Centers (Only Patient Groups, With Questionnaires, k = 9), Bar Plot Indicating Composition of Clusters



3.2.3 PPA and BvFTD

3.2.3.1 Diagnostic Groups. Superficially, the analysis with PPA and bvFTD patient subgroups shows that with increasing number of clusters first bvFTD and PPA patients cluster apart and then svPPA patients form separate clusters from nfvPPA/ lvPPA patients (*e.g.,* $C1_5$, $C5_5$ contain mainly all three groups of PPA patients compared to $C3_5$, $C4_5$ containing mainly bvFTD patients in the analysis without questionnaires). In the analysis without questionnaire scores (Figure 12) the separation of bvFTD patients in separate clusters is incomplete with several clusters consisting to a similar proportion of bvFTD and PPA patients (*e.g.,* $C2_5$ consists of 57% bvFTD and 43% PPA patients). Clustering results become more homogeneous when including questionnaire scores (Figure 13). With questionnaires at k = 3 a clear bvFTD cluster emerges ($C3_3$) and at k = 4 there are two bvFTD ($C2_4$, $C3_4$) and two PPA clusters ($C1_4$, $C4_4$). At k greater than four, percentage of bvFTD patients clustering separately from PPA subgroups surpasses 70% and is highest for the analysis with five clusters reaching 87%. At the maximum number of clusters (k = 8 and k = 9), there does not seem to be a meaningful difference between clustering results with and without questionnaires.

Interestingly, across both analyses some clusters persist that despite consisting mainly of bvFTD patients also have a considerable proportion of PPA patients (*e.g.,* analysis without questionnaires: $C1_9$, $C6_9$; analysis with questionnaires: $C4_9$, $C9_9$). Inspection of these clusters may suggest that some bvFTD patients are more similar to svPPA or to nfvPPA/ lvPPA patients, as they seem to cluster either with one or the other patient group. Additionally, both when including or excluding questionnaires, a majority of svPPA patients cluster together and separately from lvPPA/ nfvPPA patients, while lvPPA and nfvPPA patients do not seem to cluster apart. Further, except for the analysis with questionnaire scores at k = 9, in all other analyses at least one cluster persists that contains all three PPA groups to a similar proportion.

Generally, similarity in cluster structure emerging in the analysis with and without questionnaire scores is observed. For example some clusters may be attributed to bvFTD (*e.g.,* without questionnaire scores $C6_8$, $C7_8$, $C5_9$, $C8_9$, $C9_9$; with questionnaire scores $C1_8$, $C7_8$, $C8_8$, $C1_9$, $C2_9$, $C5_9$, $C8_9$), svPPA (*e.g.,* without questionnaire scores $C1_8$, $C7_9$; with questionnaire scores: $C4_8$, $C6_9$, $C9_9$), lvPPA/ nfvPPA (*e.g.*, without questionnaire scores $C5_8$, $C8_8$, $C3_9$, $C4_9$; with questionnaire scores $C2_8$, $C3_8$, $C3_9$, $C4_9$, $C7_9$) , mixed PPA (*e.g.,* without questionnaire scores $C1_8$, $C2_9$; with questionnaire scores $C5_8$). Other mixed clusters are less comparable across the two analyses such as a cluster of mixed bvFTD and svPPA patients in the analysis without questionnaire scores (*e.g.*, $C2_8$, $C1_9$).

Figure 12

Visualizing Clusters by their Composition of Patient Groups (Including PPA and bvFTD, Without Questionnaires)

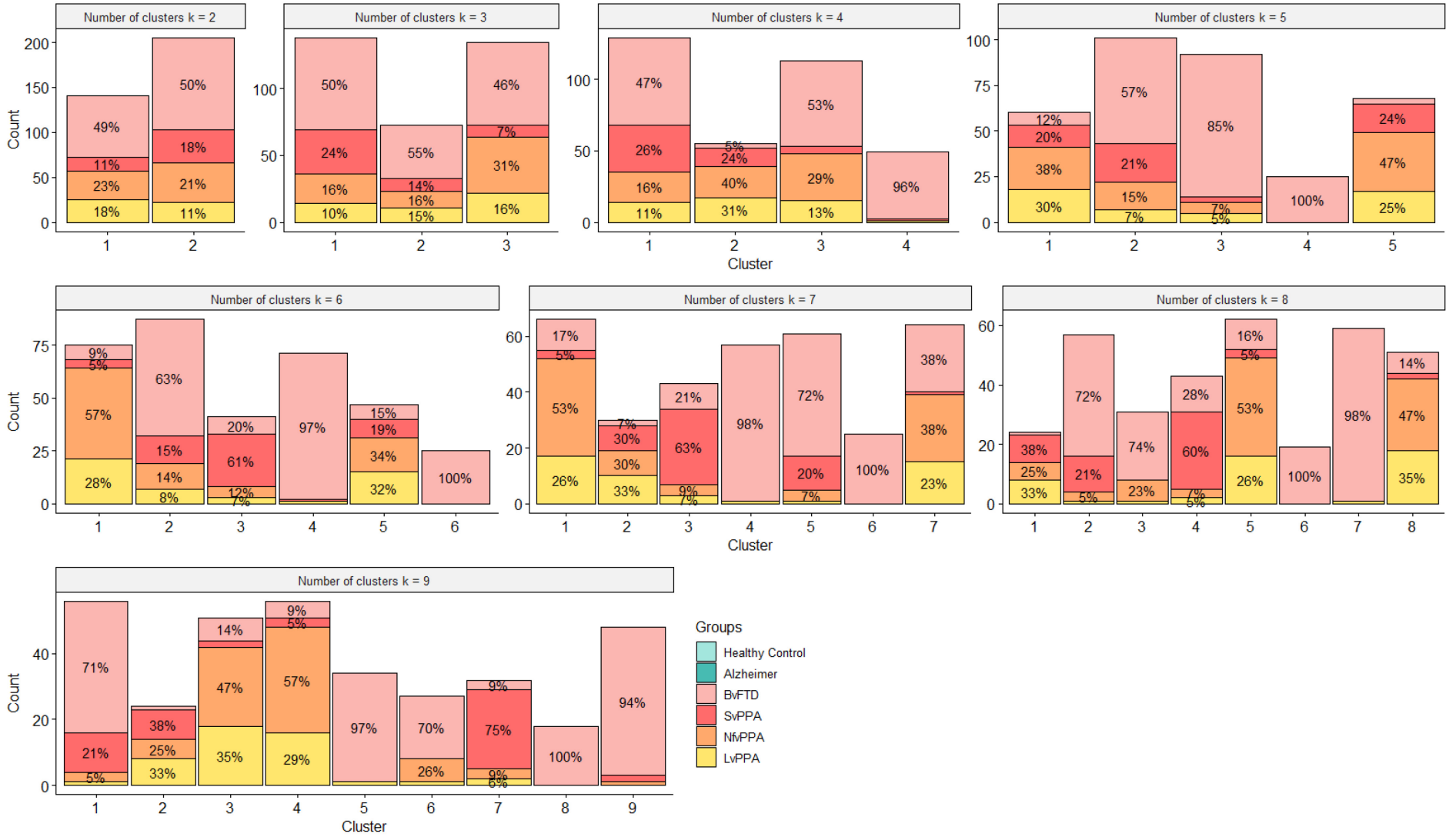
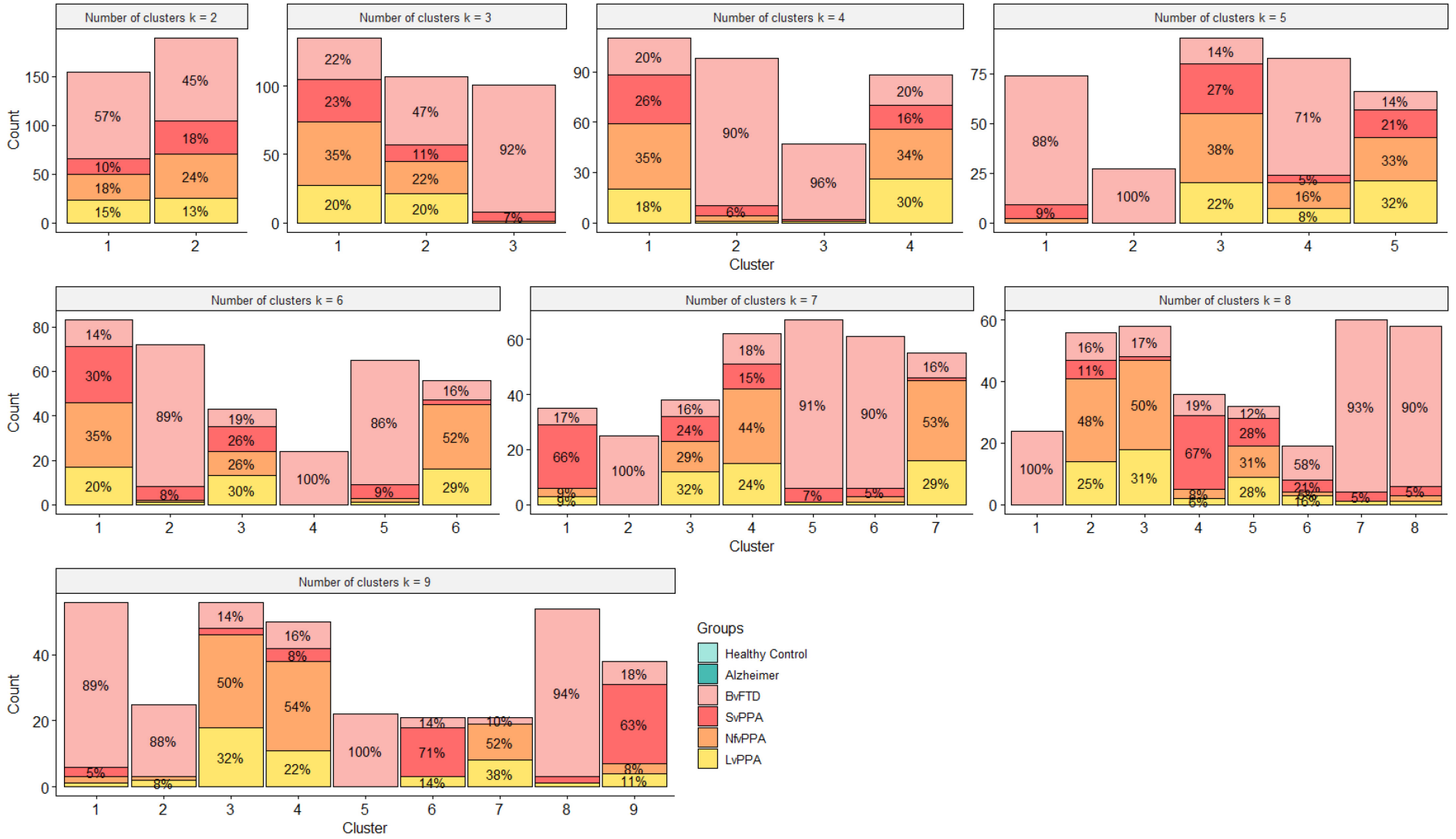


Figure 13

Visualizing Clusters by their Composition of Patient Groups (Including PPA and bvFTD, With Questionnaires)



3.2.3.2 Neuropsychological Profile. Comparing the different clusters which mainly consist of lvPPA/ nfvPPA patients (*i.e.,* without questionnaire scores $C3_9$, $C4_9$; with questionnaire scores $C3_9$, $C4_9$) suggests that the two clusters differ in severity. Compared to the other clusters, patients in the lvPPA/ nfvPPA clusters reach better scores on the FrSBe *Executive Dysfunction* subscale. Comparing these clusters to the svPPA clusters (*i.e.,* without questionnaire scores $C7_9$; with questionnaire scores $C6_9$, $C9_9$), svPPA patients seem to score lower on the point task, the BNT and the delayed memory task of word list recall and recognition but neither for immediate word list recall nor for figure recall. SvPPA clusters scored relatively better on some executive function, non-verbal memory, and processing speed scores such as Stroop, digit and block tapping span tasks.

The bvFTD clusters (*i.e.,* without questionnaire scores $C5_9$, $C8_9$, $C9_9$; with questionnaire scores $C1_9$, $C2_9$, $C5_9$, $C8_9$) seem to differ in severity. In the analysis without questionnaires one cluster (*i.e.,* $C5_9$) shows relatively preserved delayed memory recall both for words and figures. Possibly this indicates a subgroup of bvFTD patients who show relatively preserved memory function. In comparison to the svPPA clusters, the bvFTD clusters have better scores on the BNT and FTLD-CDR *Language* sub-score and lower mean scores on the variables of the Apathy Scale and FrSBe as well as the FTLD-CDR sub-scores *Orientation* and *Personal care*.

Figure 14

Radar Plot Summarizing Cluster Centers (PPA and bvFTD, Without Questionnaires, k = 9), Bar Plot Indicating Composition of Clusters

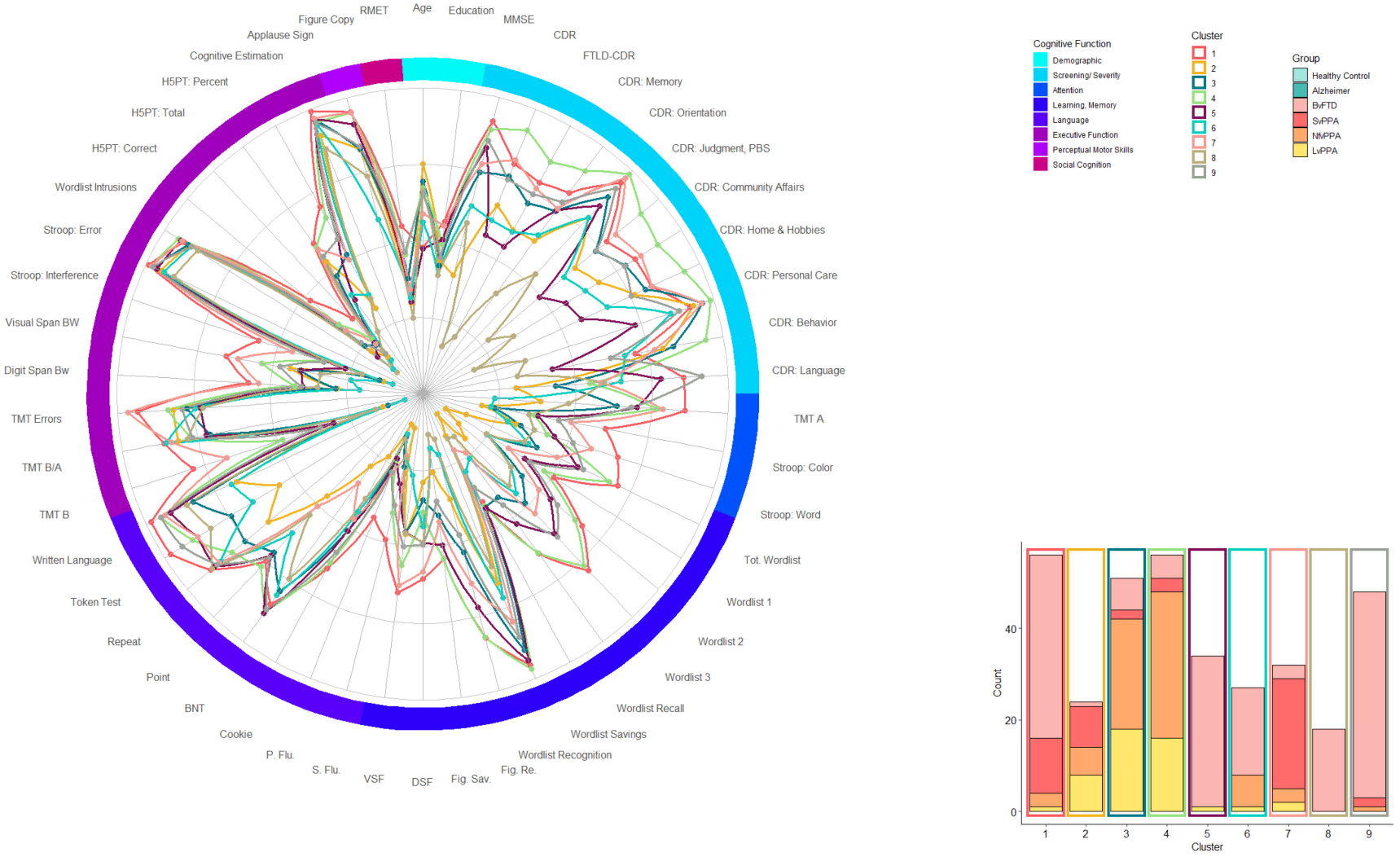
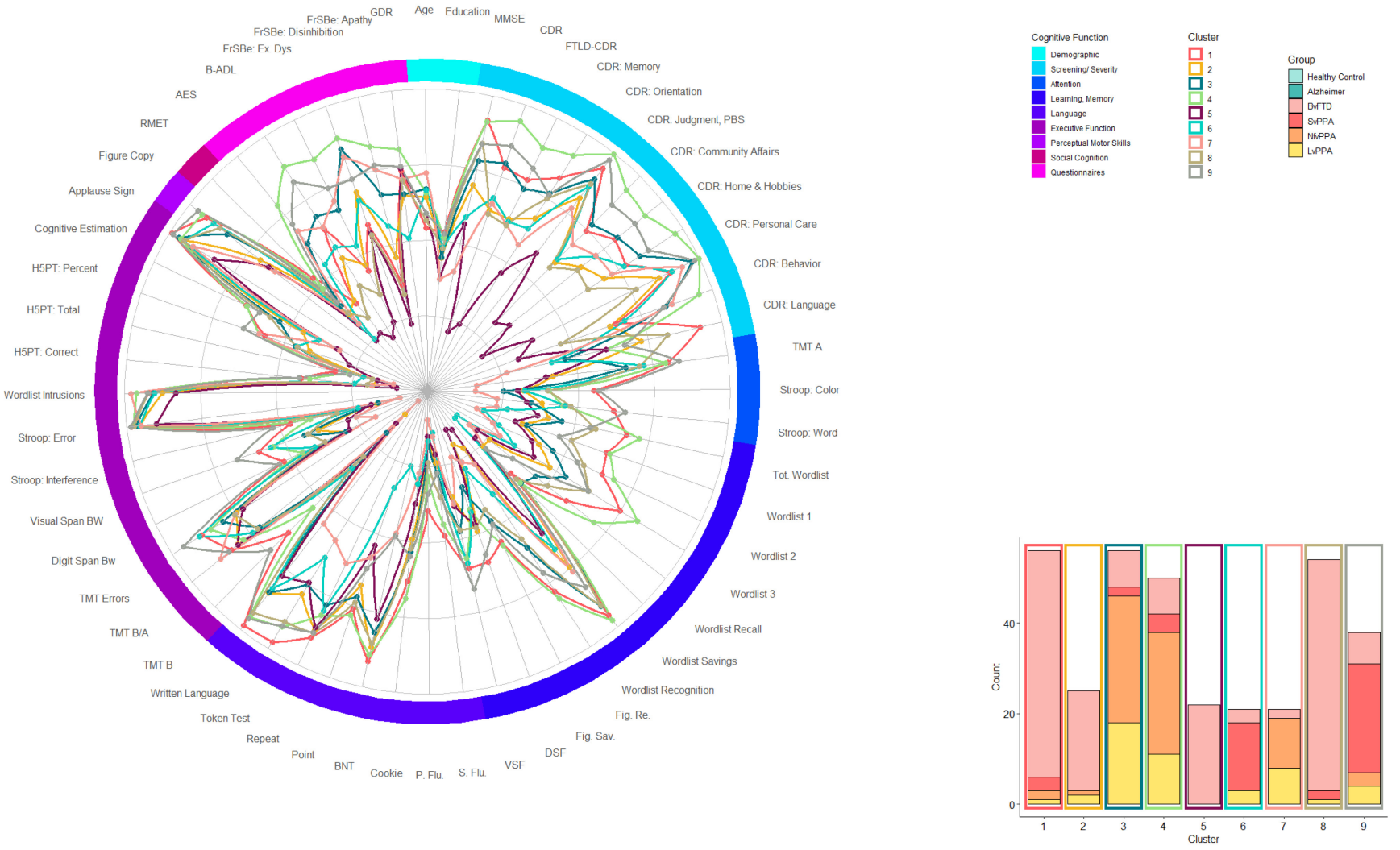


Figure 15

Radar Plot Summarizing Cluster Centers (PPA and bvFTD, With Questionnaires, k = 9), Bar Plot Indicating Composition of Clusters



3.2.4 PPA

3.2.4.1 Diagnostic Groups. Similar to the previous analysis, excluding bvFTD patients suggests separation of svPPA patients and mixed lvPPA/ nfvPPA patients in different clusters. Inclusion of questionnaire scores does not seem to impact clustering results in a meaningful way (Figure 17). Further, despite most clusters showing a predominance of either svPPA or lvPPA/ nfvPPA patients, even at the maximum number of clusters (k = 9) a mixed cluster persists consisting of a similar proportion of all three PPA variants (*e.g.,* without questionnaires $C1_9$, $C9_9$).

Figure 16

Visualizing Clusters by their Composition of Patient Groups (Including PPA, Without Questionnaires)

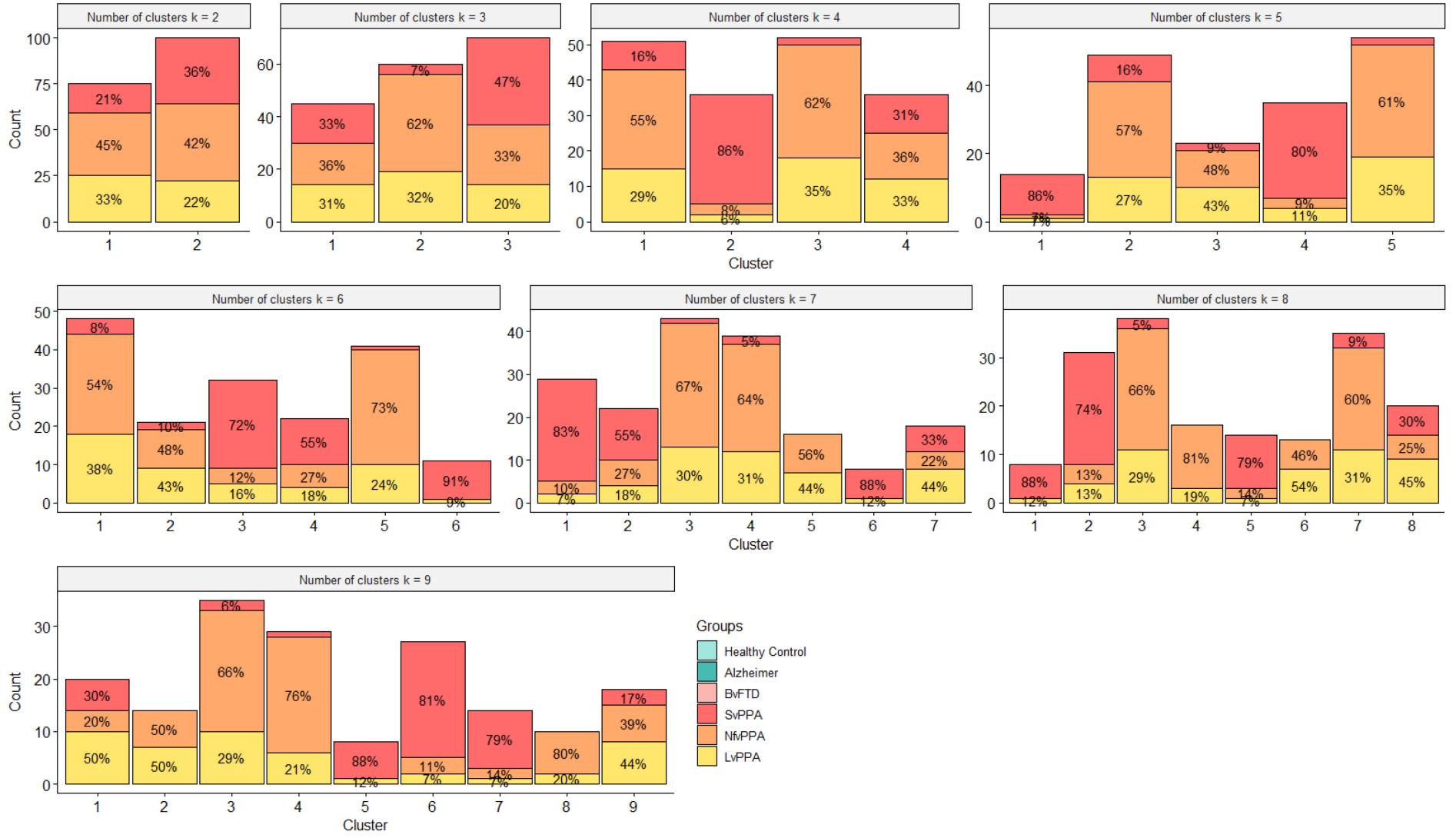
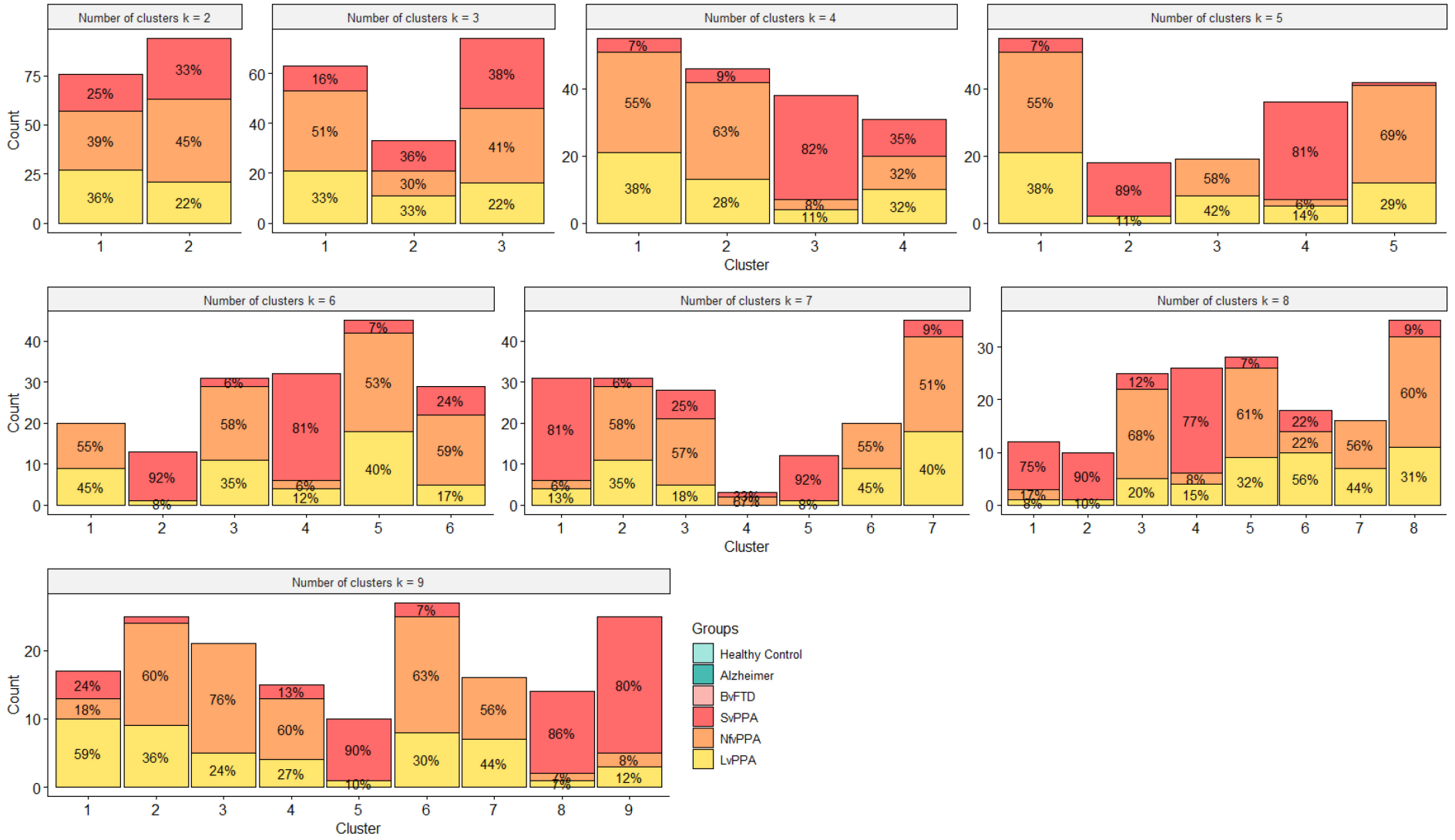


Figure 17

Visualizing Clusters by their Composition of Patient Groups (Including PPA, With Questionnaires)



3.2.4.2 Neuropsychological Profile. Examples are provided from the analysis without questionnaire scores (Figures 18 and 19). Overall, comparing the svPPA clusters (*e.g.,* $C1_5$, $C4_5$; $C5_9$, $C6_9$, $C7_9$) supports the idea that patients were clustered based on disease severity. Interestingly different clusters show highly similar results on some variables such as the repeat task, a figure copying task, the error score on the TMT and the block tapping task. In case clustering results indeed correspond to severity, cognitive function for these tests may be well preserved even in more severe forms of svPPA. Additionally, when the number of clusters was set higher (*i.e.,* k = 7-9), one cluster that may correspond to intermediate severity (*i.e.,* $C2_8$, $C6_9$) compared to the other two svPPA clusters shows the smallest impairment of svPPA clusters across FTLD-CDR sub-scores. This may suggest existence of a svPPA subgroup with relatively spared behavioral symptoms. Support for this idea comes from a corresponding cluster in the analysis with questionnaires (*i.e.,* $C4_8$, $C9_9$) which shows comparatively high scores on the questionnaires. However, these clusters are less homogeneous including up to 26% of participants with a diagnosis of nfvPPA or lvPPA. Patients with nfvPPA or lvPPA may show lower behavioral impairment and drive the observed difference.

In a similar way the comparison of the mixed nfvPPA/ lvPPA clusters (*e.g.,* $C1_4$, $C3_4$; $C2_5$, $C3_5$, $C5_5$) suggests clustering by disease severity. Scores on some variables are similar across clusters but this finding is inconsistent and varies when increasing the number of clusters.

Focusing on the contrast between nfvPPA/ lvPPA and svPPA clusters the only consistent difference is that patients in the svPPA clusters perform better on various variables from the TMT, especially on the error score and the block tapping and digit span tasks. Additionally, results corroborate the previous finding that svPPA clusters show a large drop in performance from the repeat to the point part of the repeat and point test and that this is not the case for nfvPPA/ lvPPA clusters.

Figure 18

Radar Plot Summarizing Cluster Centers (PPA, Without Questionnaires, k = 5), Bar Plot Indicating Composition of Clusters

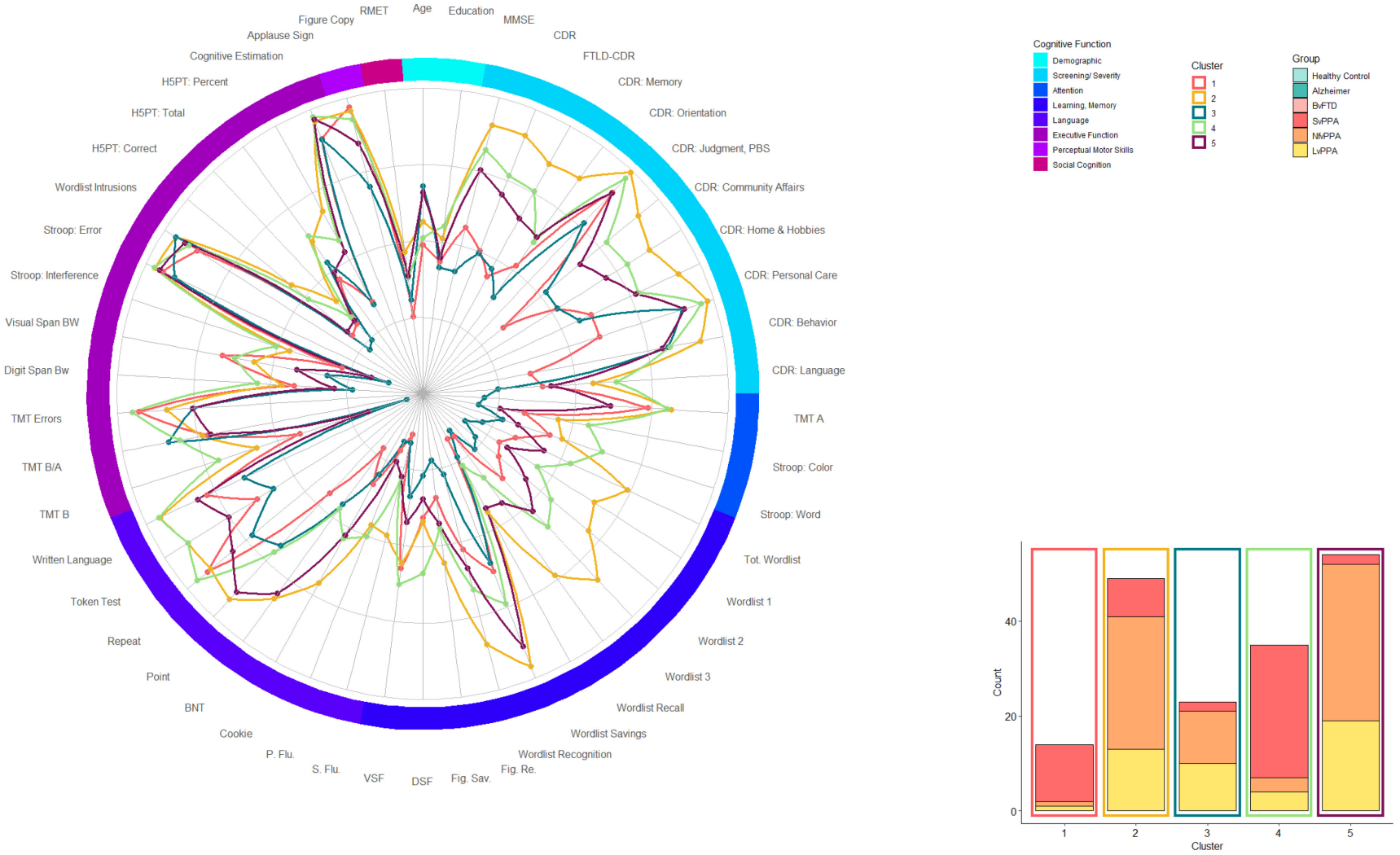
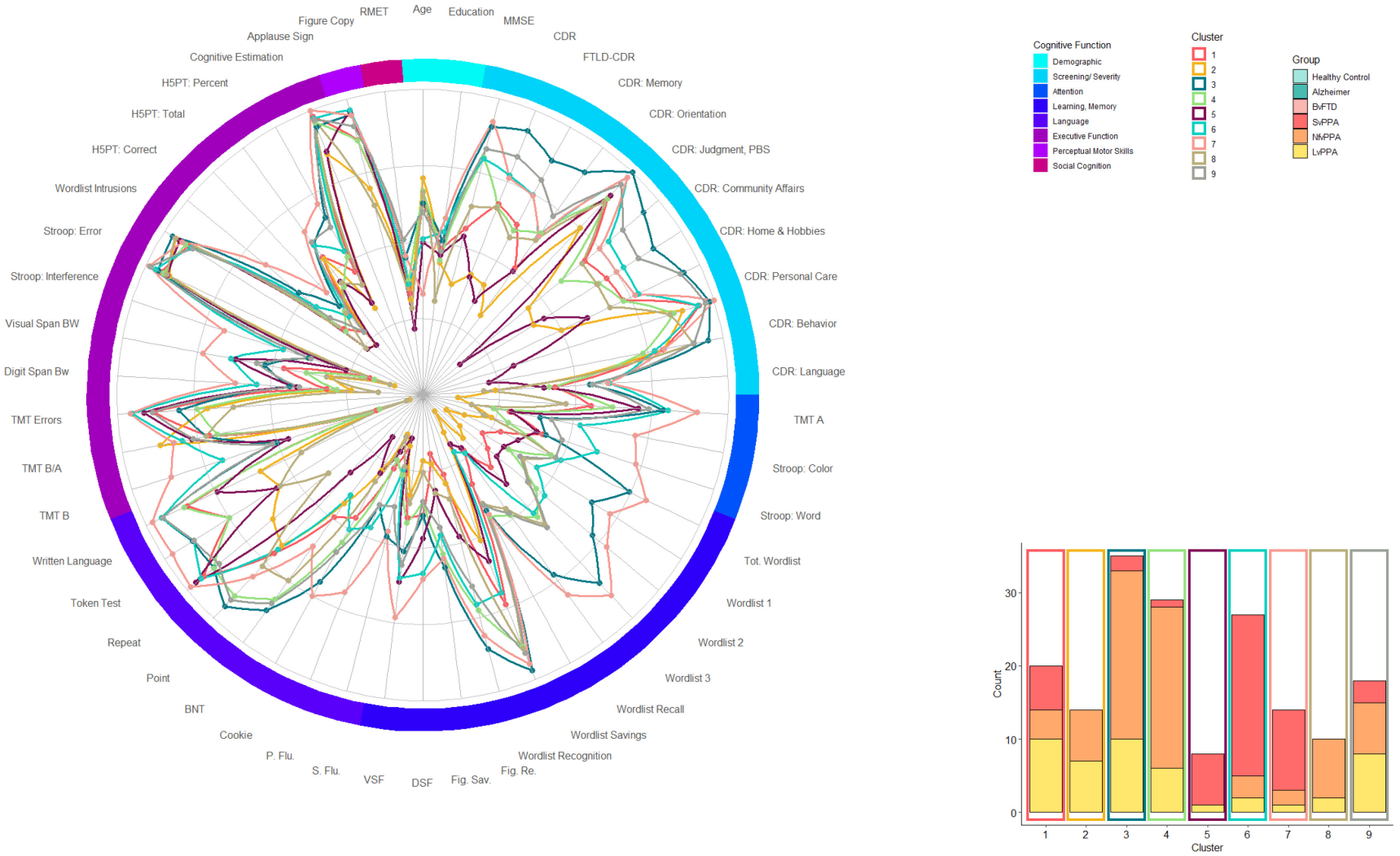


Figure 19

Radar Plot Summarizing Cluster Centers (PPA, Without Questionnaires, k = 9), Bar Plot Indicating Composition of Clusters



3.2.5 BvFTD and AD

3.2.5.1 Diagnostic Groups. The last analysis performed included only bvFTD and AD patients. Two types of clusters emerged. Several clusters consist nearly exclusively of bvFTD patients. The remaining clusters contain mainly AD patients but also a large proportion of bvFTD patients. This does not change even when increasing the number of clusters in the analysis to k = 9. Thus, a majority of bvFTD patients can be separated neatly from AD patients while a smaller proportion instead does not separate from AD patients. The maximum homogeneity yielded for AD clusters is a cluster which contains 74% AD patients. Therefore, all AD clusters have at least one quarter of patients diagnosed with bvFTD. Including questionnaire scores suggests a slight improvement of results both for k = 3 and k = 7-9. In the analysis with questionnaires, maximum homogeneity of AD clusters reaches 78% and 85% at cluster sizes k = 8 and k = 9 respectively. In general results seem comparable in the analysis including and excluding questionnaire scores (Figure 21).

Figure 20

Visualizing Clusters by their Composition of Patient Groups (Including AD and bvFTD, Without Questionnaires)

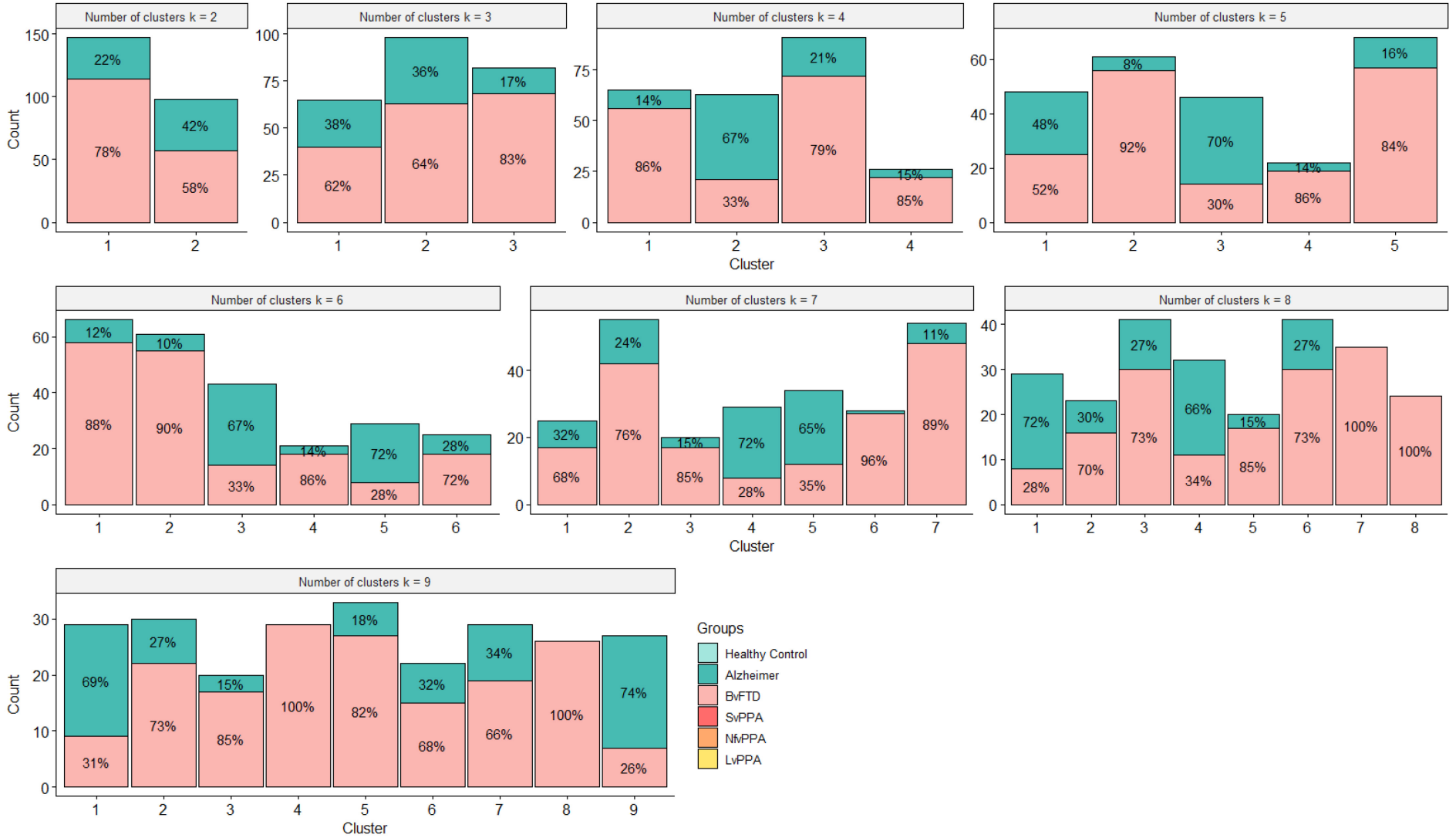
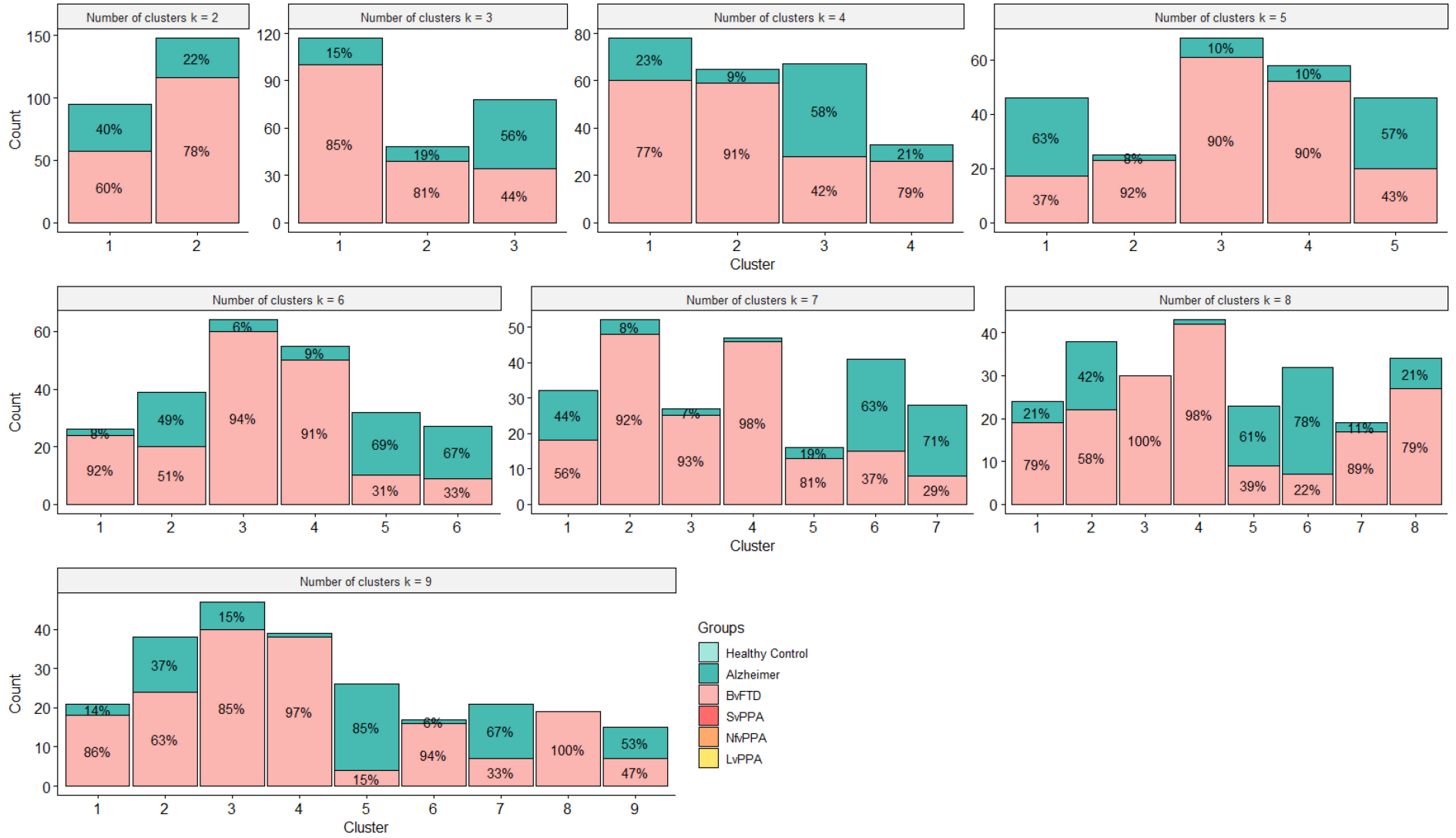


Figure 21

Visualizing Clusters by their Composition of Patient Groups (Including AD and bvFTD, With Questionnaires)



3.2.5.2 Neuropsychological Profile. Inspecting clusters which consist mainly of bvFTD patients suggests that the analysis with lower number of clusters (*e.g.*, k = 3 - 5) is based on disease severity with clusters yielding scores on the different variables that remain approximately in parallel (*e.g.,* without questionnaire scores $C1_4$, $C3_4$, $C4_4$; with questionnaire scores $C1_4$, $C2_4$, $C4_4$). Few exceptions of this are observed on the proportion of words recalled after a delay compared to immediate recall (*i.e., savings words*), the ratio of time needed on the TMT version B and version A. When including questionnaire scores, the FrSBe *Executive Dysfunction* scores and Depression Scale were either similar across clusters or did not correspond to the order of severity as observed on the other variables. With a greater number of clusters (*e.g.,* k = 7-9) clustering results were inconsistent. However, results may suggest existence of bvFTD subgroups with different neuropsychological pattern. One subgroup may show greater behavioral impairment as indicated by particularly low FTLD-CDR and questionnaire sub-scores while another subgroup may be characterized by greater attention, language and executive impairments (*i.e.,* without questionnaires $C4_6$, $C6_6$, $C1_7$, $C3_7$, $C3_9$, $C6_9$; with questionnaires $C1_8$, $C7_8$, $C1_9$, $C6_9$).

Inspection of AD clusters partly supports that clusters differ with respect to disease severity. However, similar mean scores across clusters are observed on several variables (*e.g.,* FrSBe *Personal care*, FrSBe *Behavior, comportment, & personality*, word discriminability, figure recall, TMT B/A, TMT error score, block tapping forward). Parallelity of AD clusters is not maintained for other variables such as the number of intrusions and the scores for recalling figures copied (*e.g.,* without questionnaire scores $C3_6$, $C5_6$, $C4_7$, $C5_7$, $C1_8$, $C4_8$; with questionnaire scores $C6_7$, $C7_7$, $C5_8$, $C6_8$, $C5_9$, $C7_9$).

Comparing clusters that are composed mainly of bvFTD or mainly of AD patients shows that the clusters containing most AD patients score better on the *Behavior, comportment, & personality* score of the FTLD-CDR (*e.g.,* without questionnaire scores $C2_4$, $C1_5$, $C3_5$, $C3_6$, $C5_6$). Considered alone, memory scores do not seem to differentiate AD from bvFTD clusters. However, in comparison to scores on other neuropsychological variables, AD clusters seem relatively more impacted on the memory scores compared to the bvFTD clusters. AD clusters mimic the most severely affected bvFTD cluster on memory scores while being more similar to intermediate bvFTD clusters on other variables. Including questionnaires shows that bvFTD clusters have low scores on the Apathy scale and the FrSBe subscales even when they do not demonstrate large impairment on other variables (*e.g.,* $C4_5$, $C4_6$). A general pattern is observed on the FTLD-CDR with patients from bvFTD clusters consistently scoring lower on the *Behavior, comportment, & personality* scale compared to the *Language*.

Figure 22

Radar Plot Summarizing Cluster Centers (AD and bvFTD, Without Questionnaires, k = 4), Bar Plot Indicating Composition of Clusters

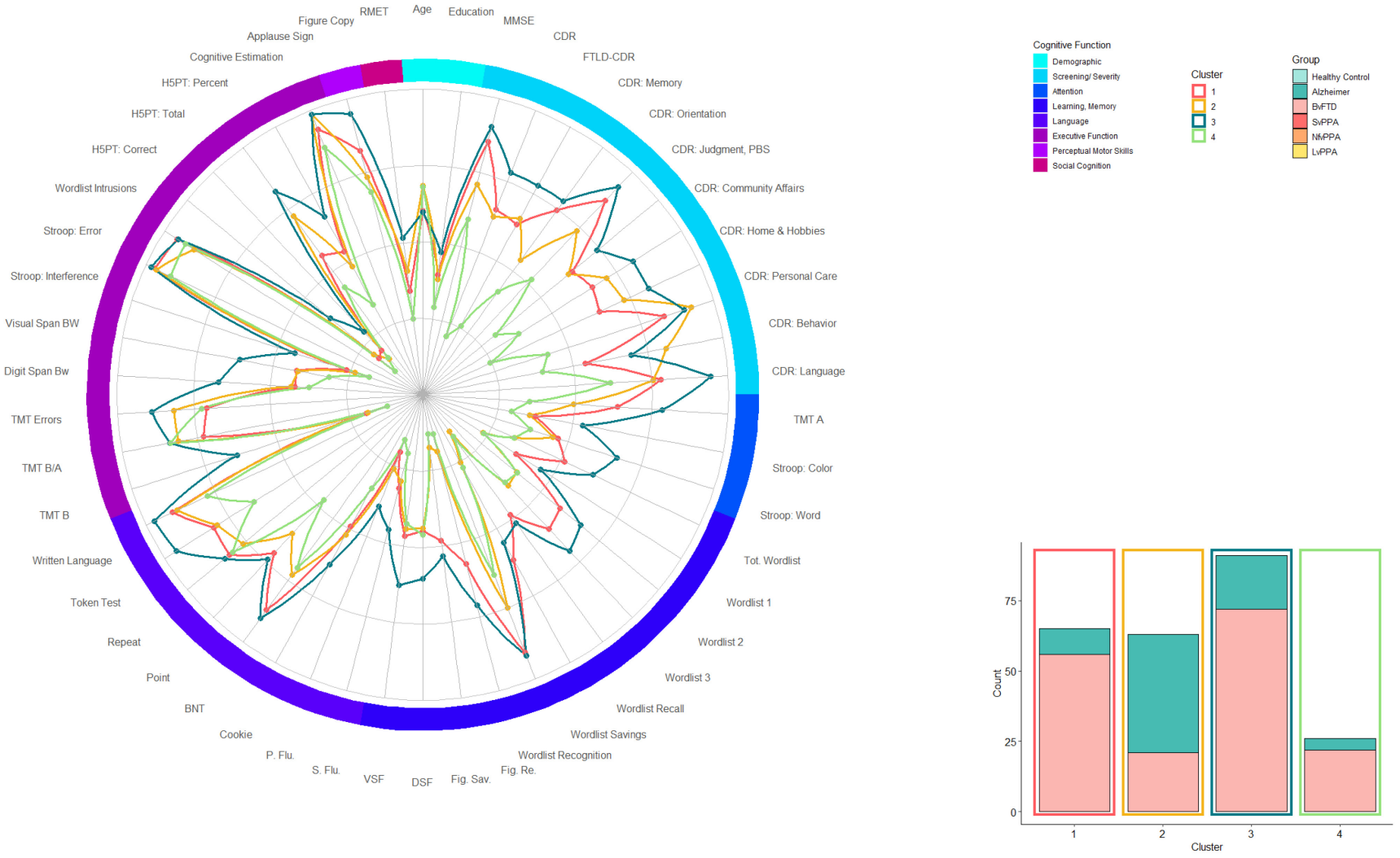
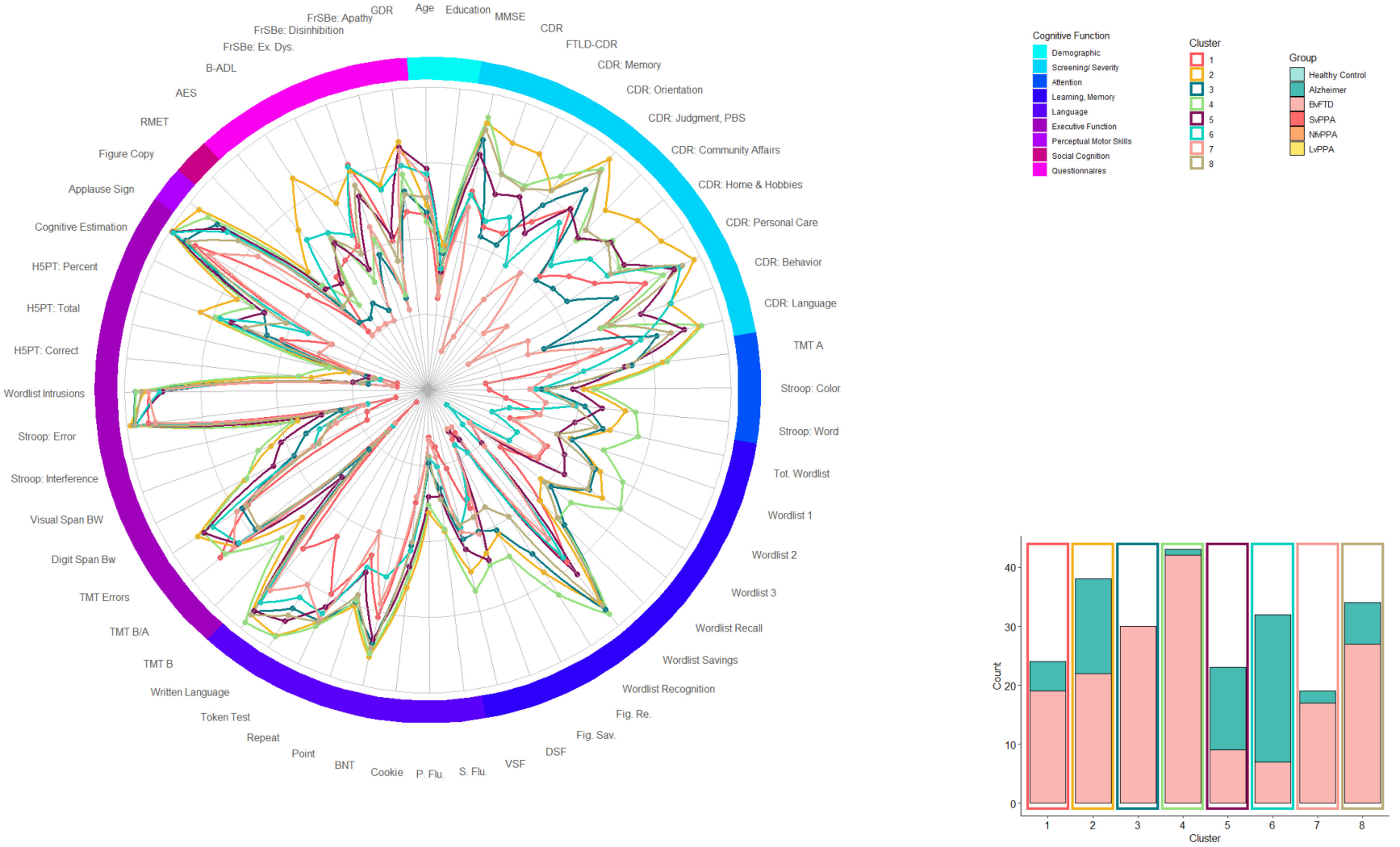


Figure 23

Radar Plot Summarizing Cluster Centers (AD and bvFTD, With Questionnaires, k = 8), Bar Plot Indicating Composition of Clusters



4 Discussion

The present study used an unsupervised ML algorithm to analyze neuropsychological data from patients with FTD. Comparison groups included in the study were healthy controls for validation purposes and patients with AD or the atypical AD variant lvPPA. Differential diagnosis of these patient groups is difficult, and misdiagnoses are common (Rascovsky & Grossman, 2013). To date, diagnostic certainty can be reached only via pathological confirmation post-mortem (Deuschl et al., 2016) and there is no gold-standard for diagnosis prior to death (Arevalo-Rodriguez et al., 2021). However, diagnostic decisions highly influence the treatment options considered by doctors. No cure exists neither for AD nor for FTD (Farouk & Rady, 2020; Panza et al., 2020). Pharmacological treatments have proven useful to reduce progression or to minimize specific symptoms for AD but not for FTD (Deuschl et al., 2016). When drugs approved for AD were administered to FTD patients, no consistent effects were found and treatment with Donepezil, a common drug for treating AD (Deuschl et al., 2016), was associated with worsening of neuropsychiatric symptoms (Ljubenkov & Boxer, 2021). The goal of the current analysis was to investigate the heterogeneity within diagnostic groups as well as differences between them. These differences may inform future clinical decisions and support the search for effective therapeutic strategies (Musa et al., 2020).

Various studies investigate the applications ML algorithms can find as decision support systems in the medical context (Álvarez et al., 2019; Cohen et al., 2021; de Bruijne, 2016). In contrast to supervised learning which is trained to find a pattern linking input variables and class labels in the data, unsupervised learning is independent of labelling. In the case of dementia, where diagnostic uncertainty persists, it may be a main advantage to investigate patterns in the data without being restricted by existing labels. Rather than simply minimizing the error of classification in comparison to the current diagnostic standard, this exploratory approach may help to find new patterns and put current diagnostic procedures into question in the search of more clinically relevant diagnostic standards (Escudero et al., 2011; Farouk & Rady, 2020). Restricting the analysis to behavioral data allows for implementation of findings in a clinical setting with sparse resources. This study was exploratory in nature and conclusions need to be interpreted with care and tested more rigorously in future studies.

4.1 Summary of Results

To gain insight into the structure of the data, the analysis was performed repeatedly varying a) the patient groups included, b) whether, additionally to demographic and neuropsychological test scores, the analysis also included questionnaire scores, and c) the number of clusters defined (*i.e.,* varying from k = 2 to k = 9 for each analysis). Several

observations proved consistent across these different analyses. Firstly, coherence between patients' diagnosis and the cluster results of our analyses was partly shown, especially for bvFTD and the distinction of svPPA and the two other PPA variants. This suggests that there is a correspondence between grouping of participants when using only behavioral data and their diagnosis. Unsupervised clustering can thus characterize the patients on cognitive functioning and support the current diagnoses without being trained on diagnostic labels (Escudero et al., 2011).

Second, the inclusion of questionnaire scores may be useful particularly, if not singularly to distinguish bvFTD from other patient groups. Questionnaire scores that seem of special relevance are the FrSBe and the Apathy Scale. Interestingly, when including only bvFTD and AD patients, clustering results seem very similar in both, the analyses with and the analyses without questionnaire scores. Thus, in the current analyses these scores do not seem to influence separation between AD and bvFTD meaningfully. Questionnaire scores may instead be of particular use to separate bvFTD from PPA and this may be even more valid at low levels of disease severity when cognitive functioning in bvFTD is largely intact. Usefulness of the other two questionnaire scores included (*i.e.*, B.ADL and Depression Scale) as well as usefulness for differential diagnoses between non-bvFTD patients seems limited.

Thirdly, by varying both the patient groups included and the number of clusters defined (*i.e.,* parameter k) in the analysis we observed that increasing the number of clusters in the k-means clustering analysis has an effect of "zooming in" to the data structure. This allows to understand the grouping of different participants independent of diagnosis but also to assess which patient groups separate more easily from the remaining participants based on the data used. By inspecting cluster centers, one may then infer the factors influencing the separation of different clusters. The zooming effect may be of special value, when including a greater number of diagnostic groups. When including healthy control participants, we observed that all healthy control subjects grouped together. However, in opposition to what might have been expected, they did not separate well from other patient groups. Instead, they were clustered with participants from all other patient groups who were characterized by very low cognitive symptoms. By increasing the number of clusters in the analysis, healthy controls progressively separated better from patient groups, forming a more homogeneous cluster. In a similar way, some patterns that are observed when including fewer patient groups can already be observed when increasing the number of clusters in the analysis with all patients. For example, in the analysis without questionnaire scores, defining k = 9, one cluster of mixed lvPPA/ nfvPPA patients (*i.e.,* $C5_9$) and one cluster of mainly svPPA patients (*i.e.,* $C6_9$) emerges. This

observation that svPPA patients cluster separately from the other two PPA groups is replicated when including only PPA patients. One conclusion from this observation could be that k-means clustering is very suitable for analyzing data from various patient groups at once. Despite such analyses having high clinical relevance, including patient groups from various etiologies in unsupervised learning is not common and to our knowledge the current study was the first one to jointly investigate AD, bvFTD and PPA in this way. The emergence of large-scale consortium data may facilitate such research in the coming years.

Fourth, heterogeneity observed between patients was large. This was noted between but also within diagnostic groups and made separation of diagnostic groups fuzzy. Clusters containing patients from more than one diagnostic group were observed in every analysis performed. Additionally, focusing on mean scores of homogeneous clusters containing primarily patients from one diagnostic group, showed that patients with the same diagnosis may yield very different scores on neuropsychological assessments. Results on most cognitive tests did not support the existence of disorder-specific impairments. Instead, additional to the disorder, disease severity was most probably a main driving factor for differences in performance between participants. At low number of clusters ($k = 2$ and $k = 3$) cluster centers did not show patterns specific to variables or cognitive domains assessed but instead could be characterized by low, intermediate, or high scores on virtually all variables. In these cases, disease severity may have been the main factor driving clustering results. When the number of clusters was set higher, patterns other than disease severity could be observed.

Based on the patterns observed some hypotheses can be formulated concerning cognitive function across diagnostic groups. These need to be tested in future studies to investigate whether they may indeed facilitate distinction of patient groups.

a) Differentiation of svPPA and nfvPPA/ lvPPA based on the repeat and point task: Not a single score but instead a drop of performance from the repeat to the point task may be indicative of svPPA symptomatology. This was recently suggested by several studies (M. Henry & Grasso, 2018; Seckin et al., 2022).
b) Differentiation of bvFTD and PPA based on companion-reported behavior and cognitive assessments: Results suggest that bvFTD is characterized by low scores on the companion-rated versions of the Apathy Scale and the sub-scores of the FrSBe Scale. Based on our findings, this may be true independent of disease severity. Relatively low scores on the questionnaires despite preserved cognitive functioning as assessed by formal neuropsychological testing, may distinguish bvFTD from PPA even at low levels of disease severity. Current diagnostic criteria

for bvFTD include the presence of apathy and disinhibition as behavioral symptoms indicative of bvFTD (Rascovsky et al., 2011). Based on the current study, both the Apathy Scale and the FrSBe seem appropriate to assess the presence of these symptoms.

c) Executive problems in bvFTD are not well-reflected using neuropsychological assessments: In the current study, bvFTD was related to lower scores on the FrSBe *Executive Dysfunction* scale while neuropsychological assessments of executive function yielded inconsistent results. As mentioned previously the inclusion of a summed executive error index may help to detect executive problems in bvFTD (Kamath et al., 2019; Kramer et al., 2003). Classically four distinct cognitive abilities - working memory, inhibition, set shifting and fluency – are considered executive functions and are required for goal-oriented behavior (Rabinovici et al., 2015). Possibly, patients with bvFTD do not necessarily show executive problems in formal neuropsychological tests of executive function. However, the effects of changed executive functioning may become apparent for close others when observing the patient in their daily life. Behavioral rather than cognitive measures may therefore be more appropriate to assess dysexecutive symptoms. For clarity, diagnostic criteria may be adapted to include dysexecutive behavior rather than a neuropsychological profile characterized by executive deficits.

d) The usefulness of social cognition assessments for distinction of AD and FTD patient groups is limited: Despite previous studies suggesting social cognition to differentiate bvFTD from healthy controls (Schroeter et al., 2018) and other neurodegenerative or psychiatric diagnoses (Ducharme et al., 2020; Gossink et al., 2018) the current study does not find supporting evidence. Current results should however be interpreted with care, as social cognition is a complex construct involving various cognitive functions (Dodich et al., 2021). Additionally, in the current study it was assessed with a single variable, namely the RMET, which had a relatively high percentage of missingness (*i.e.,*13-14% in both analyses). A previous meta-analysis did not find significant differences in performance on social cognition tasks between FTD groups but this study did not include neither AD nor lvPPA (Kamath et al., 2019). Further studies may focus on alternative tests of social cognition such as the Social Cognition and Emotional Assessment (SEA) combining five subtests (Bertoux et al., 2012), the ultimatum game (Hinterbuchinger et al., 2018), various social cognition tasks presented by Rankin (2021) for the study of

social cognition in FTD or spontaneous social behavior (Rankin et al., 2008).

e) Greater memory deficits relative to general cognitive impairment may be specific to AD: The results from our analysis suggest that memory impairments in bvFTD patients may well be as pronounced as in AD. However, it may be possible to dissociate the two disorders by evaluating memory relative to general cognitive functioning as assessed by the (FTLD-) CDR. Our results suggest a dissociation between AD and bvFTD. BvFTD patients show relatively preserved memory and (FTLD-) CDR scores or high impairment on both dimensions. AD instead seems to be relatively more impaired on memory functions. AD clusters in our study mimicked better performing bvFTD clusters for the (FTLD-) CDR scores while mimicking more impaired bvFTD clusters for memory scores. Results from this study need to be interpreted with care as AD patients did not form highly homogeneous clusters. Most AD clusters contained around 20%-30% of patients with a diagnosis other than AD, which may have distorted findings on the neuropsychological cluster centers. Future studies may investigate possible within-participant contrasts to express memory-dysfunction with relation to general cognitive functioning. A retrieval-specific memory dysfunction in bvFTD compared to AD as previously suggested (Ahmed et al., 2021; Musa et al., 2020) was not supported by the current study. This is consistent with results reported in a study by Glosser et al. (2002), in which patients with bvFTD did not show to profit from cues at recall to a greater extent than patients with AD.

f) Contrasting FTLD-CDR scores of *Language* and *Behavior, comportment, & personality* may be useful to screen for bvFTD but not for PPA variants: Comparing FTLD-CDR sub-scores bvFTD clusters seemed to consistently show a drop in mean scores from the *Language* to the *Behavior, comportment, & personality* sub-score. For other patient groups FTLD-CDR scores were more mixed. We hypothesize that a difference score of these two FTLD-CDR sub-scores may be sensitive to bvFTD and useful as a screening tool.

g) Involvement of non-language deficits in PPA is highly heterogeneous: The hallmark of PPA is a language deficit and diagnostic criteria for being considered with a diagnosis of PPA include language as the main, most pronounced and most debilitating of the symptoms in the initial phase of the syndrome (Gorno-Tempini et al., 2011). Results from the current analysis suggest that cognitive involvement in the PPA variants is larger than may be expected. Impairment in PPA being specific

to language function is not supported based on our results from formal neuropsychological testing. This was found despite data in the current study being restricted to patient's first visit and average delay since symptom onset for PPA patients being less than three years. Language deficits may have influenced results on tests of other cognitive domains such as reduced scores on the verbal learning task in svPPA but this does not explain the entirety of results.

One consistent finding of the current study was that separation of svPPA and nfvPPA/ lvPPA clusters seemed, except for differences on the repeat and point task, mainly driven by differences on assessments of processing speed, short term and working memory. Studies investigating neuropsychological impairment in the PPA variants suggest lvPPA to show the greatest cognitive deficits of the three variants and that they show faster progression of cognitive decline (Kamath et al., 2020). A meta-analysis found greater impairment of lvPPA on measures of attention and mathematical skills compared to both nfvPPA and svPPA, on memory assessments compared to nfvPPA and on executive functioning, processing speed and visuospatial skills compared to svPPA (Kamath et al., 2020).

Several studies suggest working memory impairment in lvPPA and nfvPPA (Grossman, 2010; Harris et al., 2019; Libon et al., 2007; Ruksenaite et al., 2021). Specifically, the difference of verbal and visuospatial working memory impairment in these patient groups as assessed by the digit span and the visuospatial span tasks has been investigated repeatedly. LvPPA is consistently shown to be impaired on verbal memory span tasks (Foxe et al., 2020; Foxe et al., 2013; Gorno-Tempini et al., 2008; Meyer et al., 2015). Concerning impairment on visuospatial span tasks results are less consistent. While some studies do find patients with lvPPA to be impaired on assessments of visuospatial memory span (Foxe et al., 2020; Foxe, Cheung, et al., 2021), other studies do not support this (Foxe et al., 2013; Gorno-Tempini et al., 2008). The latter explain the difference of verbal and visuospatial memory deficits by an impairment specific to phonological processing in lvPPA. Patients with nfvPPA seem more impaired on tasks assessing verbal memory (Foxe et al., 2020; Foxe, Cheung, et al., 2021; Kamath et al., 2020) while svPPA patients are shown to be spared from impairment on both types of tasks (Foxe et al., 2020; Foxe, Cheung, et al., 2021).

Interestingly, previous studies have focused on these tests to discern lvPPA from the other two variants. The current study does not support the distinction of

lvPPA and nfvPPA based on neuropsychological assessments. Findings however are in accordance with previous studies that find patients with svPPA to be spared on both verbal and visuospatial memory span tasks compared to the other two PPA groups. Additionally, visuospatial memory function assessed by figure copying and delayed recall tasks have been suggested for differentiation of the three variants (Marshall et al., 2018; Watson et al., 2018). This was not supported by our findings. Further, studies have suggested behavioral changes to be typical of svPPA (Grossman, 2010; Harris et al., 2019; Ruksenaite et al., 2021). This could not be consistently observed in the current study. Future studies may include variables assessing apraxia (Harris et al., 2019) and mathematical skills (Kamath et al., 2020) which have both been related to differences between the PPA variants.

h) Hierarchical diagnostic procedure: A particularly high score on the *Language* item of the FTLD-CDR, indicating high levels of impairment, may be sensitive to PPA. Formal, quantitative neuropsychological testing of language deficits may instead be useful and required to distinguish the different PPA variants, particularly svPPA from nfvPPA/ lvPPA. These combined findings may suggest that the diagnostic procedure could profit from a hierarchical framework. In a first step, general language function could distinguish PPA from other patient groups. In a second step, more specific neuropsychological testing would then be required for diagnostic precision.

4.2 Limitations of the Current Study

For appropriate interpretation of the current results, several limitations need to be considered. There are some methodological decisions made that may have influenced the results. The common approach in k-means clustering analysis is to evaluate the number of clusters for the analysis based on a defined criterion such as the maximization of the Silhouette (*e.g.,* Khedairia & Khadir, 2022) or Dunn index (Șenbabaoğlu et al., 2015) or the minimization of the Davies-Bouldin index (Matias-Guiu et al., 2018), to name few internal criteria available. We repeated the k-means clustering analysis by systematically varying the number of clusters between two and nine. This approach may be criticized, as internal validation of the number of clusters defined was not performed. In contrast to previous studies, the current study did not aim to find the "true" number of clusters in the data. The goal was not to find possible subgroups in the included patient groups. Instead, we intended to explore data structure. By evaluating results on a number of cluster analyses, we were able to formulate hypotheses that were observed independently of specific number of clusters.

Results from previous studies may show inconsistencies due to differences in clustering algorithm, number of clusters inspected, and clinical subgroups included. Systematically varying both patient groups included in the analysis and the number of clusters defined, creates replications of findings independent of the exact parameters. However, based on previous studies, one may argue that the maximum number of clusters set to nine is not sufficient to adequately inspect clustering of patients in the analysis with multiple diagnostic groups. Previous unsupervised clustering studies suggest the existence of several subgroups for bvFTD (Cerami et al., 2016; Ranasinghe et al., 2016; Whitwell et al., 2009), lvPPA (Machulda et al., 2013; Owens et al., 2018), nfvPPA (Matias-Guiu et al., 2019) and AD (Alexander et al., 2021; Scheltens et al., 2016; van der Vlies et al., 2009). SvPPA instead may be characterized by a single cluster without further segregation into subgroups (Matias-Guiu et al., 2018, 2019). In conclusion, repeating the analysis with higher number of clusters may be useful when including all participant or all patient groups. With less patient groups, this may not provide additional information, instead complicating interpretability of results.

For simplicity reasons, gender was not included as demographic variable in the analysis. K-means clustering is operationalized to minimize clustering error, most commonly assessed via the sum of the squared Euclidean distance. This measure is numeric and does not easily allow for integration of categorical variables in the analysis (Likas et al., 2003). Corresponding algorithms which allow for the inclusion of categorical data, such as k-modes clustering may be used (Huang, 1998). Recent studies suggest gender-specific differences in dementia syndromes on both clinical symptoms and extent of brain atrophy, so called sexual dimorphism. This was shown for PPA (Rogalski et al., 2007; Sebastian et al., 2018), bvFTD (Illán-Gala et al., 2021) and AD (Gamberger et al., 2016). Although results for bvFTD and PPA patients are not conclusive, and further studies are required to establish the relationship between gender and dementia phenotypes, these studies suggest that gender as an influencing factor should at least not be completely discarded. Future studies should thus consider this knowledge and include gender as additional factor for the analysis. Alternatively, future studies may perform the analysis for each gender separately. While investigating the effect of gender on dementia syndromes, this may additionally allow replication of results in two independent cohorts.

Criticism may be expressed concerning the dataset used. Large differences in the sample size of participant outcome groups may be problematic for two reasons. Firstly, it may have influenced clustering results observed (Fränti & Sieranoja, 2018). The computational aspects of k-means clustering which aim to minimize clustering error have the effect that spherical clusters with similar cluster sizes are favored. This is also referred to as the “uniform effect”

(Hui Xiong et al., 2009; Mittal et al., 2019; Zhou & Yang, 2020). In accordance with bvFTD forming the largest and lvPPA forming the smallest participant group, the observation that bvFTD and lvPPA clusters present as most and least homogeneous, respectively, could be an artifact of differences in group size. However, clustering results remain similar also when including only patients diagnosed with PPA. PPA groups are much more similar in size. To ensure that results are independent of group size, replication of the results may be performed by resampling the groups to balance them for group sizes. Second, the analysis may not reflect well the prevalence of diagnostic groups, thus questioning the validity of results for implementation in a clinical setting. While the data was acquired from clinics over many years as they were encountering diverse patients, the inconsistency of group size to population prevalence may particularly be true with regards to the two control groups in the dataset, namely AD which is much more prevalent than its proportion of the current study and healthy controls, which may not present in neurologic clinics.

Further criticism may concern the high amount of missingness in the current dataset and the way missing data was addressed in the analysis. Compared to single-center prospective studies, retrospective consortium studies use data from many participants assessed on many variables and often followed longitudinally. This is particularly valuable for rare disorders, such as PPA, for which prospective studies are usually restricted by a small sample size. ML algorithms are optimal for analyzing such highly dimensional data (*e.g.,* Parums, 2021). Multi-centric collaboration however comes at the cost of less clean and less complete data due to the involvement of a multitude of parties with varying interests and differences in available resources (*e.g.,* García-Laencina et al., 2010). To apply the k-means clustering algorithm, it is required to deal with the missing data prior to analysis. We decided to combine two approaches to deal with missing data. First, participants who had more than 20% missingness on the variables used for the analysis were excluded. This approach is referred to as list-wise deletion or complete-case analysis (Molenberghs et al., 2014). Missing data on the remaining participants was imputed using multiple imputation (Campion & Rubin, 1987). As there is not a single best way to treat missing data, this approach may be criticized. We argue however, that the chosen approach represents a reasonable trade-off to accommodate the limitations of both methods by themselves. List-wise deletion leads to a selection bias in cases where data is not missing completely at random or concerns only a very small percentage (Basagaña et al., 2013; Enders, 2010). Additionally, for the dataset used, list-wise deletion would imply that only specified variables could be used as virtually none of the participants had complete data. The advantage of applying unsupervised learning algorithms lies in its ability to be suitable for

highly dimensional data and not having to decide a priori about the relevance of specific variables. Imputing data for participants who have missing values on most variables is at the risk of considerable amount of additional noise. Information from other variables may not be sufficient to accurately impute the missing values, meaning that the assumption of ignorability may not be met (van Buuren, 2018). Multiple imputation is thought to be well suited to deal with missing data in the clinical setting (*e.g.,* Hayati Rezvan et al., 2015; Tran et al., 2017). One major advantage is that noise introduced by the process of imputation is reflected in the variation of results. Maruta et al. (2015), prior to applying diverse clustering algorithms to clinical data from patients diagnosed with PPA, used a similar two step approach to address missing data. Instead of applying list-wise deletion for participants with a large proportion of missingness, they removed variables exceeding 30% missingness and imputed remaining values using the mean or mode value. In our study, questionnaires showed the largest proportion of missingness. In favor of inclusion of questionnaire scores assessing behavioral change we decided against this method and chose list-wise deletion instead. Imputation using multiple imputation rather than single imputation poses less risk to produce spurious effects caused by biased parameter estimates or underestimated standard errors (Enders, 2010). Overall, we consider the missing data and imputation method chosen a limitation of the study but with respect to the goal of the study and the current standard adopted in research for dealing with missing data, we argue that the selected approach was reasonably adequate.

Finally, limitations referring to the clustering results found should be noted. Clusters in all comparisons emerged that remained mixed and were difficult to interpret. The segregation of patients diagnosed with AD and nfvPPA or lvPPA seemed particularly fuzzy. AD patients mixed with all other patient groups to a similar extent. NfvPPA and lvPPA patients grouped together but separately from the other groups. Prior studies may suggest certain diagnostic groups to be more difficult to separate based on neuropsychological data only. Similar cognitive symptoms are expected for AD and bvFTD (*e.g.,* Musa et al., 2020) as well as for nfvPPA and lvPPA patients (*e.g.,* Bürger et al., 2017; Tippett, 2020). In these cases, data from other modalities may be required to observe data-driven separation of patient groups. This prior knowledge does not explain however clusters that contained patients from all or most diagnostic groups. In these cases, inspection of cluster centers did not help interpreting the results. It may demonstrate the heterogeneity of cognitive symptoms observed for patients of the same diagnostic group. Additionally, it is conceivable that these clusters contain misdiagnoses. Pathological data existed only for a small minority of participants included in the study and thus diagnoses were not definite. Future studies may investigate these mixed clusters more closely

to understand whether additional, multimodal data could help the distinction of these patients and their respective diagnoses. Pathological data such as amyloid-marker as indicated by positive amyloid-PET may be useful to distinguish AD and its atypical variant lvPPA from the FTD spectrum disorders (Deuschl et al., 2016). LvPPA is a relatively new diagnostic category and not yet well understood. Some studies question the existence of lvPPA as separate diagnostic group (Maruta et al., 2015; Sajjadi et al., 2012). Pathological markers in lvPPA remain ambiguous and overlap in pathology underlying AD and FTD syndromes was shown (Elahi & Miller, 2017). Despite lvPPA belonging to the atypical AD variants, not all patients diagnosed with lvPPA show AD typical pathology (Matias-Guiu et al., 2019). Lastly, these findings may also be in accordance with a debate that dementia syndromes need to be conceptualized on a spectrum, with possible overlaps between them. Single, deterministic diagnoses may not be adequate to characterize dementia syndromes which instead may be mixed or evolve over time. This has been suggested by several studies for PPA (Gil-Navarro et al., 2013; Sajjadi et al., 2012; Wicklund et al., 2014) and AD (*e.g.,* Price et al., 2015). To conclude, interpretation of mixed clusters observed is ambiguous and may be caused by methodological limitations, noisy data or instead may reflect limitations of current diagnostic standards. This remains to be investigated in future studies.

4.3 Future Studies

Future studies are necessary to answer open questions, replicate the results and prove the clinical relevance of the current findings. Building on our results, we point to possible targets for future studies and make suggestions to improve the analysis performed. Hypotheses pointed out previously need to be tested in more rigorous studies. Instead of focusing on significant differences between patient groups, these studies should assess the number of participants that can be attributed to the correct diagnostic group using specific tests. This is common for supervised ML studies but also regression analyses. Patient-centered prediction is an important goal, particularly in the case of high within-group heterogeneity as was observed in our study. One interesting finding is the possible use of FTLD-CDR sub-scores to distinguish diagnostic groups. While the FTLD-CDR total scores are commonly used to assess or control for severity, we found indications that the FTLD-CDR sub-scores may be used to distinguish patients with PPA from the other groups. We performed a preliminary analysis by creating a difference score between the *Language* sub-score and the average of the remaining FTLD-CDR sub-scores. A simple logistic regression resulted in a specificity and sensitivity around 0.90 to distinguish PPA from non-PPA groups. The threshold for separation was -0.36. This may highlight the possible use of a hierarchical framework for diagnosis. In a first step, PPAs may

be separated from the remaining groups and in a second step using for example the repeat and point test, verbal and visuospatial memory assessments patients with svPPA may be distinguished from the two other PPA variants. Based on the current study it is difficult to differentiate lvPPA and nfvPPA and pathological or neuroimaging data may be necessary. Similarly, performing a logistic regression on a difference score between the *Behavior, comportment, & personality* sub-score and the *Language* sub-scores to distinguish bvFTD patients from all other groups yields a sensitivity score of 0.73 and a specificity score of 0.96 at a threshold of 0.25.

Generally, replications are necessary to prove robustness of results across different samples. Such studies may include cohorts from other cultures (Bachli et al., 2020), apply different clustering methods or use other neuropsychological and behavioral assessments. For example, in the current study tests examining executive functions did not seem to have a strong influence on clustering results. Error scores commonly showed floor effects and a combined error score may prove more useful (Kamath et al., 2019; Kramer et al., 2003). Social functioning was assessed in the current study using the RMET only. In contrast to previous studies (Gossink et al., 2018) for possible differentiation of bvFTD, the current study did not find impairment on social functioning to be specific for bvFTD. As for executive functioning, social functioning is a complex construct, and a single assessment may not reflect social functioning impairment. Future studies need to investigate the usefulness of other social function tasks (Dodich et al., 2021). Further, for the distinction of svPPA from the other two PPA variants, visuospatial memory tasks have been suggested to prove useful (D. Foxe, Irish, et al., 2021). Visuospatial memory function was assessed in our study using the visual memory span and a figure drawing task. Greater impairment in lvPPA/ nfvPPA compared to svPPA was partly supported but needs further investigation. One finding that could be of interest is that nfvPPA/ lvPPA seemed to exhibit worse performance on TMT and the Stroop task. This may be related to fluency and processing speed and requires further investigation. Additionally, impairments on the TMT and digit span task may be related to differences in number processing. Support for greater impairment of number processing in lvPPA/ nfvPPA compared to svPPA may come from a study by Hardy et al. (2018) assessing processing of degraded speech output of number or place words. Further, lvPPA specifically has been related to difficulties with calculation (Kamath et al., 2020; Rohrer et al., 2010). These need to be investigated in future studies. Additionally, exploration of differences between patient groups of the same cluster exceeded the scope of the current research. This may be a focus of future studies and provide fine-grained insight into the clinical phenotypes. Are there differences between diagnostic groups in joint clusters that the

algorithm did not pick-up on?

Further, robustness of results needs to be investigated by evaluating results with respect to correspondence to clustering results from other modalities. Is there a relationship between clustering results found from behavioral and cognitive data and clustering results from neuroimaging or pathological data? Coherence between different modalities would strengthen the findings and the use of neuropsychological data as a cheaper and less invasive method.

Once selected tests have repeatedly shown usefulness for distinction of different patient groups, clinical applicability needs to be tested. To facilitate applicability, the number of variables should be reduced to a minimum. Additionally, the emergence of longitudinal data, within the FTLD consortium or other patient cohorts may provide validation for the findings and their relevance for diagnosis. Findings from longitudinal data support progression of behavioral symptoms to differ in bvFTD and svPPA (O'Connor et al., 2016). Additionally, differences in temporal progression of cognitive impairment was found for AD and FTD (Libon et al., 2009; Xie et al., 2010). The current study focuses only on patients first assessment. Possible correspondence of greater disease severity as observed in our study with temporal progression of the syndromes remains to be shown.

4.3.1 Heterogeneity and severity

One major finding was that a large proportion of clustering results may be explained by differences in overall disease severity. This was particularly true for low numbers of clusters (*i.e.,* k = 2 or k = 3) or when more than one cluster contained a majority of participants from the same diagnostic group. This information needs to be integrated into future studies.

Commonly, CDR, MMSE or both are used to assess clinical or cognitive disease severity (Cerami et al., 2016; Rogalski et al., 2007; Themistocleous et al., 2021). For the use in patient cohorts belonging to the FTD spectrum, the FTLD-CDR scale was developed by adding two sub-scores evaluating language and behavioral impairment to the CDR scale. Based on the findings of the current study, bvFTD patients seem to score relatively highly on the MMSE while patients with PPA or AD score comparatively high on the (FTLD-) CDR score. This corresponds to the conceptualization of bvFTD patients showing greater behavioral than cognitive changes while the PPAs and AD are associated mainly with cognitive symptoms. The MMSE was also suspected to overestimate severity in PPA patients due to reliance on language function (Henry & Grasso, 2018; Osher et al., 2008). MMSE and CDR scale do not seem to have a high agreement (Juva et al., 2009).

As severity may be a driving factor for the large heterogeneity between patients of the same diagnostic groups, it is necessary that future studies include measures of severity,

optimally of both cognitive and behavioral symptoms. Additional to the use of the MMSE and the CDR score, a global z-score may be indicative of severity (Machulda et al., 2013). Previous studies have shown that demographic differences such as duration since onset of the symptoms is not a good proxy for disease severity (Fan et al., 2020; Kamath et al., 2020; Machulda et al., 2013). In contrast to the current study, we advise that severity scores get incorporated in the statistical analysis. Accounting for differences in severity within the analysis would have two major advantages. Firstly, it may reduce heterogeneity within patient groups thereby facilitating differential diagnosis. For example, as our results suggest, the hallmark of AD being memory impairment may not be particularly apparent when not first controlling for disease severity. Secondly, it may enhance current understanding of the syndromes across different severities. The simplest way to incorporate severity in the analysis would be by stratification. A more complex model was proposed by Young et al. (2018), called by the authors *Subtype and Stage Inference* (SuStaIn) using clustering and progression modelling of neuroimaging data from mutation-carriers related to AD and FTD. Additionally, the inclusion of longitudinal data could provide information about the existence of possible subtypes, or whether instead greater severity corresponds to a later stage of the same subtype. Some studies have suggested the existence of different cognitive profiles (O'Connor et al., 2017; Ramanan et al., 2020) or patterns of severity (Machulda et al., 2013; Ziegler et al., 2020) within the same disorder. This could not be supported by the current study but requires further investigation.

4.3.1.1 Preliminary Results of a Clustering Analysis Stratifying Participants Based on Disease Severity

We propose a new data-driven approach that may result in patient stratification by severity to allow for more detailed exploration of patient phenotypes. Stratification based on specific cut-offs on the MMSE, FTLD-CDR or a global z-score may be rather arbitrary. Instead, we propose a data-driven two-step clustering approach. We made use of the observation drawn from the current study that the major factor influencing clustering results may have been disease severity. In a first step we used the k-means clustering algorithm to group participants by disease severity. In a second step the resulting clusters were further analyzed repeating k-means clustering. In this way participants from various diagnostic groups may be compared across similar disease severities. This approach may be relevant in case the sample to be analyzed does show large variations in disease severity. Further, it is based on the hypothesis that k-means clustering results will be highly influenced by differences in severity if those exist. In these cases, the approach may adapt flexibly to the variables of interest and does not require an explicit measure to approximate disease severity. This method does not aim to be a clinically

relevant measure of severity. Instead, it aims to make exploration of participant subgroups more straightforward by being focused on participants with similar severities. We performed a preliminary analysis using the proposed method.

Results were similar to the previous analysis: No clear AD clusters emerged and lvPPA and nfvPPA patients did not seem to separate in coherent ways. This further analysis proved relevant however for the segregation of svPPA and the two other PPA variants as well as for the segregation of PPA and bvFTD from other patient subgroups. Patterns that were previously observed could be replicated and observed in a more consistent and understandable way. Increasing the number of clusters to more than six in this second run analysis did not seem to add informative value.

Concerning the separation of bvFTD from the remaining groups, relevance of behavioral questionnaires and FTLD-CDR sub-scores could be replicated. The possible contrast between a relatively high *Language* and relatively low *Behavior, comportment, & personality* sub-score may be specific to bvFTD. One observation that was not made previously is the possible usefulness of the H5PT. BvFTD showed a tendency for a smaller percentage of correct figures drawn, than the other patient groups. Considering that bvFTD patients did not seem to have lower total scores on the H5PT suggests that participants with bvFTD had problems keeping track of the figures already drawn and made repetition errors. This indicates an executive impairment reflected also in a heightened error on the TMT. As suggested previously, possible executive impairment of patients with bvFTD may be best characterized using an overall error score (Kamath et al., 2019; Kramer et al., 2003).

Only clusters of patients with PPA showed bad performance on language and verbal memory despite preserved FTLD-CDR scores. In most cases this was mirrored by a low *Language* sub-score on the FTLD-CDR compared to the other sub-scores. Verbal memory but not figural memory was impaired both while learning and during recall.

When compared to all other patient groups AD clusters seemed to show greater impairment on measures of delayed recall than on immediate learning of a verbal task. This was not however observed when only AD and bvFTD were compared and it is in contrast to literature describing hippocampal-dependent memory deficits in AD to be characterized by both encoding and retrieval problems (Ahmed et al., 2021; Hutchinson & Mathias, 2007). Memory impairments of AD were reflected by low scores on both verbal and figural tasks. Language specific assessments seemed relatively preserved in AD. Particularly the Cookie theft task may be useful to differentiate AD from PPAs.

To differentiate svPPA from nfvPPA and lvPPA language assessments may not be the

most valuable. Instead, a consistent pattern showed lower scores of nfvPPA/ lvPPA clusters on assessments involving processing speed such as the TMT and the Stroop task. The difference on the TMT may be particularly pronounced for the version B. Additionally, differences were observed on both tasks of the WMS-R, the digit span and the visual memory span. Further, greater behavioral changes in svPPA compared to nfvPPA/ lvPPA clusters were suggested by results from the questionnaires. This was not observed in the previous analysis but is consistent with literature suggesting behavioral changes in svPPA to be similar to seen inpatients with bvFTD (Grossman, 2010; Harris et al., 2019; Ruksenaite et al., 2021) As previously mentioned, a clear distinction may be made between svPPA and nfvPPA/ lvPPA clusters based on the repeat and point task. While patients with svPPA seemed to reach considerably fewer points on the point than on the repeat task, patients in nfvPPA/ lvPPA showed similar impairment across both parts or in some cases showed the opposite pattern of greater impairment on the repeat than on the point part. Further tendencies observed, although less consistent, were greater deficits on the verbal memory assessments, the BNT, and the RMET in svPPA. This pattern may be explained by loss of semantics. Rather than reflecting deficits in social functioning, we hypothesize that deficits observed on the RMET may be explained by difficulties understanding the meaning of emotion words. Further, compared to nfvPPA/ lvPPA clusters greater impairments on the percent of correct figures of the H5PT was partly observed in svPPA. This may be in accordance with previous studies suggesting executive impairment in participants with svPPA (Kamath et al., 2020). However, on other measures of executive dysfunction such as the TMT version B or the Stroop, nfvPPA/ lvPPA clusters showed greater impairment.

Additionally, the analyses all together may suggest relevance to the FTLD-CDR sub-scores. In fact, despite rather broad and unspecific they seem to reflect well the overall pattern of domains impaired such as *Language* being highly impaired in the PPAs, *Memory* partly more impaired in AD and behavioral changes in bvFTD reflected by low *Behavior, comportment & personality*. As a measure of severity the FTLD-CDR score may be useful within but not necessarily between patient groups as patients with bvFTD may score comparatively low.

Results from this cluster analysis make clear that impairments on single variables are not indicative of a specific disorder. Instead, the goal should be to establish specific within participant patterns or contrasts to distinguish the disorders. In accordance with current diagnostic procedures a hierarchical framework may first separate PPA from non-PPA syndromes depending on the presence of language impairments before in a next step distinguishing bvFTD and AD or instead svPPA from nfvPPA/ lvPPA. The current study does not allow for conclusions concerning the distinction of nfvPPA and lvPPA. Further, it needs to

be investigated how different disease severities observed in the current study relate to progression of the studied syndromes. Specific impairments may be more straightforward to interpret in the case of preserved FTLD-CDR scores. In cases of increased disease severity instead it may be difficult to discern specific patterns of heightened impairments in some compared to other domains.

4.3.2 ML Applications for Differential Diagnosis

Chekroud et al. (2021) summarizes the main advantages of ML applications compared to classical statistical methods as being threefold. First, prediction is performed on the level of the individual. Second, ML methods can easily integrate the information from a large number of variables as well as their combined effects, even when effects of single variables contribute only marginally. Third, ML methods may find more complex patterns in the data, that are not linear. For implementation in the clinical setting, it is necessary for tools developed to maximize explanatory power while at the same time remaining comparatively simple. Studies have shown that simpler algorithms increase patient and clinician trust due to greater transparency (Grote & Berens, 2020; Holzinger et al., 2022; Vayena et al., 2018). Additionally, while research may investigate characteristics of specific patient groups using highly sophisticated methods such as resting-state or functional MRI, this is not common in clinical settings. Instead, access to neuroimaging is relatively limited and expensive. While it should be stressed that access to neuroimaging methods is important for accurate diagnosis we here also want to stress that the possible usefulness of neuropsychological data has not yet been exploited to its fullest.

The use of ML techniques may come to full potential with the development of continuously larger consortium studies. Consortium studies have the advantage of exploring patient's characteristics on a breadth of variables and to include many patients from different centers. A recent study demonstrated the need for relatively large sample sizes especially of test data for supervised ML techniques. With small samples a high risk of overestimating classification performance was noted (Flint et al., 2021). Additional to the emergence of consortium studies, the sharing of open-source models could facilitate development and application of ML tools for implementation in the clinics. Open-source models could spur quality and applicability by being tested on independent datasets without the need to share the sensible patient data (Flint et al., 2021). Lastly, "ensembling" techniques may allow to yield more robust results by combining predictions from a variety of analyses on the same data (Chekroud et al., 2021). In our study, ensembling was used to handle the missing data. More commonly, ensemble techniques are used to combine results from different algorithms or using different hyperparameter specifications.

The use of unsupervised ML techniques in previous studies of FTD spectrum disorders has focused on two main goals: diagnostic segregation and subgroup discovery. Diagnostic segregation commonly focused on few groups and included either aphasic or bvFTD patients but not both together. Studies aiming at subgroup discovery instead investigate single diagnostic groups to find subgroups with different cognitive phenotypes (O'Connor et al., 2017; Owens et al., 2018) or atrophy patterns (Bruun et al., 2019; Whitwell et al., 2009). Other applications of unsupervised clustering in the clinical setting are to aid preprocessing or compression of imaging data. A classifier may then be trained on the clustered rather than the raw data (Li & Liu, 2018; Sampath & Saradha, 2014). Our study may indicate that the use of k-means clustering could be broader than previously expected. It seems suitable for exploration of data including several diagnostic groups. In this study, six different groups, including healthy controls participants, were included. We observed relatively stable results with increasing number of clusters, across different comparisons of patient groups and despite high levels of missing data. Specifically, many results observed when comparing few diagnostic groups were already seen when more groups were included in the analysis.

The use of ML is of particular interest in cases where uncertainty persists. From this perspective differential diagnosis, as was investigated in the current study may be the most pressing issue to be explored using ML techniques. In contrast, distinction from healthy controls is less uncertain and may provide little additional information. Further, the investigation for subgroups within existing diagnostic groups may be of little clinical use to date. However, this type of research may prove relevant in the future by yielding new information about the observed heterogeneity within diagnostic groups. One question that persists is whether clear split diagnostic groups exist in the FTD and AD spectrum. Existence of overlap may however not necessarily be problematic for the application of k-means clustering. In fact, a study by Fränti & Sieranoja (2018) suggests that overlap between groups in the data may allow for better clustering results. In case of overlap the k-means clustering algorithm may be less likely to stabilize on a local minimum and instead more flexibly alter between clustering solutions in the search of a global minimum. Even in the most extreme case where no regions of increased density or separation between groups exists but instead phenotypes from dementia syndromes exist on a spectrum, k-means clustering may allow to simplify the data for exploration. In this case grouping may not claim added-value for diagnosis. However, by clustering participants that lie nearby on the dimensions assessed, interpretation of highly dimensional data may be facilitated.

4.4 Further Considerations

One key aspect of clustering used for data exploration is the visualization of results. Interpretation of results is performed by a human observer and thus the information gained from a clustering analysis may be highly dependent on appropriateness and novelty of the visualization used (von Luxburg et al., 2012). Additionally, some authors have suggested that clustering can only be evaluated with respect to its goal and that no "natural" kinds (von Luxburg et al., 2012) or "true" clusters (Hennig, 2015) to be uncovered exist. In this way, the various existing internal or external validation methods for clustering may all be insufficient and validation of resulting clusters depends on the specific goal. Combined, these two aspects indicate that clustering despite having the potential of being independent of labels, will be biased by the goal pursued and the expectations of the researcher. In our study, the goal of the clustering algorithm was differential diagnosis between FTD and AD spectrum syndromes. Thus, visualization was highly focused on the groups that existed in the data. Validation of the clustering results instead needs to prove by usefulness in future diagnosis or specification for treatment options.

At this point one may question the utility of the diagnostic procedure all together. Philosophically, categories are a human construct and no "true" grouping exists (Hennig, 2015). Categorization may still be useful to navigate the world by simplifying it in meaningful ways. Diagnostic utility may be particularly high in case it carries information about treatment advice. In the case of FTD where no clear conclusions about treatment follow a diagnosis, utility may lie in homogenizing patient groups to research possible treatments (Elahi & Miller, 2017). As such diagnostic precision may always present a trade-off between simplifying the clinical picture of a patient while taking into account as many details to be maximally meaningful for treatment. This makes multimodality of the diagnostic process including pathology, atrophy and behavioral changes important. Rather than a description of reality, diagnoses may be seen as aiming at the most useful simplification of reality. It is thus important for researchers investigating patient cohorts and for clinicians working on the level of the individual patients, to regularly put diagnoses into question. It is important to recognize the large heterogeneity within diagnostic groups and the overlap between diagnostic groups. Diagnosis may be malleable and should be adapted when more evidence is acquired. Similarly, changes in available treatments may also affect the usefulness of patient segregation. For example, it may be useful to distinguish patients diagnosed with bvFTD that could profit from a treatment with Donepezil, a drug commonly used for AD patients (Deuschl et al., 2016), and those in whom Donepezil may lead to worsening of symptoms. A distinction of AD and bvFTD may in that case be less meaningful. This is particularly true for FTD where the pathological process

influencing the emergence of the FTD phenotypes is not well understood. Another aspect that should not be underestimated is the effect a diagnosis can have for patients and caregivers. On the one hand, they may resolve patients' and caregivers' questions and uncertainties, thereby improving quality of life (Musa et al., 2020; Weder et al., 2007). On the other hand, diagnoses, particularly if inaccurate, carry potential for social stigma (Sachdev et al., 2014; van Vliet et al., 2013).

5 Conclusions

To summarize, using an unsupervised clustering approach we found homogeneous grouping of patients with svPPA and bvFTD while grouping of the remaining patient groups included was more mixed. LvPPA and nfvPPA did not seem separable based on the variables included in the study. Overlap in neuropsychological and behavioral impairment between diagnostic groups was large. High levels of heterogeneity within diagnostic groups were found and partly related to differences in disease severity. A new data-driven method to stratify patients based on severity prior to further analysis was proposed and its usefulness may be probed in the future. Further, specific hypotheses to be tested in future studies were posited. Companion-rated questionnaires assessing apathy, executive dysfunction and disinhibition seemed particularly useful in the distinction of bvFTD. Semantic impairment in svPPA was shown by low scores on both language and memory tests. NfvPPA and lvPPA patients instead showed impairment in processing speed and short term and working memory assessments. We maintain the relevance of neuropsychological and behavioral variables in the distinction of the groups included in the current study. It is important that future studies keep a patient-centered approach. By being independent from data labelling, k-means clustering may prove particularly useful to analyze data from highly heterogeneous or unclassifiable patients.

References

Abdi, H., Williams, L. J., Beaton, D., Posamentier, M. T., Harris, T. S., Krishnan, A., & Devous, M. D. S. (2012). Analysis of regional cerebral blood flow data to discriminate among Alzheimer's disease, frontotemporal dementia, and elderly controls: a multi-block barycentric discriminant analysis (MUBADA) methodology. *Journal of Alzheimer's Disease : JAD*, *31 Suppl 3*(0 3), S189-201. https://doi.org/10.3233/JAD-2012-112111

Ahmad, M. A., Eckert, C., & Teredesai, A. (2018). Interpretable Machine Learning in Healthcare. *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, *21*, 559–560. https://doi.org/10.1145/3233547.3233667

Ahmed, R. M., Hodges, J. R., & Piguet, O. (2021). *Behavioural Variant Frontotemporal Dementia: Recent Advances in the Diagnosis and Understanding of the Disorder* (pp. 1–15). https://doi.org/10.1007/978-3-030-51140-1_1

Alashwal, H., El Halaby, M., Crouse, J. J., Abdalla, A., & Moustafa, A. A. (2019). The Application of Unsupervised Clustering Methods to Alzheimer's Disease. *Frontiers in Computational Neuroscience*, *13*(May), 1–9. https://doi.org/10.3389/fncom.2019.00031

Alexander, N., Alexander, D. C., Barkhof, F., & Denaxas, S. (2021). Identifying and evaluating clinical subtypes of Alzheimer's disease in care electronic health records using unsupervised machine learning. *BMC Medical Informatics and Decision Making*, *21*(1), 343. https://doi.org/10.1186/s12911-021-01693-6

Álvarez, J. D., Matias-Guiu, J. A., Cabrera-Martín, M. N., Risco-Martín, J. L., & Ayala, J. L. (2019). An application of machine learning with feature selection to improve diagnosis and classification of neurodegenerative disorders. *BMC Bioinformatics*, *20*(1), 491. https://doi.org/10.1186/s12859-019-3027-7

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Arevalo-Rodriguez, I., Smailagic, N., Roqué-Figuls, M., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., Pedraza, O. L., Bonfill Cosp, X., & Cullum, S. (2021). Mini-Mental State Examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews*, *2021*(7), CD010783. https://doi.org/10.1002/14651858.CD010783.pub3

Ash, S., Jester, C., York, C., Kofman, O. L., Langey, R., Halpin, A., Firn, K., Dominguez Perez, S., Chahine, L., Spindler, M., Dahodwala, N., Irwin, D. J., McMillan, C., Weintraub, D., & Grossman, M. (2017). Longitudinal decline in speech production in Parkinson's disease

spectrum disorders. *Brain and Language*, *171*, 42–51. https://doi.org/10.1016/j.bandl.2017.05.001

Atkinson, R. C., & Shiffrin, R. M. (1971). The Control of Short-Term Memory. *Scientific American*, *225*(2), 82–91.

Austin, P. C., White, I. R., Lee, D. S., & van Buuren, S. (2021). Missing Data in Clinical Research: A Tutorial on Multiple Imputation. *Canadian Journal of Cardiology*, *37*(9), 1322–1331. https://doi.org/10.1016/j.cjca.2020.11.010

Baborie, A., Griffiths, T. D., Jaros, E., Momeni, P., McKeith, I. G., Burn, D. J., Keir, G., Larner, A. J., Mann, D. M., & Perry, R. (2012). Frontotemporal Dementia in Elderly Individuals. *Archives of Neurology*, *69*(8), 1052. https://doi.org/10.1001/archneurol.2011.3323

Bachli, M. B., Sedeño, L., Ochab, J. K., Piguet, O., Kumfor, F., Reyes, P., Torralva, T., Roca, M., Cardona, J. F., Campo, C. G., Herrera, E., Slachevsky, A., Matallana, D., Manes, F., García, A. M., Ibáñez, A., & Chialvo, D. R. (2020). Evaluating the reliability of neurocognitive biomarkers of neurodegenerative diseases across countries: A machine learning approach. *NeuroImage*, *208*(October 2019), 116456. https://doi.org/10.1016/j.neuroimage.2019.116456

Baez, S., Pinasco, C., Roca, M., Ferrari, J., Couto, B., García-Cordero, I., Ibañez, A., Cruz, F., Reyes, P., Matallana, D., Manes, F., Cetcovich, M., & Torralva, T. (2019). Brain structural correlates of executive and social cognition profiles in behavioral variant frontotemporal dementia and elderly bipolar disorder. *Neuropsychologia*, *126*, 159–169. https://doi.org/10.1016/j.neuropsychologia.2017.02.012

Bang, J., Spina, S., & Miller, B. L. (2015). Non-Alzheimer's dementia 1 Frontotemporal dementia. *The Lancet*, *386*(10004), 1672–1682. http://dx.doi.org/10.1016/S0140-6736(15)00461-4

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry*, *42*(2), S0021963001006643. https://doi.org/10.1017/S0021963001006643

Basagaña, X., Barrera-Gómez, J., Benet, M., Antó, J. M., & Garcia-Aymerich, J. (2013). A Framework for Multiple Imputation in Cluster Analysis. *American Journal of Epidemiology*, *177*(7), 718–725. https://doi.org/10.1093/aje/kws289

Beber, B. C., & Chaves, M. L. F. (2013). Evaluation of patients with behavioral and cognitive complaints: Misdiagnosis in frontotemporal dementia and Alzheimer's disease. *Dementia & Neuropsychologia*, *7*(1), 60–65. https://doi.org/10.1590/S1980-57642013DN70100010

Beeldman, E., Raaphorst, J., Klein Twennaar, M., Govaarts, R., Pijnenburg, Y. A. L., de Haan, R. J., de Visser, M., & Schmand, B. A. (2018). The cognitive profile of behavioural variant FTD and its similarities with ALS: a systematic review and meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, *89*(9), 995–1002. https://doi.org/10.1136/jnnp-2017-317459

Benussi, A., Grassi, M., Palluzzi, F., Koch, G., Di Lazzaro, V., Nardone, R., Cantoni, V., Dell'Era, V., Premi, E., Martorana, A., di Lorenzo, F., Bonnì, S., Ranieri, F., Capone, F., Musumeci, G., Cotelli, M. S., Padovani, A., & Borroni, B. (2020). Classification Accuracy of Transcranial Magnetic Stimulation for the Diagnosis of Neurodegenerative Dementias. *Annals of Neurology*, *87*(3), 394–404. https://doi.org/10.1002/ana.25677

Bertoux, M., Delavest, M., de Souza, L. C., Funkiewiez, A., Lépine, J.-P., Fossati, P., Dubois, B., & Sarazin, M. (2012). Social Cognition and Emotional Assessment differentiates frontotemporal dementia from depression. *Journal of Neurology, Neurosurgery & Psychiatry*, *83*(4), 411–416. https://doi.org/10.1136/jnnp-2011-301849

Bhardwaj, R., Nambiar, A. R., & Dutta, D. (2017). A Study of Machine Learning in Healthcare. *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, *2*, 236–241. https://doi.org/10.1109/COMPSAC.2017.164

Bonanni, L., Franciotti, R., Martinotti, G., Vellante, F., Flacco, M. E., Di Giannantonio, M., Thomas, A., & Onofrj, M. (2018). Post Traumatic Stress Disorder Heralding the Onset of Semantic Frontotemporal Dementia. *Journal of Alzheimer's Disease*, *63*(1), 203–215. https://doi.org/10.3233/JAD-171134

Bouts, M. J. R. J. R. J., Möller, C., Hafkemeijer, A., van Swieten, J. C., Dopper, E., van der Flier, W. M., Vrenken, H., Wink, A. M., Pijnenburg, Y. A. L. L., Scheltens, P., Barkhof, F., Schouten, T. M., de Vos, F., Feis, R. A., van der Grond, J., de Rooij, M., & Rombouts, S. A. R. B. R. B. (2018). Single Subject Classification of Alzheimer's Disease and Behavioral Variant Frontotemporal Dementia Using Anatomical, Diffusion Tensor, and Resting-State Functional Magnetic Resonance Imaging. *Journal of Alzheimer's Disease*, *62*(4), 1827–1839. https://doi.org/10.3233/JAD-170893

Bridel, C., van Wieringen, W. N., Zetterberg, H., Tijms, B. M., Teunissen, C. E., Alvarez-Cermeño, J. C., Andreasson, U., Axelsson, M., Bäckström, D. C., Bartos, A., Bjerke, M., Blennow, K., Boxer, A., Brundin, L., Burman, J., Christensen, T., Fialová, L., Forsgren, L., Frederiksen, J. L., … Wild, E. J. (2019). Diagnostic Value of Cerebrospinal Fluid Neurofilament Light Protein in Neurology: A Systematic Review and Meta-analysis. *JAMA Neurology*, *76*(9), 1035–1048. https://doi.org/10.1001/jamaneurol.2019.1534

Brigadoi, S., Cutini, S., Meconi, F., Castellaro, M., Sessa, P., Marangon, M., Bertoldo, A., Jolicœur, P., & Dell'Acqua, R. (2017). On the Role of the Inferior Intraparietal Sulcus in Visual Working Memory for Lateralized Single-feature Objects. *Journal of Cognitive Neuroscience*, *29*(2), 337–351. https://doi.org/10.1162/jocn_a_01042

Bron, E. E., Smits, M., Papma, J. M., Steketee, R. M. E., Meijboom, R., de Groot, M., van Swieten, J. C., Niessen, W. J., & Klein, S. (2017). Multiparametric computer-aided differential diagnosis of Alzheimer's disease and frontotemporal dementia using structural and advanced MRI. *European Radiology*, *27*(8), 3372–3382. https://doi.org/10.1007/s00330-016-4691-x

Bron, E. E., Steketee, R. M. E., Houston, G. C., Oliver, R. A., Achterberg, H. C., Loog, M., van Swieten, J. C., Hammers, A., Niessen, W. J., Smits, M., & Klein, S. (2014). Diagnostic classification of arterial spin labeling and structural MRI in presenile early stage dementia. *Human Brain Mapping*, *35*(9), 4916–4931. https://doi.org/10.1002/hbm.22522

Bruun, M., Koikkalainen, J., Rhodius-Meester, H. F. M., Baroni, M., Gjerum, L., van Gils, M., Soininen, H., Remes, A. M., Hartikainen, P., Waldemar, G., Mecocci, P., Barkhof, F., Pijnenburg, Y., van der Flier, W. M., Hasselbalch, S. G., Lötjönen, J., & Frederiksen, K. S. (2019). Detecting frontotemporal dementia syndromes using MRI biomarkers. *NeuroImage: Clinical*, *22*(November 2018), 101711. https://doi.org/10.1016/j.nicl.2019.101711

Bruun, M., Rhodius-Meester, H. F. M., Koikkalainen, J., Baroni, M., Gjerum, L., Lemstra, A. W., Barkhof, F., Remes, A. M., Urhemaa, T., Tolonen, A., Rueckert, D., Gils, M., Frederiksen, K. S., Waldemar, G., Scheltens, P., Mecocci, P., Soininen, H., Lötjönen, J., Hasselbalch, S. G., & Flier, W. M. (2018). Evaluating combinations of diagnostic tests to discriminate different dementia types. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*(1), 509–518. https://doi.org/10.1016/j.dadm.2018.07.003

Buhl, C., Stokholm, J., & Gade, A. (2013). Clinical Utility of Short Social Cognitive Tests in Early Differentiation of Behavioral Variant Frontotemporal Dementia from Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders Extra*, *3*(1), 376–385. https://doi.org/10.1159/000355123

Bürger, K., Arzberger, T., Stephan, J., Levin, J., & Edbauer, D. (2017). Pathomechanismen und klinische Aspekte der frontotemporalen Lobärdegeneration. *Der Nervenarzt*, *88*(2), 163–172. https://doi.org/10.1007/s00115-016-0259-x

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). mice : Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67.

https://doi.org/10.18637/jss.v045.i03

Caixeta, L., & Caixeta, M. (2011). Primary progressive aphasia beginning with a psychiatric disorder. *Clinics*, *66*(8), 1505–1508. https://doi.org/10.1590/S1807-59322011000800035

Callahan, A., & Shah, N. H. (2017). Machine Learning in Healthcare. In *Key Advances in Clinical Informatics* (pp. 279–291). Elsevier. https://doi.org/10.1016/B978-0-12-809523-2.00019-4

Campion, W. M., & Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. In D. B. Rubin (Ed.), *Journal of Marketing Research* (Vol. 26, Issue 4). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470316696

Cerami, C., Dodich, A., Lettieri, G., Iannaccone, S., Magnani, G., Marcone, A., Gianolli, L., Cappa, S. F., & Perani, D. (2016). Different FDG-PET metabolic patterns at single-subject level in the behavioral variant of fronto-temporal dementia. *Cortex*, *83*, 101–112. https://doi.org/10.1016/j.cortex.2016.07.008

Chagué, P., Marro, B., Fadili, S., Houot, M., Morin, A., Samper-González, J., Beunon, P., Arrivé, L., Dormont, D., Dubois, B., Teichmann, M., Epelbaum, S., & Colliot, O. (2020). Radiological classification of dementia from anatomical MRI assisted by machine learning-derived maps. *Journal of Neuroradiology = Journal de Neuroradiologie*. https://doi.org/10.1016/j.neurad.2020.04.004

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., & Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, *20*(2), 154–170. https://doi.org/10.1002/wps.20882

Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2021). Ethical Machine Learning in Healthcare. *Annual Review of Biomedical Data Science*, *4*(1), 123–144. https://doi.org/10.1146/annurev-biodatasci-092820-114757

Chen, P.-H. C., Liu, Y., & Peng, L. (2019). How to develop machine learning models for healthcare. *Nature Materials*, *18*(5), 410–414. https://doi.org/10.1038/s41563-019-0345-0

Chiu, D., & Talhouk, A. (2021). *diceR: Diverse Cluster Ensemble in R* (1.1.0). https://cran.r-project.org/package=diceR

Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*, *110*, 12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004

Cohen, J. P., Cao, T., Viviano, J. D., Huang, C.-W., Fralick, M., Ghassemi, M., Mamdani, M., Greiner, R., & Bengio, Y. (2021). Problems in the deployment of machine-learned models in health care. *Canadian Medical Association Journal*, *193*(35), E1391–E1394. https://doi.org/10.1503/cmaj.202066

Collins, J. D., Henley, S. M. D., & Suárez-González, A. (2020). A systematic review of the prevalence of depression, anxiety, and apathy in frontotemporal dementia, atypical and young-onset Alzheimer's disease, and inherited dementia. *International Psychogeriatrics*, 1–20. https://doi.org/10.1017/S1041610220001118

Constantinidis, J., Richard, J., & Tissot, R. (1974). Pick's Disease. *European Neurology*, *11*(4), 208–217. https://doi.org/10.1159/000114320

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

Cope, T. E., Wilson, B., Robson, H., Drinkall, R., Dean, L., Grube, M., Jones, P. S., Patterson, K., Griffiths, T. D., Rowe, J. B., & Petkov, C. I. (2017). Artificial grammar learning in vascular and progressive non-fluent aphasias. *Neuropsychologia*, *104*, 201–213. https://doi.org/10.1016/j.neuropsychologia.2017.08.022

Coyle-Gilchrist, I. T. S., Dick, K. M., Patterson, K., Vázquez Rodríquez, P., Wehmann, E., Wilcox, A., Lansdall, C. J., Dawson, K. E., Wiggins, J., Mead, S., Brayne, C., & Rowe, J. B. (2016). Prevalence, characteristics, and survival of frontotemporal lobar degeneration syndromes. *Neurology*, *86*(18), 1736–1743. https://doi.org/10.1212/WNL.0000000000002638

Dave, A., Hansen, N., Downey, R., & Johnson, C. (2020). FDG-PET Imaging of Dementia and Neurodegenerative Disease. *Seminars in Ultrasound, CT and MRI*, *41*(6), 562–571. https://doi.org/10.1053/j.sult.2020.08.010

de Bruijne, M. (2016). Machine learning approaches in medical image analysis: From detection to diagnosis. *Medical Image Analysis*, *33*(October), 94–97. https://doi.org/10.1016/j.media.2016.06.032

Deleon, J., & Miller, B. L. (2018). Frontotemporal dementia. In *Handbook of Clinical Neurology* (1st ed., Vol. 148, pp. 409–430). Elsevier B.V. https://doi.org/10.1016/B978-0-444-64076-5.00027-2

DeLozier, S. J., & Davalos, D. (2016). A Systematic Review of Metacognitive Differences Between Alzheimer's Disease and Frontotemporal Dementia. *American Journal of Alzheimer's Disease and Other Dementias*, *31*(5), 381–388.

https://doi.org/10.1177/1533317515618899

Dementia. (2021). https://www.who.int/news-room/fact-sheets/detail/dementia

Denk, J., Oberhauser, F., Kornhuber, J., Wiltfang, J., Fassbender, K., Schroeter, M. L., Volk, A. E., Diehl-Schmid, J., Prudlo, J., Danek, A., Landwehrmeyer, B., Lauer, M., Otto, M., & Jahn, H. (2018). Specific serum and CSF microRNA profiles distinguish sporadic behavioural variant of frontotemporal dementia compared with Alzheimer patients and cognitively healthy controls. *PLOS ONE*, *13*(5), e0197329. https://doi.org/10.1371/journal.pone.0197329

Deo, R. C. (2015). Machine Learning in Medicine. *Circulation*, *132*(20), 1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593

Desmarais, P., Weidman, D., Wassef, A., Bruneau, M.-A. A., Friedland, J., Bajsarowicz, P., Thibodeau, M.-P. P., Herrmann, N., & Nguyen, Q. D. (2020). The Interplay Between Post-traumatic Stress Disorder and Dementia: A Systematic Review. *American Journal of Geriatric Psychiatry*, *28*(1), 48–60. https://doi.org/10.1016/j.jagp.2019.08.006

Deuschl, G., Maier, W., Jessen, F., & Spottke, A. (2016). S3-Leitlinie Demenzen. *Deutsche Gesellschaft Für Neurologie, Hrsg. Leitlinien Für Diagnostik Und Therapie in Der Neurologie*. www.dgn.org/leitlinien

Dev, S. I., Dickerson, B. C., & Touroutoglou, A. (2021). *Neuroimaging in Frontotemporal Lobar Degeneration: Research and Clinical Utility* (pp. 93–112). https://doi.org/10.1007/978-3-030-51140-1_7

Devineni, B., & Onyike, C. U. (2015). Young-Onset Dementia Epidemiology Applied to Neuropsychiatry Practice. *Psychiatric Clinics of North America*, *38*(2), 233–248. https://doi.org/10.1016/j.psc.2015.02.003

Diehl, J., Monsch, A. U., Aebi, C., Wagenpfeil, S., Krapp, S., Grimmer, T., Seeley, W., Förstl, H., & Kurz, A. (2005). Frontotemporal Dementia, Semantic Dementia, and Alzheimer's Disease: The Contribution of Standard Neuropsychological Tests to Differential Diagnosis. *Journal of Geriatric Psychiatry and Neurology*, *18*(1), 39–44. https://doi.org/10.1177/0891988704272309

Dodich, A., Crespi, C., Santi, G. C., Cappa, S. F., & Cerami, C. (2021). Evaluation of Discriminative Detection Abilities of Social Cognition Measures for the Diagnosis of the Behavioral Variant of Frontotemporal Dementia: a Systematic Review. *Neuropsychology Review*, *31*(2), 251–266. https://doi.org/10.1007/s11065-020-09457-1

Donnelly-Kehoe, P. A., Pascariello, G. O., García, A. M., Hodges, J. R., Miller, B., Rosen, H., Manes, F., Landin-Romero, R., Matallana, D., Serrano, C., Herrera, E., Reyes, P.,

Santamaria-Garcia, H., Kumfor, F., Piguet, O., Ibanez, A., & Sedeño, L. (2019). Robust automated computational approach for classifying frontotemporal neurodegeneration: Multimodal/multicenter neuroimaging. *Alzheimer's & Dementia (Amsterdam, Netherlands)*, *11*, 588–598. https://doi.org/10.1016/j.dadm.2019.06.002

Dottori, M., Sedeño, L., Martorell Caro, M., Alifano, F., Hesse, E., Mikulan, E., García, A. M., Ruiz-Tagle, A., Lillo, P., Slachevsky, A., Serrano, C., Fraiman, D., & Ibanez, A. (2017). Towards affordable biomarkers of frontotemporal dementia: A classification study via network's information sharing. *Scientific Reports*, *7*(1), 3822. https://doi.org/10.1038/s41598-017-04204-8

Dubois, B., Defontaines, B., Deweer, B., Malapani, C., & Pillon, B. (1995). Cognitive and behavioral changes in patients with focal lesions of the basal ganglia. In *Behavioral neurology of movement disorders.* (pp. 29–41). Raven Press.

Ducharme, S., Dols, A., Laforce, R., Devenney, E., Kumfor, F., van den Stock, J., Dallaire-Théroux, C., Seelaar, H., Gossink, F., Vijverberg, E., Huey, E., Vandenbulcke, M., Masellis, M., Trieu, C., Onyike, C., Caramelli, P., de Souza, L. C., Santillo, A., Waldö, M. L., … Pijnenburg, Y. (2020). Recommendations to distinguish behavioural variant frontotemporal dementia from psychiatric disorders. *Brain*, *143*(6), 1632–1650. https://doi.org/10.1093/brain/awaa018

Elahi, F. M., & Miller, B. L. (2017). A clinicopathological approach to the diagnosis of dementia. *Nature Reviews Neurology*, *13*(8), 457–476. https://doi.org/10.1038/nrneurol.2017.96

Enders, C. K. (2010). *Applied Missing Data Analysis*. Guilford Press.

Escudero, J., Zajicek, J. P., & Ifeachor, E. (2011). Early detection and characterization of Alzheimer's disease in clinical scenarios using Bioprofile concepts and K-means. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 6470–6473. https://doi.org/10.1109/IEMBS.2011.6091597

Fan, J. M., Gorno-Tempini, M. L., Dronkers, N. F., Miller, B. L., Berger, M. S., & Chang, E. F. (2020). Data-Driven, Visual Framework for the Characterization of Aphasias Across Stroke, Post-resective, and Neurodegenerative Disorders Over Time. *Frontiers in Neurology*, *11*(December), 1–10. https://doi.org/10.3389/fneur.2020.616764

Farouk, Y., & Rady, S. (2020). Early Diagnosis of Alzheimer's Disease using Unsupervised Clustering. *International Journal of Intelligent Computing and Information Sciences*, *20*(2), 112–124. https://doi.org/10.21608/ijicis.2021.51180.1044

Feczko, E., Miranda-Dominguez, O., Marr, M., Graham, A. M., Nigg, J. T., & Fair, D. A.

(2019). The Heterogeneity Problem: Approaches to Identify Psychiatric Subtypes. *Trends in Cognitive Sciences*, *23*(7), 584–601. https://doi.org/10.1016/j.tics.2019.03.009

Feis, R. A., Bouts, M. J. R. J., de Vos, F., Schouten, T. M., Panman, J. L., Jiskoot, L. C., Dopper, E. G. P., van der Grond, J., van Swieten, J. C., & Rombouts, S. A. R. B. (2019). A multimodal MRI-based classification signature emerges just prior to symptom onset in frontotemporal dementia mutation carriers. *Journal of Neurology, Neurosurgery, and Psychiatry*, *90*(11), 1207–1214. https://doi.org/10.1136/jnnp-2019-320774

Feis, R. A., Bouts, M. J. R. J. R. J., Panman, J. L., Jiskoot, L. C., Dopper, E. G. P. P., Schouten, T. M., de Vos, F., van der Grond, J., van Swieten, J. C., & Rombouts, S. A. R. B. R. B. (2019). Single-subject classification of presymptomatic frontotemporal dementia mutation carriers using multimodal MRI. *NeuroImage. Clinical*, *20*, 188–196. https://doi.org/10.1016/j.nicl.2018.07.014

Ficiarà, E., Boschi, S., Ansari, S., D'Agata, F., Abollino, O., Caroppo, P., Di Fede, G., Indaco, A., Rainero, I., & Guiot, C. (2021). Machine Learning Profiling of Alzheimer's Disease Patients Based on Current Cerebrospinal Fluid Markers and Iron Content in Biofluids. *Frontiers in Aging Neuroscience*, *13*, 607858. https://doi.org/10.3389/fnagi.2021.607858

Fillenbaum, G. G., Belle, G. Van, Morris, J. C., Mohs, R. C., Mirra, S. S., Davis, P. C., Tariot, P. N., Silverman, J. M., Clark, C. M., Welsh-bohmer, K. A., & Heyman, A. (2008). Consortium to Establish a Registry for Alzheimer's Disease (CERAD): The first twenty years. *Alzheimer's & Dementia*, *4*, 96–109. https://doi.org/10.1016/j.jalz.2007.08.005

Flint, C., Cearns, M., Opel, N., Redlich, R., Mehler, D. M. A., Emden, D., Winter, N. R., Leenings, R., Eickhoff, S. B., Kircher, T., Krug, A., Nenadic, I., Arolt, V., Clark, S., Baune, B. T., Jiang, X., Dannlowski, U., & Hahn, T. (2021). Systematic misestimation of machine learning performance in neuroimaging studies of depression. *Neuropsychopharmacology*, *46*(8), 1510–1517. https://doi.org/10.1038/s41386-021-01020-7

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Foran, A., Mathias, J., & Bowden, S. (2021). Effectiveness of sorting tests for detecting cognitive decline in older adults with dementia and other common neurodegenerative disorders: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *120*(October 2020), 442–454. https://doi.org/10.1016/j.neubiorev.2020.10.013

Foster, N. L., Heidebrink, J. L., Clark, C. M., Jagust, W. J., Arnold, S. E., Barbas, N. R.,

DeCarli, C. S., Scott Turner, R., Koeppe, R. A., Higdon, R., & Minoshima, S. (2007). FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease. *Brain*, *130*(10), 2616–2635. https://doi.org/10.1093/brain/awm177

Foxe, D., Cheung, S. C., Cordato, N. J., Burrell, J. R., Ahmed, R. M., Taylor-Rubin, C., Irish, M., & Piguet, O. (2021). Verbal Short-Term Memory Disturbance in the Primary Progressive Aphasias: Challenges and Distinctions in a Clinical Setting. *Brain Sciences*, *11*(8), 1060. https://doi.org/10.3390/brainsci11081060

Foxe, D., Irish, M., Hodges, J. R., & Piguet, O. (2013). Verbal and Visuospatial Span in Logopenic Progressive Aphasia and Alzheimer's Disease. *Journal of the International Neuropsychological Society*, *19*(3), 247–253. https://doi.org/10.1017/S1355617712001269

Foxe, D., Irish, M., D'Mello, M., Barhon, L., Burrell, J. R., Kessels, R. P. C., & Piguet, O. (2021). The Box Task: A novel tool to differentiate the primary progressive aphasias. *European Journal of Neurology*, *28*(12), 3945–3954. https://doi.org/10.1111/ene.15035

Foxe, D., Irish, M., Roquet, D., Scharfenberg, A., Bradshaw, N., Hodges, J. R., Burrell, J. R., & Piguet, O. (2020). Visuospatial short-term and working memory disturbance in the primary progressive aphasias: Neuroanatomical and clinical implications. *Cortex*, *132*, 223–237. https://doi.org/10.1016/j.cortex.2020.08.018

Fränti, P., & Sieranoja, S. (2018). K-means properties on six clustering benchmark datasets. *Applied Intelligence*, *48*(12), 4743–4759. https://doi.org/10.1007/s10489-018-1238-7

Fraser, K. C., Meltzer, J. A., Graham, N. L., Leonard, C., Hirst, G., Black, S. E., & Rochon, E. (2014). Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *55*, 43–60. https://doi.org/10.1016/j.cortex.2012.12.006

Gamberger, D., Ženko, B., Mitelpunkt, A., Shachar, N., & Lavrač, N. (2016). Clusters of male and female Alzheimer's disease patients in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. *Brain Informatics*, *3*(3), 169–179. https://doi.org/10.1007/s40708-016-0035-5

Gambogi, L. B., Guimarães, H. C., De Souza, L. C., & Caramelli, P. (2019). Behavioral variant frontotemporal dementia in patients with previous severe mental illness: a systematic and critical review. *Arquivos de Neuro-Psiquiatria*, *77*(9), 654–668. https://doi.org/10.1590/0004-282x20190107

García-Laencina, P. J., Sancho-Gómez, J.-L., & Figueiras-Vidal, A. R. (2010). Pattern classification with missing data: a review. *Neural Computing and Applications*, *19*(2),

263–282. https://doi.org/10.1007/s00521-009-0295-6

Garcia-Gutierrez, F., Delgado-Alvarez, A., Delgado-Alonso, C., Díaz-Álvarez, J., Pytel, V., Valles-Salgado, M., Gil, M. J., Hernández-Lorenzo, L., Matías-Guiu, J., Ayala, J. L., & Matias-Guiu, J. A. (2022). Diagnosis of Alzheimer's disease and behavioural variant frontotemporal dementia with machine learning-aided neuropsychological assessment using feature engineering and genetic algorithms. *International Journal of Geriatric Psychiatry*, *37*(2), 1–13. https://doi.org/10.1002/gps.5667

Garn, H., Coronel, C., Waser, M., Caravias, G., & Ransmayr, G. (2017). Differential diagnosis between patients with probable Alzheimer's disease, Parkinson's disease dementia, or dementia with Lewy bodies and frontotemporal dementia, behavioral variant, using quantitative electroencephalographic features. *Journal of Neural Transmission (Vienna, Austria : 1996)*, *124*(5), 569–581. https://doi.org/10.1007/s00702-017-1699-6

Garrard, P., Rentoumi, V., Gesierich, B., Miller, B., & Gorno-Tempini, M. L. (2014). Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *55*, 122–129. https://doi.org/10.1016/j.cortex.2013.05.008

Gil-Navarro, S., Lladó, A., Rami, L., Castellví, M., Bosch, B., Bargalló, N., Lomeña, F., Reñé, R., Montagut, N., Antonell, A., Molinuevo, J. L., & Sánchez-Valle, R. (2013). Neuroimaging and Biochemical Markers in the Three Variants of Primary Progressive Aphasia. *Dementia and Geriatric Cognitive Disorders*, *35*(1–2), 106–117. https://doi.org/10.1159/000346289

Glosser, G., Gallo, J. L., Clark, C. M., & Grossman, M. (2002). Memory encoding and retrieval in frontotemporal dementia and Alzheimer's disease. *Neuropsychology*, *16*(2), 190–196. https://doi.org/10.1037/0894-4105.16.2.190

Goldman, J. S., Farmer, J. M., Wood, E. M., Johnson, J. K., Boxer, A., Neuhaus, J., Lomen-Hoerth, C., Wilhelmsen, K. C., Lee, V. M. Y., Grossman, M., & Miller, B. L. (2005). Comparison of family histories in FTLD subtypes and related tauopathies. *Neurology*, *65*(11), 1817–1819. https://doi.org/10.1212/01.wnl.0000187068.92184.63

Goodglass, H., & Kaplan, E. (1972). The assessment of aphasia and related disorders. *Lea & Febiger*.

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

Gorno-Tempini, M. L., Brambati, S. M., Ginex, V., Ogar, J., Dronkers, N. F., Marcone, A., Perani, D., Garibotto, V., Cappa, S. F., & Miller, B. L. (2008). The logopenic/phonological

variant of primary progressive aphasia. *Neurology*, *71*(16), 1227–1234. https://doi.org/10.1212/01.wnl.0000320506.79811.da

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., Ogar, J. M., Rohrer, J. D., Black, S., Boeve, B. F., Manes, F., Dronkers, N. F., Vandenberghe, R., Rascovsky, K., Patterson, K., Miller, B. L., Knopman, D. S., Hodges, J. R., Mesulam, M. M., & Grossman, M. (2011). Classification of primary progressive aphasia and its variants. *Neurology*, *76*(11), 1006–1014. https://doi.org/10.1212/WNL.0b013e31821103e6

Gossink, F., Schouws, S., Krudop, W., Scheltens, P., Stek, M., Pijnenburg, Y., & Dols, A. (2018). Social Cognition Differentiates Behavioral Variant Frontotemporal Dementia From Other Neurodegenerative Diseases and Psychiatric Disorders. *The American Journal of Geriatric Psychiatry*, *26*(5), 569–579. https://doi.org/10.1016/j.jagp.2017.12.008

Grace, J., & Malloy, P. H. (2001). Frontal systems behavior scale (FrSBe): Professional manual. *Psychological Assessment Resources (PAR).*

Graham, A. (2005). Pathologically proven frontotemporal dementia presenting with severe amnesia. *Brain*, *128*(3), 597–605. https://doi.org/10.1093/brain/awh348

Grossman, M. (2010). Primary progressive aphasia: clinicopathological correlations. *Nature Reviews Neurology*, *6*(2), 88–97. https://doi.org/10.1038/nrneurol.2009.216

Grote, T., & Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics*, *46*(3), 205–211. https://doi.org/10.1136/medethics-2019-105586

Guillén, E. F., Rosales, J. J., Lisei, D., Grisanti, F., Riverol, M., & Arbizu, J. (2020). Current role of 18F-FDG-PET in the differential diagnosis of the main forms of dementia. *Clinical and Translational Imaging*, *8*(3), 127–140. https://doi.org/10.1007/s40336-020-00366-0

Habes, M., Grothe, M. J., Tunc, B., McMillan, C., Wolk, D. A., & Davatzikos, C. (2020). Disentangling Heterogeneity in Alzheimer's Disease and Related Dementias Using Data-Driven Methods. *Biological Psychiatry*, *88*(1), 70–82. https://doi.org/10.1016/j.biopsych.2020.01.016

Hall, B., Mak, E., Cervenka, S., Aigbirhio, F. I., Rowe, J. B., & O'Brien, J. T. (2017). In vivo tau PET imaging in dementia: Pathophysiology, radiotracer quantification, and a systematic review of clinical findings. *Ageing Research Reviews*, *36*, 50–63. https://doi.org/10.1016/j.arr.2017.03.002

Hardy, C. J. D., Marshall, C. R., Bond, R. L., Russell, L. L., Dick, K., Ariti, C., Thomas, D. L., Ross, S. J., Agustus, J. L., Crutch, S. J., Rohrer, J. D., Bamiou, D.-E., & Warren, J. D.

(2018). Retained capacity for perceptual learning of degraded speech in primary progressive aphasia and Alzheimer's disease. *Alzheimer's Research & Therapy*, *10*(1), 70. https://doi.org/10.1186/s13195-018-0399-2

Harper, L., Fumagalli, G. G., Barkhof, F., Scheltens, P., O'Brien, J. T., Bouwman, F., Burton, E. J., Rohrer, J. D., Fox, N. C., Ridgway, G. R., & Schott, J. M. (2016). MRI visual rating scales in the diagnosis of dementia: evaluation in 184 post-mortem confirmed cases. *Brain : A Journal of Neurology*, *139*(Pt 4), 1211–1225. https://doi.org/10.1093/brain/aww005

Harris, J. M., Gall, C., Thompson, J. C., Richardson, A. M. T., Neary, D., du Plessis, D., Pal, P., Mann, D. M. A., Snowden, J. S., & Jones, M. (2013). Classification and pathology of primary progressive aphasia. *Neurology*, *81*(21), 1832–1839. https://doi.org/10.1212/01.wnl.0000436070.28137.7b

Harris, J. M., Saxon, J. A., Jones, M., Snowden, J. S., & Thompson, J. C. (2019). Neuropsychological differentiation of progressive aphasic disorders. *Journal of Neuropsychology*, *13*(2), 214–239. https://doi.org/10.1111/jnp.12149

Hayati Rezvan, P., Lee, K. J., & Simpson, J. A. (2015). The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Medical Research Methodology*, *15*(1), 30. https://doi.org/10.1186/s12874-015-0022-1

Heitkamp, N., Leiss, E., & Danek, A. (2010). Repeat & Point German Adaptation of a tool for differentiating semantic dementia and primary progressive aphasia. *Klinische Neurophysiologie*, *41*(01).

Hennig, C. (2015). What are the true clusters? *Pattern Recognition Letters*, *64*, 53–62. https://doi.org/10.1016/j.patrec.2015.04.009

Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (Eds.). (2015). *Handbook of Cluster Analysis* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b19706

Henry, M., & Grasso, S. (2018). Assessment of Individuals with Primary Progressive Aphasia. *Seminars in Speech and Language*, *39*(03), 231–241. https://doi.org/10.1055/s-0038-1660782

Henry, M., & Gorno-Tempini, M. L. (2010). The logopenic variant of primary progressive aphasia. *Current Opinion in Neurology*, *23*(6), 633–637. https://doi.org/10.1097/WCO.0b013e32833fb93e

Hindmarch, I., Lehfeld, H., de Jongh, P., & Erzigkeit, H. (1998). The Bayer Activities of Daily Living Scale (B-ADL). *Dementia and Geriatric Cognitive Disorders*, *9*(2), 20–26. https://doi.org/10.1159/000051195

Hinterbuchinger, B., Kaltenboeck, A., Baumgartner, J. S., Mossaheb, N., & Friedrich, F. (2018). Do patients with different psychiatric disorders show altered social decision-making? A systematic review of ultimatum game experiments in clinical populations. *Cognitive Neuropsychiatry*, *23*(3), 117–141. https://doi.org/10.1080/13546805.2018.1453791

Hodges, J. R., Martinos, M., Woollams, A. M., Patterson, K., & Adlam, A.-L. R. (2008). Repeat and Point: Differentiating semantic dementia from progressive non-fluent aphasia. *Cortex*, *44*(9), 1265–1270. https://doi.org/10.1016/j.cortex.2007.08.018

Hogan, D. B., Jetté, N., Fiest, K. M., Roberts, J. I., Pearson, D., Smith, E. E., Roach, P., Kirk, A., Pringsheim, T., & Maxwell, C. J. (2016). The Prevalence and Incidence of Frontotemporal Dementia: a Systematic Review. *The Canadian Journal of Neurological Sciences. Le Journal Canadien Des Sciences Neurologiques*, *43 Suppl 1*, S96–S109. https://doi.org/10.1017/cjn.2016.25

Holzinger, A., Saranti, A., Molnar, C., Biecek, P., & Samek, W. (2022). *Explainable AI Methods - A Brief Overview* (pp. 13–38). https://doi.org/10.1007/978-3-031-04083-2_2

Horn, J.-F., Habert, M.-O., Kas, A., Malek, Z., Maksud, P., Lacomblez, L., Giron, A., & Fertil, B. (2009). Differential automatic diagnosis between Alzheimer's disease and frontotemporal dementia based on perfusion SPECT images. *Artificial Intelligence in Medicine*, *47*(2), 147–158. https://doi.org/10.1016/j.artmed.2009.05.001

Hornberger, M., Piguet, O., Graham, A. J., Nestor, P. J., & Hodges, J. R. (2010). How preserved is episodic memory in behavioral variant frontotemporal dementia? *Neurology*, *74*(6), 472–479. https://doi.org/10.1212/WNL.0b013e3181cef85d

Hornberger, M., & Piguet, O. (2012). Episodic memory in frontotemporal dementia: a critical review. *Brain*, *135*(3), 678–692. https://doi.org/10.1093/brain/aws011

Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. Data Mining and Knowledge Discovery 2, 283-304. *Data Mining and Knowledge Discovery*, *2*(3), 283–304. https://www.researchgate.net/publication/220451944_Huang_Z_Extensions_to_the_k-Means_Algorithm_for_Clustering_Large_Data_Sets_with_Categorical_Values_Data_Mining_and_Knowledge_Discovery_2_283-304

Hui Xiong, Junjie Wu, & Jian Chen. (2009). K-Means Clustering Versus Validation Measures: A Data-Distribution Perspective. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *39*(2), 318–331. https://doi.org/10.1109/TSMCB.2008.2004559

Hutchinson, A. D., & Mathias, J. L. (2007). *Neuropsychological deficits in frontotemporal*

dementia and Alzheimer's disease: a meta-analytic review. https://doi.org/10.1136/jnnp.2006.100669

Illán-Gala, I., Casaletto, K. B., Borrego-Écija, S., Arenaza-Urquijo, E. M., Wolf, A., Cobigo, Y., Goh, S. Y. M., Staffaroni, A. M., Alcolea, D., Fortea, J., Blesa, R., Clarimon, J., Iulita, M. F., Brugulat-Serrat, A., Lladó, A., Grinberg, L. T., Possin, K., Rankin, K. P., Kramer, J. H., … Rosen, H. J. (2021). Sex differences in the behavioral variant of frontotemporal dementia: A new window to executive and behavioral reserve. *Alzheimer's & Dementia*, *17*(8), 1329–1341. https://doi.org/10.1002/alz.12299

Ingram, R. U., Halai, A. D., Pobric, G., Sajjadi, S., Patterson, K., & Lambon Ralph, M. A. (2020). Graded, multidimensional intra- and intergroup variations in primary progressive aphasia and post-stroke aphasia. *Brain*, *143*(10), 3121–3135. https://doi.org/10.1093/brain/awaa245

Isaacs, B., & Kennie, A. T. (1973). The Set Test as an Aid to the Detection of Dementia in Old People. *British Journal of Psychiatry*, *123*(575), 467–470. https://doi.org/10.1192/bjp.123.4.467

Ishihara, T., & Terada, S. (2001). Geriatric Depression Scale (GDS). In *Management of Dementia: Vol. 69 Suppl 8* (Issue 4, pp. 152–152). CRC Press. https://doi.org/10.3109/9780203213896-32

Jiskoot, L. C., Panman, J. L., Meeter, L. H., Dopper, E. G. P., Donker Kaat, L., Franzen, S., van der Ende, E. L., van Minkelen, R., Rombouts, S. A. R. B., Papma, J. M., & 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, J. K., Diehl, J., Mendez, M. F., Neuhaus, J., Shapira, J. S., Forman, M., Chute, D. J., Roberson, E. D., Pace-Savitsky, C., Neumann, M., Chow, T. W., Rosen, H. J., Forstl, H., Kurz, A., & Miller, B. L. (2005). Frontotemporal Lobar Degeneration. *Archives of Neurology*, *62*(6), 87–114. https://doi.org/10.1001/archneur.62.6.925

Josephs, K. A., Whitwell, J. L., Knopman, D. S., Boeve, B. F., Vemuri, P., Senjem, M. L., Parisi, J. E., Ivnik, R. J., Dickson, D. W., Petersen, R. C., & 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

Juva, K., Sulkava, R., Erkinjuntti, T., Ylikoski, R., Valvanne, J., & Tilvis, R. (2009). Staging the severity of dementia: comparison of clinical (CDR, DSM-III-R), functional (ADL, IADL) and cognitive (MMSE) scales. *Acta Neurologica Scandinavica*, *90*(4), 293–298.

https://doi.org/10.1111/j.1600-0404.1994.tb02724.x

Kamath, V., Chaney, G.-A. S., DeRight, J., & Onyike, C. U. (2019). A meta-analysis of neuropsychological, social cognitive, and olfactory functioning in the behavioral and language variants of frontotemporal dementia. *Psychological Medicine*, *49*(16), 2669–2680. https://doi.org/10.1017/S0033291718003604

Kamath, V., Sutherland, E. R., & Chaney, G.-A. (2020). A Meta-Analysis of Neuropsychological Functioning in the Logopenic Variant of Primary Progressive Aphasia: Comparison with the Semantic and Non-Fluent Variants. *Journal of the International Neuropsychological Society*, *26*(3), 322–330. https://doi.org/10.1017/S1355617719001115

Kaplan, E., Goodglass, H., & Weintraub, S. (1978). The Boston Naming Test. *Veterans Administration Medical Center*.

Kertesz, A., Davidson, W., McCabe, P., & Munoz, D. (2003). Behavioral Quantitation Is More Sensitive Than Cognitive Testing in Frontotemporal Dementia. *Alzheimer Disease & Associated Disorders*, *17*(4), 223–229. https://doi.org/10.1097/00002093-200310000-00005

Khedairia, S., & Khadir, M. T. (2022). A multiple clustering combination approach based on iterative voting process. *Journal of King Saud University - Computer and Information Sciences*, *34*(1), 1370–1380. https://doi.org/10.1016/j.jksuci.2019.09.013

Khvostikov, A., Aderghal, K., Benois-Pineau, J., Krylov, A., & Catheline, G. (2018). *3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies*. http://arxiv.org/abs/1801.05968

Kim, E.-J., Sidhu, M., Gaus, S. E., Huang, E. J., Hof, P. R., Miller, B. L., DeArmond, S. J., & Seeley, W. W. (2012). Selective Frontoinsular von Economo Neuron and Fork Cell Loss in Early Behavioral Variant Frontotemporal Dementia. *Cerebral Cortex*, *22*(2), 251–259. https://doi.org/10.1093/cercor/bhr004

Kim, J. P., Kim, J., Park, Y. H., Park, S. B., Lee, J. S., Yoo, S., Kim, E.-J., Kim, H. J., Na, D. L., Brown, J. A., Lockhart, S. N., Seo, S. W., & Seong, J.-K. (2019). Machine learning based hierarchical classification of frontotemporal dementia and Alzheimer's disease. *NeuroImage. Clinical*, *23*, 101811. https://doi.org/10.1016/j.nicl.2019.101811

Klöppel, S., Kotschi, M., Peter, J., Egger, K., Hausner, L., Frölich, L., Förster, A., Heimbach, B., Normann, C., Vach, W., Urbach, H., & Abdulkadir, A. (2018). Separating Symptomatic Alzheimer's Disease from Depression based on Structural MRI. *Journal of Alzheimer's Disease : JAD*, *63*(1), 353–363. https://doi.org/10.3233/JAD-170964

Kloppel, S., Stonnington, C. M., Barnes, J., Chen, F., Chu, C., Good, C. D., Mader, I., Mitchell, L. A., Patel, A. C., Roberts, C. C., Fox, N. C., Jack, C. R., Ashburner, J., & Frackowiak, R. S. J. (2008). Accuracy of dementia diagnosis--a direct comparison between radiologists and a computerized method. *Brain*, *131*(11), 2969–2974. https://doi.org/10.1093/brain/awn239

Knibb, J. A., Xuereb, J. H., Patterson, K., & Hodges, J. R. (2006). Clinical and pathological characterization of progressive aphasia. *Annals of Neurology*, *59*(1), 156–165. https://doi.org/10.1002/ana.20700

Knopman, D. S., Kramer, J. H., Boeve, B. F., Caselli, R. J., Graff-Radford, N. R., Mendez, M. F., Miller, B. L., & 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

Knopman, D. S., & Roberts, R. O. (2011). Estimating the Number of Persons with Frontotemporal Lobar Degeneration in the US Population. *Journal of Molecular Neuroscience*, *45*(3), 330–335. https://doi.org/10.1007/s12031-011-9538-y

Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, *23*(1), 89–109. https://doi.org/10.1016/S0933-3657(01)00077-X

Koutsouleris, N., Meisenzahl, E. M., Davatzikos, C., Bottlender, R., Frodl, T., Scheuerecker, J., Schmitt, G., Zetzsche, T., Decker, P., Reiser, M., Möller, H.-J., & Gaser, C. (2009). Use of Neuroanatomical Pattern Classification to Identify Subjects in At-Risk Mental States of Psychosis and Predict Disease Transition. *Archives of General Psychiatry*, *66*(7), 700. https://doi.org/10.1001/archgenpsychiatry.2009.62

Kramer, J. H., Jurik, J., Sha, S. J., Rankin, K. P., Rosen, H. J., Johnson, J. K., & Miller, B. L. (2003). Distinctive Neuropsychological Patterns in Frontotemporal Dementia, Semantic Dementia, And Alzheimer Disease. *Cognitive and Behavioral Neurology*, *16*(4), 211–218. https://doi.org/10.1097/00146965-200312000-00002

Kuring, J. K., Mathias, J. L., & Ward, L. (2018). Prevalence of Depression, Anxiety and PTSD in People with Dementia: a Systematic Review and Meta-Analysis. *Neuropsychology Review*, *28*(4), 393–416. https://doi.org/10.1007/s11065-018-9396-2

Lage, C., López-García, S., Bejanin, A., Kazimierczak, M., Aracil-Bolaños, I., Calvo-Córdoba, A., Pozueta, A., García-Martínez, M., Fernández-Rodríguez, A., Bravo-González, M., Jiménez-Bonilla, J., Banzo, I., Irure-Ventura, J., Pegueroles, J., Illán-Gala, I., Fortea, J., Rodríguez-Rodríguez, E., Lleó-Bisa, A., García-Cena, C. E., & Sánchez-Juan, P. (2020).

Distinctive Oculomotor Behaviors in Alzheimer's Disease and Frontotemporal Dementia. *Frontiers in Aging Neuroscience*, *12*, 603790. https://doi.org/10.3389/fnagi.2020.603790

Lanata, S. C., & Miller, B. L. (2016). The behavioural variant frontotemporal dementia (bvFTD) syndrome in psychiatry. *Journal of Neurology, Neurosurgery & Psychiatry*, *87*(5), 501–511. https://doi.org/10.1136/jnnp-2015-310697

Lecerf, S., Leroy, M., Lebouvier, T., Lebert, F., Deramecourt, V., Maurage, C. A., & Pasquier, F. (2020). Alzheimer's disease phenotypes misdiagnosed with frontotemporal lobar degeneration: A retrospective neuropathologic study. *Alzheimer's & Dementia*, *16*(S2), 39692. https://doi.org/10.1002/alz.039692

Lei, D., Pinaya, W. H. L., van Amelsvoort, T., Marcelis, M., Donohoe, G., Mothersill, D. O., Corvin, A., Gill, M., Vieira, S., Huang, X., Lui, S., Scarpazza, C., Young, J., Arango, C., Bullmore, E., Qiyong, G., McGuire, P., & Mechelli, A. (2019). Detecting schizophrenia at the level of the individual: relative diagnostic value of whole-brain images, connectome-wide functional connectivity and graph-based metrics. *Psychological Medicine*, 1–10. https://doi.org/10.1017/S0033291719001934

Leyton, C. E., Ballard, K. J., Piguet, O., & Hodges, J. R. (2014). Phonologic errors as a clinical marker of the logopenic variant of PPA. *Neurology*, *82*(18), 1620–1627. https://doi.org/10.1212/WNL.0000000000000387

Li, F., & Liu, M. (2018). Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. *Computerized Medical Imaging and Graphics*, *70*, 101–110. https://doi.org/10.1016/j.compmedimag.2018.09.009

Libon, D. J., Xie, S. X., Moore, P., Farmer, J., Antani, S., McCawley, G., Cross, K., & Grossman, M. (2007). Patterns of neuropsychological impairment in frontotemporal dementia. *Neurology*, *68*(5), 369–375. https://doi.org/10.1212/01.wnl.0000252820.81313.9b

Libon, D. J., Xie, S. X., Wang, X., Massimo, L., Moore, P., Vesely, L., Khan, A., Chatterjee, A., Coslett, H. B., Hurtig, H. I., Liang, T. W., & Grossman, M. (2009). Neuropsychological Decline in Frontotemporal Lobar Degeneration: A Longitudinal Analysis. *Neuropsychology*, *23*(3), 337–346. https://doi.org/10.1037/a0014995

Likas, A., Vlassis, N., & J. Verbeek, J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, *36*(2), 451–461. https://doi.org/10.1016/S0031-3203(02)00060-2

Lin, C.-H., Chiu, S.-I., Chen, T.-F., Jang, J.-S. R., & Chiu, M.-J. (2020). Classifications of Neurodegenerative Disorders Using a Multiplex Blood Biomarkers-Based Machine Learning Model. *International Journal of Molecular Sciences*, *21*(18).

https://doi.org/10.3390/ijms21186914

Livinț Popa, L., Dragoș, H.-M. M., Strilciuc, Ștefan, Pantelemon, C., Mureșanu, I., Dina, C., Văcăraș, V., & Mureșanu, D. (2021). Added Value of QEEG for the Differential Diagnosis of Common Forms of Dementia. *Clinical EEG and Neuroscience*, *52*(3), 201–210. https://doi.org/10.1177/1550059420971122

Ljubenkov, P. A., & Boxer, A. L. (2021). *FTLD Treatment: Current Practice and Future Possibilities* (pp. 297–310). https://doi.org/10.1007/978-3-030-51140-1_18

Ma, D., Lu, D., Popuri, K., & Beg, M. F. (2021). Differential Diagnosis of Frontotemporal Dementia and Alzheimer's Disease using Generative Adversarial Network. *ArXiv Preprint*, 1–9. https://doi.org/https://doi.org/10.48550/arXiv.2109.05627

Machulda, M. M., Whitwell, J. L., Duffy, J. R., Strand, E. A., Dean, P. M., Senjem, M. L., Jack, C. R., & Josephs, K. A. (2013). Identification of an atypical variant of logopenic progressive aphasia. *Brain and Language*, *127*(2), 139–144. https://doi.org/10.1016/j.bandl.2013.02.007

Macoir, J., Légaré, A., & Lavoie, M. (2021). Contribution of the Cognitive Approach to Language Assessment to the Differential Diagnosis of Primary Progressive Aphasia. *Brain Sciences*, *11*(6), 815. https://doi.org/10.3390/brainsci11060815

MacPherson, S. E., Wagner, G. P., Murphy, P., Bozzali, M., Cipolotti, L., & Shallice, T. (2014). Bringing the Cognitive Estimation Task into the 21st Century: Normative Data on Two New Parallel Forms. *PLoS ONE*, *9*(3), e92554. https://doi.org/10.1371/journal.pone.0092554

Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology*, *110*, 63–73. https://doi.org/10.1016/j.jclinepi.2019.02.016

Marin, R. S., Biedrzycki, R. C., & Firinciogullari, S. (1991). Reliability and validity of the apathy evaluation scale. *Psychiatry Research*, *38*(2), 143–162. https://doi.org/10.1016/0165-1781(91)90040-V

Marshall, C. R., Hardy, C. J. D., Volkmer, A., Russell, L. L., Bond, R. L., Fletcher, P. D., Clark, C. N., Mummery, C. J., Schott, J. M., Rossor, M. N., Fox, N. C., Crutch, S. J., Rohrer, J. 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

Maruta, C., Pereira, T., Madeira, S. C., De Mendonça, A., & Guerreiro, M. (2015). Classification of primary progressive aphasia: Do unsupervised data mining methods support a logopenic variant? *Amyotrophic Lateral Sclerosis and Frontotemporal*

Degeneration, *16*(3–4), 147–159. https://doi.org/10.3109/21678421.2015.1026266

Massimo, L., Kales, H. C., & Kolanowski, A. (2018). State of the Science: Apathy As a Model for Investigating Behavioral and Psychological Symptoms in Dementia. *Journal of the American Geriatrics Society*, *66*(Suppl 1), S4–S12. https://doi.org/10.1111/jgs.15343

Matias-Guiu, J. A., Díaz-Álvarez, J., Ayala, J. L., Risco-Martín, J. L., Moreno-Ramos, T., Pytel, V., Matias-Guiu, J., Carreras, J. L., & Cabrera-Martín, M. N. (2018). Clustering Analysis of FDG-PET Imaging in Primary Progressive Aphasia. *Frontiers in Aging Neuroscience*, *10*, 230. https://doi.org/10.3389/fnagi.2018.00230

Matias-Guiu, J. A., Díaz-Alvarez, J., Cuetos, F., Cabrera-Martín, M. N., Segovia-Ríos, I., Pytel, V., Moreno-Ramos, T., Carreras, J. L., Matías-Guiu, J., & Ayala, J. L. (2019). Machine learning in the clinical and language characterisation of primary progressive aphasia variants. *Cortex*, *119*, 312–323. https://doi.org/https://doi.org/10.1016/j.cortex.2019.05.007

Matias-Guiu, J. A., Suárez-Coalla, P., Pytel, V., Cabrera-Martín, M. N., Moreno-Ramos, T., Delgado-Alonso, C., Delgado-Álvarez, A., Matías-Guiu, J., & Cuetos, F. (2020). Reading prosody in the non-fluent and logopenic variants of primary progressive aphasia. *Cortex*, *132*, 63–78. https://doi.org/10.1016/j.cortex.2020.08.013

Mesulam, M.-M. (1982). Slowly progressive aphasia without generalized dementia. *Annals of Neurology*, *11*(6), 592–598. https://doi.org/10.1002/ana.410110607

Mesulam, M.-M., Coventry, C., Bigio, E. H., Geula, C., Thompson, C., Bonakdarpour, B., Gefen, T., Rogalski, E. J., & Weintraub, S. (2021). *Nosology of Primary Progressive Aphasia and the Neuropathology of Language* (pp. 33–49). https://doi.org/10.1007/978-3-030-51140-1_3

Mesulam, M. M. (2001). Primary progressive aphasia. *Annals of Neurology*, *49*(4), 425–432. https://doi.org/10.4414/sanp.2020.03101

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

Meyer, A. M., Snider, S. F., Campbell, R. E., & Friedman, R. B. (2015). Phonological short-term memory in logopenic variant primary progressive aphasia and mild Alzheimer's disease. *Cortex*, *71*, 183–189. https://doi.org/10.1016/j.cortex.2015.07.003

Meyer, S., Mueller, K., Stuke, K., Bisenius, S., Diehl-Schmid, J., Jessen, F., Kassubek, J., Kornhuber, J., Ludolph, A. C., Prudlo, J., Schneider, A., Schuemberg, K., Yakushev, I., Otto, M., & Schroeter, M. L. (2017). Predicting behavioral variant frontotemporal

dementia with pattern classification in multi-center structural MRI data. *NeuroImage: Clinical*, *14*, 656–662. https://doi.org/10.1016/j.nicl.2017.02.001

Micanovic, C., & Pal, S. (2014). The diagnostic utility of EEG in early-onset dementia: a systematic review of the literature with narrative analysis. *Journal of Neural Transmission*, *121*(1), 59–69. https://doi.org/10.1007/s00702-013-1070-5

Mittal, M., Goyal, L. M., Hemanth, D. J., & Sethi, J. K. (2019). Clustering approaches for high-dimensional databases: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(3), 1–14. https://doi.org/10.1002/widm.1300

Modirrousta, M., Price, B. H., & Dickerson, B. C. (2013). Neuropsychiatric symptoms in primary progressive aphasia: phenomenology, pathophysiology, and approach to assessment and treatment. *Neurodegenerative Disease Management*, *3*(2), 133–146. https://doi.org/10.2217/nmt.13.6

Moguilner, S., García, A. M., Mikulan, E., Hesse, E., García-Cordero, I., Melloni, M., Cervetto, S., Serrano, C., Herrera, E., Reyes, P., Matallana, D., Manes, F., Ibáñez, A., & Sedeño, L. (2018). Weighted Symbolic Dependence Metric (wSDM) for fMRI resting-state connectivity: A multicentric validation for frontotemporal dementia. *Scientific Reports*, *8*(1), 11181. https://doi.org/10.1038/s41598-018-29538-9

Moguilner, S., García, A. M., Perl, Y. S., Tagliazucchi, E., Piguet, O., Kumfor, F., Reyes, P., Matallana, D., Sedeño, L., & Ibáñez, A. (2021). Dynamic brain fluctuations outperform connectivity measures and mirror pathophysiological profiles across dementia subtypes: A multicenter study. *NeuroImage*, *225*, 117522. https://doi.org/10.1016/j.neuroimage.2020.117522

Mohs, R. C., Kim, Y., Johns, C. A., Dunn, D. D., & Davis, K. L. (1986). Assessing changes in Alzheimer's disease: Memory and language. In *Handbook for clinical memory assessment of older adults.* (pp. 149–155). American Psychological Association. https://doi.org/10.1037/10057-012

Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., & Verbeke, G. (2014). *Handbooks of Modern Statistical Methods Handbook of Missing Data Methodology*.

Moms, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., Mellits, E. D., & Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assesment of Alzheimer's disease. *Neurology*, *39*(9), 1159–1159. https://doi.org/10.1212/WNL.39.9.1159

Morris, J. C. (1993). The Clinical Dementia Rating (CDR) Current version and scoring rules. *Neurology*, *43*(11), 2412.2-2412-a. https://doi.org/10.1212/WNL.43.11.2412-a

Mulder-Heijstra, M. M. P., Jokel, R. R., Chertkow, H. H., Conn, D. D. K., & Mah, L. L. (2021). Primary Progressive Aphasia Presenting With Neuropsychiatric Symptoms. *Journal of Geriatric Psychiatry and Neurology*, 089198872110361. https://doi.org/10.1177/08919887211036189

Muñoz-Neira, C., Tedde, A., Coulthard, E., Thai, N. J., & Pennington, C. (2019). Neural correlates of altered insight in frontotemporal dementia: a systematic review. *NeuroImage. Clinical*, *24*, 102066. https://doi.org/10.1016/j.nicl.2019.102066

Murley, A. G., Coyle-Gilchrist, I., Rouse, M. A., Jones, P. S., Li, W., Wiggins, J., Lansdall, C., Rodríguez, P. V., Wilcox, A., Tsvetanov, K. A., Patterson, K., Lambon Ralph, M. A., & Rowe, J. B. (2020). Redefining the multidimensional clinical phenotypes of frontotemporal lobar degeneration syndromes. *Brain*, *143*(5), 1555–1571. https://doi.org/10.1093/brain/awaa097

Musa, G., Slachevsky, A., Muñoz-Neira, C., Méndez-Orellana, C., Villagra, R., González-Billault, C., Ibáñez, A., Hornberger, M., & Lillo, P. (2020). Alzheimer's Disease or Behavioral Variant Frontotemporal Dementia? Review of Key Points Toward an Accurate Clinical and Neuropsychological Diagnosis. *Journal of Alzheimer's Disease*, *73*(3), 833–848. https://doi.org/10.3233/JAD-190924

Neary, D., Brun, A., Englund, B., Gustafson, L., Passant, U., Mann, D. M. A., & Snowden, J. S. (1994). Clinical and neuropathological criteria for frontotemporal dementia. The Lund and Manchester Groups. *Journal of Neurology, Neurosurgery & Psychiatry*, *57*(4), 416–418. https://doi.org/10.1136/jnnp.57.4.416

Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., Freedman, M., Kertesz, A., Robert, P. H., Albert, M., Boone, K., Miller, B. L., Cummings, J., & Benson, D. F. (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology*, *51*(6), 1546–1554. https://doi.org/10.1212/WNL.51.6.1546

O'Connor, C. M., Clemson, L., Hornberger, M., Leyton, C. E., Hodges, J. R., Piguet, O., & Mioshi, E. (2016). Longitudinal change in everyday function and behavioral symptoms in frontotemporal dementia. *Neurology: Clinical Practice*, *6*(5), 419–428. https://doi.org/10.1212/CPJ.0000000000000264

O'Connor, C. M., Landin-Romero, R., Clemson, L., Kaizik, C., Daveson, N., Hodges, J. R., Hsieh, S., Piguet, O., & Mioshi, E. (2017). Behavioral-variant frontotemporal dementia. *Neurology*, *89*(6), 570–577. https://doi.org/10.1212/WNL.0000000000004215

Olney, N. T., Spina, S., & Miller, B. L. (2017). Frontotemporal Dementia. *Neurologic Clinics*, *35*(2), 339–374. https://doi.org/10.1016/j.ncl.2017.01.008

Onyike, C. U., & Diehl-Schmid, J. (2013). The epidemiology of frontotemporal dementia. *International Review of Psychiatry*, *25*(2), 130–137. https://doi.org/10.3109/09540261.2013.776523

Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neuroscience and Biobehavioral Reviews*, *36*(4), 1140–1152. https://doi.org/10.1016/j.neubiorev.2012.01.004

Osher, J. E., Wicklund, A. H., Rademaker, A., Johnson, N., & Weintraub, S. (2008). The Mini-Mental State Examination in Behavioral Variant Frontotemporal Dementia and Primary Progressive Aphasia. *American Journal of Alzheimer's Disease & Other Dementiasr*, *22*(6), 468–473. https://doi.org/10.1177/1533317507307173

Otto, M., Ludolph, A. C., Landwehrmeyer, B., Förstl, H., Diehl-Schmid, J., Neumann, M., Kretzschmar, H. A., Schroeter, M., Kornhuber, J., & Danek, A. (2011). Konsortium zur Erforschung der frontotemporalen Lobärdegeneration. *Der Nervenarzt*, *82*(8), 1002–1005. https://doi.org/10.1007/s00115-011-3261-3

Overbeek, J. M., Korten, N., Gossink, F., Fieldhouse, J., van de Beek, M., Reus, L., Dols, A., Pijnenburg, Y., & Schouws, S. (2020). The Value of Neuropsychological Assessment in the Differentiation Between Behavioral Variant Frontotemporal Dementia and Late-Onset Psychiatric Disorders. *The Journal of Clinical Psychiatry*, *81*(1), 1–8. https://doi.org/10.4088/JCP.19m12811

Owens, T. E., Machulda, M. M., Duffy, J. R., Strand, E. A., Clark, H. M., Boland, S., Martin, P. R., Lowe, V. J., Jack, C. R., Whitwell, J. L., & Josephs, K. A. (2018). Patterns of Neuropsychological Dysfunction and Cortical Volume Changes in Logopenic Aphasia. *Journal of Alzheimer's Disease*, *66*(3), 1015–1025. https://doi.org/10.3233/JAD-171175

Pan, P. L., Song, W., Yang, J., Huang, R., Chen, K., Gong, Q. Y., Zhong, J. G., Shi, H. C., & Shang, 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*(2–3), 141–148. https://doi.org/10.1159/000338176

Panza, F., Lozupone, M., Seripa, D., Daniele, A., Watling, M., Giannelli, G., & Imbimbo, B. P. (2020). Development of disease-modifying drugs for frontotemporal dementia spectrum disorders. *Nature Reviews Neurology*, *16*(4), 213–228. https://doi.org/10.1038/s41582-020-0330-x

Parums, D. V. (2021). Editorial: The National COVID Cohort Collaborative Consortium Combines Population Data with Machine Learning to Evaluate and Predict Risk Factors

for the Severity of COVID-19. *Medical Science Monitor*, *27*, e934171. https://doi.org/10.12659/MSM.934171

Peet, B. T., Castro-Suarez, S., & Miller, B. L. (2021). *The Neuropsychiatric Features of Behavioral Variant Frontotemporal Dementia* (pp. 17–31). https://doi.org/10.1007/978-3-030-51140-1_2

Pick, A. (1901). Senile hirnatrophie als grundlage von Herderscheinungen. *Wiener Klinische Wochenschrift*, *14*, 16–17.

Pick, A. (1904). Zur Symptomatologie der linksseitigen Schläfenlappenatrophie. *Monatschrift Für Psychiatrie Und Neurologie*, *16*, 378–388.

Pick, Arnold. (1892). Uber die Beziehungen der senilen Hirnatrophie zur Aphasie. *Prag Med Wochenschr*, *17*, 165–167.

Piguet, O., Hornberger, M., Mioshi, E., & Hodges, J. R. (2011). Behavioural-variant frontotemporal dementia: diagnosis, clinical staging, and management. *The Lancet Neurology*, *10*(2), 162–172. https://doi.org/10.1016/S1474-4422(10)70299-4

Poonam, K., Guha, R., & Chakrabarti, P. P. (2021). *Artificial intelligence methods based hierarchical classification of frontotemporal dementia to improve diagnostic predictability*.

Postema, M. C., Rooij, D. van, Anagnostou, E., Arango, C., Auzias, G., Behrmann, M., Filho, G. B., Calderoni, S., Calvo, R., Daly, E., Deruelle, C., Martino, A. Di, Dinstein, I., Duran, F. L. S., Durston, S., Ecker, C., Ehrlich, S., Fair, D., Fedor, J., … Francks, C. (2019). *Altered structural brain asymmetry in autism spectrum disorder: large-scale analysis via the ENIGMA Consortium*. https://doi.org/10.1101/570655

Premi, E., Cauda, F., Costa, T., Diano, M., Gazzina, S., Gualeni, V., Alberici, A., Archetti, S., Magoni, M., Gasparotti, R., Padovani, A., & Borroni, B. (2016). Looking for Neuroimaging Markers in Frontotemporal Lobar Degeneration Clinical Trials: A Multi-Voxel Pattern Analysis Study in Granulin Disease. *Journal of Alzheimer's Disease : JAD*, *51*(1), 249–262. https://doi.org/10.3233/JAD-150340

Price, C. C., Tanner, J. J., Schmalfuss, I. M., Brumback, B., Heilman, K. M., & Libon, D. J. (2015). Dissociating Statistically-Determined Alzheimer's Disease/Vascular Dementia Neuropsychological Syndromes Using White and Gray Neuroradiological Parameters. *Journal of Alzheimer's Disease*, *48*(3), 833–847. https://doi.org/10.3233/JAD-150407

Primativo, S., Clark, C., Yong, K. X. X., Firth, N. C., Nicholas, J., Alexander, D., Warren, J. D., Rohrer, J. D., & Crutch, S. J. (2017). Eyetracking metrics reveal impaired spatial anticipation in behavioural variant frontotemporal dementia. *Neuropsychologia*, *106*,

328–340. https://doi.org/10.1016/j.neuropsychologia.2017.10.014

Qayyum, A., Qadir, J., Bilal, M., & Al-Fuqaha, A. (2021). Secure and Robust Machine Learning for Healthcare: A Survey. *IEEE Reviews in Biomedical Engineering*, *14*, 156–180. https://doi.org/10.1109/RBME.2020.3013489

R Core Team. (2021). *R: A language and environment for statistical computing*.

Rabinovici, G. D., & Miller, B. L. (2010). Frontotemporal Lobar Degeneration. *CNS Drugs*, *24*(5), 375–398. https://doi.org/10.2165/11533100-000000000-00000

Rabinovici, G. D., Stephens, M. L., & Possin, K. L. (2015). Executive Dysfunction. *CONTINUUM: Lifelong Learning in Neurology*, *21*(3), 646–659. https://doi.org/10.1212/01.CON.0000466658.05156.54

Raghunathan, T., Lepkowski, J., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, *27*(1), 85–96.

Ramanan, S., Roquet, D., Goldberg, Z., Hodges, J. R., Piguet, O., Irish, M., & Lambon Ralph, M. A. (2020). Establishing two principal dimensions of cognitive variation in logopenic progressive aphasia. *Brain Communications*. https://doi.org/10.1093/braincomms/fcaa125

Ranasinghe, K. G., Rankin, K. P., Pressman, P. S., Perry, D. C., Lobach, I. V., Seeley, W. W., Coppola, G., Karydas, A. M., Grinberg, L. T., Shany-Ur, T., Lee, S. E., Rabinovici, G. D., Rosen, H. J., Gorno-Tempini, M. L., Boxer, A. L., Miller, Z. A., Chiong, W., DeMay, M., Kramer, J. H., … Miller, B. L. (2016). Distinct Subtypes of Behavioral Variant Frontotemporal Dementia Based on Patterns of Network Degeneration. *JAMA Neurology*, *73*(9), 1078. https://doi.org/10.1001/jamaneurol.2016.2016

Rankin, K. P. (2021). *Measuring Behavior and Social Cognition in FTLD* (pp. 51–65). https://doi.org/10.1007/978-3-030-51140-1_4

Rankin, K. P., Santos-Modesitt, W., Kramer, J. H., Pavlic, D., Beckman, V., & Miller, B. L. (2008). Spontaneous Social Behaviors Discriminate Behavioral Dementias From Psychiatric Disorders and Other Dementias. *The Journal of Clinical Psychiatry*, *69*(1), 60–73. https://doi.org/10.4088/JCP.v69n0109

Rascovsky, K., & Grossman, M. (2013). Clinical diagnostic criteria and classification controversies in frontotemporal lobar degeneration. *International Review of Psychiatry*, *25*(2), 145–158. https://doi.org/10.3109/09540261.2013.763341

Rascovsky, K., Hodges, J. R., Knopman, D., Mendez, M. F., Kramer, J. H., Neuhaus, J., van Swieten, J. C., Seelaar, H., Dopper, E. G. P., Onyike, C. U., Hillis, A. E., Josephs, K. A., Boeve, B. F., Kertesz, A., Seeley, W. W., Rankin, K. P., Johnson, J. K., Gorno-Tempini,

M.-L., Rosen, H., … Miller, B. L. (2011). Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain*, *134*(9), 2456–2477. https://doi.org/10.1093/brain/awr179

Ratnavalli, E., Brayne, C., Dawson, K., & Hodges, J. R. (2002). The prevalence of frontotemporal dementia. *Neurology*, *58*(11), 1615–1621. https://doi.org/10.1212/WNL.58.11.1615

Regard, M., Strauss, E., & Knapp, P. (1982). Children's production on verbal and non-verbal fluency tasks. *Perceptual and Motor Skills*, *55*(3 Pt 1), 839–844. https://doi.org/10.2466/pms.1982.55.3.839

Reitan, R. M. (1958). Validity of the Trail Making Test as an Indicator of Organic Brain Damage. *Perceptual and Motor Skills*, *8*(3), 271–276. https://doi.org/10.2466/pms.1958.8.3.271

Reul, S., Lohmann, H., Wiendl, H., Duning, T., & Johnen, A. (2017). Can cognitive assessment really discriminate early stages of Alzheimer's and behavioural variant frontotemporal dementia at initial clinical presentation? *Alzheimer's Research & Therapy*, *9*(1), 61. https://doi.org/10.1186/s13195-017-0287-1

Riedijk, S. R., De Vugt, M. E., Duivenvoorden, H. J., Niermeijer, M. F., van Swieten, J. C., Verhey, F. R. J., & Tibben, A. (2006). Caregiver Burden, Health-Related Quality of Life and Coping in Dementia Caregivers: A Comparison of Frontotemporal Dementia and Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders*, *22*(5–6), 405–412. https://doi.org/10.1159/000095750

Rogalski, E., Rademaker, A., & Weintraub, S. (2007). Primary Progressive Aphasia: Relationship Between Gender and Severity of Language Impairment. *Cognitive and Behavioral Neurology*, *20*(1), 38–43. https://doi.org/10.1097/WNN.0b013e31802e3bae

Rohrer, J. D., Caso, F., Mahoney, C., Henry, M., Rosen, H. J., Rabinovici, G., Rossor, M. N., Miller, B., Warren, J. D., Fox, N. C., Ridgway, G. R., & Gorno-Tempini, M. L. (2013). Patterns of longitudinal brain atrophy in the logopenic variant of primary progressive aphasia. *Brain and Language*, *127*(2), 121–126. https://doi.org/10.1016/j.bandl.2012.12.008

Rohrer, J. D., Ridgway, G. R., Crutch, S. J., Hailstone, J., Goll, J. C., Clarkson, M. J., Mead, S., Beck, J., Mummery, C., Ourselin, S., Warrington, E. K., Rossor, M. N., & 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., Woollacott, I. O. C., Dick, K. M., Brotherhood, E., Gordon, E., Fellows, A.,

Toombs, J., Druyeh, R., Cardoso, M. J., Ourselin, S., Nicholas, J. M., Norgren, N., Mead, S., Andreasson, U., Blennow, K., Schott, J. M., Fox, N. C., Warren, J. D., & Zetterberg, H. (2016). Serum neurofilament light chain protein is a measure of disease intensity in frontotemporal dementia. *Neurology*, *87*(13), 1329–1336. https://doi.org/10.1212/WNL.0000000000003154

Roman Meller, M., Patel, S., Duarte, D., Kapczinski, F., de Azevedo Cardoso, T., Meller, M. R., Patel, S., Duarte, D., Kapczinski, F., & Cardoso, T. de A. (2021). Bipolar disorder and frontotemporal dementia: A systematic review. *Acta Psychiatrica Scandinavica*, *144*(5), 433–447. https://doi.org/10.1111/acps.13362

Romero, B., & Wenz, M. (2002). *Konzept und Wirksamkeit eines Behandlungsprogrammes für Demenzkranke und deren Angehörige*. *35*, 118–128.

Rosen, H. J., Lengenfelder, J., & Miller, B. (2000). FRONTOTEMPORAL DEMENTIA. *Neurologic Clinics*, *18*(4), 979–992. https://doi.org/10.1016/S0733-8619(05)70235-8

Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry*, *141*(11), 1356–1364. https://doi.org/10.1176/ajp.141.11.1356

Rosness, T. A., Haugen, P. K., Passant, U., & Engedal, K. (2008). Frontotemporal dementia--a clinically complex diagnosis. *International Journal of Geriatric Psychiatry*, *23*(8), 837–842. https://doi.org/10.1002/gps.1992

Rosso, S. M., Kaat, L. D., Baks, T., Joosse, M., de Koning, I., Pijnenburg, Y., Jong, D. de, Dooijes, D., Kamphorst, W., Ravid, R., Niermeijer, M. F., Verheij, F., Kremer, H. P., Scheltens, P., Duijn, C. M. van, Heutink, P., & van Swieten, J. C. (2003). Frontotemporal dementia in The Netherlands: patient characteristics and prevalence estimates from a population-based study. *Brain*, *126*(9), 2016–2022. https://doi.org/10.1093/brain/awg204

Rossor, M. N., Fox, N. C., Mummery, C. J., Schott, J. M., & Warren, J. D. (2010). The diagnosis of young-onset dementia. *The Lancet Neurology*, *9*(8), 793–806. https://doi.org/10.1016/S1474-4422(10)70159-9

RStudio Team. (2020). RStudio: Integrated Development for R. In *RStudio, PBC*.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Ruksenaite, J., Volkmer, A., Jiang, J., Johnson, J. C., Marshall, C. R., Warren, J. D., & Hardy, C. J. (2021). Primary Progressive Aphasia: Toward a Pathophysiological Synthesis. *Current Neurology and Neuroscience Reports*, *21*(3), 7. https://doi.org/10.1007/s11910-021-01097-z

Sachdev, P. S., Blacker, D., Blazer, D. G., Ganguli, M., Jeste, D. V., Paulsen, J. S., & Petersen, R. C. (2014). Classifying neurocognitive disorders: the DSM-5 approach. *Nature Reviews Neurology*, *10*(11), 634–642. https://doi.org/10.1038/nrneurol.2014.181

Sajjadi, S. A., Patterson, K., Arnold, R. J., Watson, P. C., & Nestor, P. J. (2012). Primary progressive aphasia: A tale of two syndromes and the rest. *Neurology*, *78*(21), 1670–1677. https://doi.org/10.1212/WNL.0b013e3182574f79

Sampath, R., & Saradha, A. (2014). Classification of Alzheimer's Disease Stages Exploiting an ANFIS Classifier. *Internatinal Journal of Applied Engineering Research*, *9*(22), 16979–16990. http://www.ripublication.com

Scarpazza, C., Sartori, G., De Simone, M. S., & Mechelli, A. (2013). When the single matters more than the group: Very high false positive rates in single case Voxel Based Morphometry. *NeuroImage*, *70*, 175–188. https://doi.org/10.1016/j.neuroimage.2012.12.045

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Scheltens, N. M. E., Galindo-Garre, F., Pijnenburg, Y. A. L., van der Vlies, A. E., Smits, L. L., Koene, T., Teunissen, C. E., Barkhof, F., Wattjes, M. P., Scheltens, P., & van der Flier, W. M. (2016). The identification of cognitive subtypes in Alzheimer's disease dementia using latent class analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, *87*(3), 235–243. https://doi.org/10.1136/jnnp-2014-309582

Schönecker, S., Hell, F., Bötzel, K., Wlasich, E., Ackl, N., Süßmair, C., Otto, M., Anderl-Straub, S., Ludolph, A., Kassubek, J., Huppertz, H. J., Diehl-Schmid, J., Riedl, L., Roßmeier, C., Fassbender, K., Lyros, E., Kornhuber, J., Oberstein, T. J., Fliessbach, K., … Danek, A. (2019). The applause sign in frontotemporal lobar degeneration and related conditions. *Journal of Neurology*, *266*(2), 330–338. https://doi.org/10.1007/s00415-018-9134-y

Schroeter, M. L., Laird, A. R., Chwiesko, C., Deuschl, C., Schneider, E., Bzdok, D., Eickhoff, S. B., & 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., Pawelke, S., Bisenius, S., Kynast, J., Schuemberg, K., Polyakova, M., Anderl-Straub, S., Danek, A., Fassbender, K., Jahn, H., Jessen, F., Kornhuber, J., Lauer, M., Prudlo, J., Schneider, A., Uttner, I., Thöne-Otto, A., Otto, M., & Diehl-Schmid, J. (2018). A Modified Reading the Mind in the Eyes Test Predicts Behavioral Variant

Frontotemporal Dementia Better Than Executive Function Tests. *Frontiers in Aging Neuroscience*, *10*(JAN), 1–11. https://doi.org/10.3389/fnagi.2018.00011

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

Sebastian, R., Thompson, C. B., Wang, N.-Y., Wright, A., Meyer, A., Friedman, R. B., Hillis, A. E., & Tippett, D. C. (2018). Patterns of decline in naming and semantic knowledge in primary progressive aphasia. *Aphasiology*, *32*(9), 1010–1030. https://doi.org/10.1080/02687038.2018.1490388

Seckin, M., Ricard, I., Raiser, T., Heitkamp, N., Ebert, A., Prix, C., Levin, J., Diehl-Schmid, J., Riedl, L., Roßmeier, C., Hoen, N., Schroeter, M. L., Marschhauser, A., Obrig, H., Benke, T., Kornhuber, J., Fliessbach, K., Schneider, A., Wiltfang, J., … Danek, A. (2022). Utility of the Repeat and Point Test for Subtyping Patients With Primary Progressive Aphasia. *Alzheimer Disease & Associated Disorders*, *Publish Ah*(00), 1–8. https://doi.org/10.1097/wad.0000000000000482

Seeley, W. W., Bauer, A. M., Miller, B. L., Gorno-Tempini, M. L., Kramer, J. H., Weiner, M., & Rosen, H. J. (2005). The natural history of temporal variant frontotemporal dementia. *Neurology*, *64*(8), 1384–1390. https://doi.org/10.1212/01.WNL.0000158425.46019.5C

Seeley, William W., Carlin, D. A., Allman, J. M., Macedo, M. N., Bush, C., Miller, B. L., & DeArmond, S. J. (2006). Early frontotemporal dementia targets neurons unique to apes and humans. *Annals of Neurology*, *60*(6), 660–667. https://doi.org/10.1002/ana.21055

Seeley, William W., Crawford, R. K., Zhou, J., Miller, B. L., & Greicius, M. D. (2009). Neurodegenerative Diseases Target Large-Scale Human Brain Networks. *Neuron*, *62*(1), 42–52. https://doi.org/10.1016/j.neuron.2009.03.024

Seeley, William 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

Şenbabaoğlu, Y., Michailidis, G., & Li, J. Z. (2015). Critical limitations of consensus clustering in class discovery. *Scientific Reports*, *4*(1), 6207. https://doi.org/10.1038/srep06207

Shahid, N., Rappon, T., & Berta, W. (2019). Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PLOS ONE*, *14*(2), e0212356. https://doi.org/10.1371/journal.pone.0212356

Shailaja, K., Seetharamulu, B., & Jabbar, M. A. (2018). Machine Learning in Healthcare: A

Review. *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, *25*(2), 910–914. https://doi.org/10.1109/ICECA.2018.8474918

Shallice, T., & Evans, M. E. (1978). The Involvement of the Frontal Lobes in Cognitive Estimation. *Cortex*, *14*(2), 294–303. https://doi.org/10.1016/S0010-9452(78)80055-0

Shaw, A. D., Hughes, L. E., Moran, R., Coyle-Gilchrist, I., Rittman, T., & Rowe, J. B. (2021). In Vivo Assay of Cortical Microcircuitry in Frontotemporal Dementia: A Platform for Experimental Medicine Studies. *Cerebral Cortex (New York, N.Y. : 1991)*, *31*(3), 1837–1847. https://doi.org/10.1093/cercor/bhz024

Shinagawa, S., Nakajima, S., Plitman, E., Graff-Guerrero, A., Mimura, M., Nakayama, K., & Miller, B. L. (2014). Psychosis in frontotemporal dementia. *Journal of Alzheimer's Disease : JAD*, *42*(2), 485–499. https://doi.org/10.3233/JAD-140312

Simjanoski, M., McIntyre, A., Kapczinski, F., & Azevedo, T. de. (2021). Cognitive impairment in bipolar disorder in comparison to mild cognitive impairment and dementia: a systematic review. *Trends in Psychiatry and Psychotherapy*. https://doi.org/10.47626/2237-6089-2021-0300

Sitek, E. J., Barczak, A., & Harciarek, M. (2015). Neuropsychological Assessment and Differential Diagnosis in Young-Onset Dementias. *Psychiatric Clinics of North America*, *38*(2), 265–279. https://doi.org/10.1016/j.psc.2015.01.003

Steinacker, P., Anderl-Straub, S., Diehl-Schmid, J., Semler, E., Uttner, I., von Arnim, C. A. F., Barthel, H., Danek, A., Fassbender, K., Fliessbach, K., Foerstl, H., Grimmer, T., Huppertz, H.-J., Jahn, H., Kassubek, J., Kornhuber, J., Landwehrmeyer, B., Lauer, M., Maler, J. M., … Otto, M. (2018). Serum neurofilament light chain in behavioral variant frontotemporal dementia. *Neurology*, *91*(15), e1390–e1401. https://doi.org/10.1212/WNL.0000000000006318

Steinacker, P., Semler, E., Anderl-Straub, S., Diehl-Schmid, J., Schroeter, M. L., Uttner, I., Foerstl, H., Landwehrmeyer, B., von Arnim, C. A. F., Kassubek, J., Oeckl, P., Huppertz, H.-J., Fassbender, K., Fliessbach, K., Prudlo, J., Roßmeier, C., Kornhuber, J., Schneider, A., Volk, A. E., … Otto, M. (2017). Neurofilament as a blood marker for diagnosis and monitoring of primary progressive aphasias. *Neurology*, *88*(10), 961–969. https://doi.org/10.1212/WNL.0000000000003688

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Tellechea, P., Pujol, N., Esteve-Belloch, P., Echeveste, B., García-Eulate, M. R., Arbizu, J., &

Riverol, M. (2018). Early- and late-onset Alzheimer disease: Are they the same entity? *Neurología (English Edition)*, *33*(4), 244–253. https://doi.org/10.1016/j.nrleng.2015.08.009

Themistocleous, C., Ficek, B., Webster, K., den Ouden, D.-B., Hillis, A. E., & Tsapkini, K. (2021). Automatic Subtyping of Individuals with Primary Progressive Aphasia. *Journal of Alzheimer's Disease*, *79*(3), 1185–1194. https://doi.org/10.3233/JAD-201101

Thurstone, L. L. (1973). Primary Mental Abilities. In *The Measurement of Intelligence* (pp. 131–136). Springer Netherlands. https://doi.org/10.1007/978-94-011-6129-9_8

Tippett, D. C. (2020). Classification of primary progressive aphasia: challenges and complexities. *F1000Research*, *9*, 64. https://doi.org/10.12688/f1000research.21184.1

Tombaugh, T. N., & McIntyre, N. J. (1992). The Mini-Mental State Examination: A Comprehensive Review. *Journal of the American Geriatrics Society*, *40*(9), 922–935. https://doi.org/10.1111/j.1532-5415.1992.tb01992.x

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Torso, M., Bozzali, M., Cercignani, M., Jenkinson, M., & Chance, S. A. (2020). Using diffusion tensor imaging to detect cortical changes in fronto-temporal dementia subtypes. *Scientific Reports*, *10*(1), 11237. https://doi.org/10.1038/s41598-020-68118-8

Tran, C. T., Zhang, M., Andreae, P., & Xue, B. (2017). Multiple imputation and genetic programming for classification with incomplete data. *Proceedings of the Genetic and Evolutionary Computation Conference*, 521–528. https://doi.org/10.1145/3071178.3071181

Utianski, R. L., Botha, H., Martin, P. R., Schwarz, C. G., Duffy, J. R., Clark, H. M., Machulda, M. M., Butts, A. M., Lowe, V. J., Jack, C. R., Senjem, M. L., Spychalla, A. J., Whitwell, J. L., & Josephs, K. A. (2019). Clinical and neuroimaging characteristics of clinically unclassifiable primary progressive aphasia. *Brain and Language*, *197*(September 2011), 104676. https://doi.org/10.1016/j.bandl.2019.104676

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 & Therapy*, *11*(1), 30. https://doi.org/10.1186/s13195-019-0483-2

van't Hooft, J. J., Pijnenburg, Y. A. L., Sikkes, S. A. M., Scheltens, P., Spikman, J. M., Jaschke, A. C., Warren, J. D., & Tijms, B. M. (2021). Frontotemporal dementia, music perception and social cognition share neurobiological circuits: A meta-analysis. *Brain and Cognition*,

148, 105660. https://doi.org/10.1016/j.bandc.2020.105660

van Buuren, S. (2018). *Flexible Imputation of Missing Data, Second Edition*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429492259

van der Ende, E. L., & van Swieten, J. C. (2021). *Fluid Biomarkers of Frontotemporal Lobar Degeneration* (pp. 123–139). https://doi.org/10.1007/978-3-030-51140-1_9

van der Vlies, A. E., Verwey, N. A., Bouwman, F. H., Blankenstein, M. A., Klein, M., Scheltens, P., & van der Flier, W. M. (2009). CSF biomarkers in relationship to cognitive profiles in Alzheimer disease. *Neurology*, *72*(12), 1056–1061. https://doi.org/10.1212/01.wnl.0000345014.48839.71

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

Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLOS Medicine*, *15*(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689

Vernooij, M. W., Pizzini, F. B., Schmidt, R., Smits, M., Yousry, T. A., Bargallo, N., Frisoni, G. B., Haller, S., & Barkhof, F. (2019). Dementia imaging in clinical practice: a European-wide survey of 193 centres and conclusions by the ESNR working group. *Neuroradiology*, *61*(6), 633–642. https://doi.org/10.1007/s00234-019-02188-y

Vijverberg, E. G. B., Dols, A., Krudop, W. A., Del Campo Milan, M., Kerssens, C. J., Gossink, F., Prins, N. D., Stek, M. L., Scheltens, P., Teunissen, C. E., & Pijnenburg, Y. A. L. (2017). Cerebrospinal fluid biomarker examination as a tool to discriminate behavioral variant frontotemporal dementia from primary psychiatric disorders. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *7*(1), 99–106. https://doi.org/10.1016/j.dadm.2017.01.009

von Luxburg, U., Williamson, R. C., & Guyon, I. (2012). Clustering: Science or Art? *JMLR: Workshop and Conference Proceedings*, *27*, 6579.

Walker, A. J., Meares, S., Sachdev, P. S., & Brodaty, H. (2005). The differentiation of mild frontotemporal dementia from Alzheimer's disease and healthy aging by neuropsychological tests. *International Psychogeriatrics*, *17*(1), 57–68. https://doi.org/10.1017/S1041610204000778

Wang, J., Redmond, S. J., Bertoux, M., Hodges, J. R., & Hornberger, M. (2016). A Comparison of Magnetic Resonance Imaging and Neuropsychological Examination in the Diagnostic

Distinction of Alzheimer's Disease and Behavioral Variant Frontotemporal Dementia. *Frontiers in Aging Neuroscience*, *8*(JUN), 1–10. https://doi.org/10.3389/fnagi.2016.00119

Warren, J. D., Rohrer, J. D., & Rossor, M. N. (2013). Frontotemporal dementia. *BMJ (Online)*, *347*(7920), 1–9. https://doi.org/10.1136/bmj.f4827

Watson, C. L., Possin, K., Allen, I. E., Hubbard, H. I., Meyer, M., Welch, A. E., Rabinovici, G. D., Rosen, H., Rankin, K. P., Miller, Z., Santos-Santos, M. A., Kramer, J. H., Miller, B. L., & Gorno-Tempini, M. L. (2018). Visuospatial Functioning in the Primary Progressive Aphasias. *Journal of the International Neuropsychological Society*, *24*(3), 259–268. https://doi.org/10.1017/S1355617717000984

Weakley, A., Williams, J. A., Schmitter-Edgecombe, M., & Cook, D. J. (2015). Neuropsychological test selection for cognitive impairment classification: A machine learning approach. *Journal of Clinical and Experimental Neuropsychology*, *37*(9), 899–916. https://doi.org/10.1080/13803395.2015.1067290

Wechsler, D. A. (1987). Manual for the Wechsler Memory Scale-Revised. *New York: Psychological Corporation*.

Weder, N. D., Aziz, R., Wilkins, K., & Tampi, R. R. (2007). Frontotemporal Dementias: A Review. *Annals of General Psychiatry*, *6*(1), 15. https://doi.org/10.1186/1744-859X-6-15

Whitwell, J. L., Przybelski, S. A., Weigand, S. D., Ivnik, R. J., Vemuri, P., Gunter, J. L., Senjem, M. L., Shiung, M. M., Boeve, B. F., Knopman, D. S., Parisi, J. E., Dickson, D. W., Petersen, R. C., Jack, C. R., & Josephs, K. A. (2009). Distinct anatomical subtypes of the behavioural variant of frontotemporal dementia: a cluster analysis study. *Brain*, *132*(11), 2932–2946. https://doi.org/10.1093/brain/awp232

Wicklund, M. R., Duffy, J. R., Strand, E. A., Machulda, M. M., Whitwell, J. L., & Josephs, K. A. (2014). Quantitative application of the primary progressive aphasia consensus criteria. *Neurology*, *82*(13), 1119–1126. https://doi.org/10.1212/WNL.0000000000000261

Wiens, J., & Shenoy, E. S. (2018). Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. *Clinical Infectious Diseases*, *66*(1), 149–153. https://doi.org/10.1093/cid/cix731

Willmes, K., Poeck, K., Weniger, D., & Huber, W. (1983). Facet theory applied to the construction and validation of the Aachen Aphasia Test. *Brain and Language*, *18*(2), 259–276. https://doi.org/10.1016/0093-934X(83)90020-2

Woolley, Josh D., Khan, B. K., Murthy, N. K., Miller, B. L., & Rankin, K. P. (2011). The Diagnostic Challenge of Psychiatric Symptoms in Neurodegenerative Disease. *The Journal of Clinical Psychiatry*, *72*(02), 126–133.

https://doi.org/10.4088/JCP.10m06382oli

Woolley, Joshua D., Wilson, M. R., Hung, E., Gorno-Tempini, M. L., Miller, B. L., & Shim, J. (2007). Frontotemporal dementia and mania. *American Journal of Psychiatry*, *164*(12), 1811–1816. https://doi.org/10.1176/appi.ajp.2007.07061001

Xia, Y., Lu, S., Wen, L., Eberl, S., Fulham, M., & Feng, D. D. (2014). Automated identification of dementia using FDG-PET imaging. *BioMed Research International*, *2014*, 421743. https://doi.org/10.1155/2014/421743

Xie, S. X., Libon, D. J., Wang, X., Massimo, L., Moore, P., Vesely, L., Khan, A., Chatterjee, A., Coslett, H. B., Hurtig, H. I., Liang, T. W., & Grossman, M. (2010). Longitudinal patterns of semantic and episodic memory in frontotemporal lobar degeneration and Alzheimers disease. *Journal of the International Neuropsychological Society*, *16*(2), 278–286. https://doi.org/10.1017/S1355617709991317

Yanase, J., & Triantaphyllou, E. (2019). A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Systems with Applications*, *138*(July), 112821. https://doi.org/10.1016/j.eswa.2019.112821

Yeo, J. M., Lim, X., Khan, Z., & Pal, S. (2013). Systematic review of the diagnostic utility of SPECT imaging in dementia. *European Archives of Psychiatry and Clinical Neuroscience*, *263*(7), 539–552. https://doi.org/10.1007/s00406-013-0426-z

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1982). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research*, *17*(1), 37–49. https://doi.org/10.1016/0022-3956(82)90033-4

Young, A. L., Marinescu, R. V, Oxtoby, N. P., Bocchetta, M., Yong, K., Firth, N. C., Cash, D. M., Thomas, D. L., Dick, K. M., Cardoso, J., van Swieten, J., Borroni, B., Galimberti, D., Masellis, M., Tartaglia, M. C., Rowe, J. B., Graff, C., Tagliavini, F., Frisoni, G. B., … Alexander, D. C. (2018). Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nature Communications*, *9*(1), 4273. https://doi.org/10.1038/s41467-018-05892-0

Young, J. J., Lavakumar, M., Tampi, D., Balachandran, S., & Tampi, R. R. (2018). Frontotemporal dementia: latest evidence and clinical implications. *Therapeutic Advances in Psychopharmacology*, *8*(1), 33–48. https://doi.org/10.1177/2045125317739818

Zhou, K., & Yang, S. (2020). Effect of cluster size distribution on clustering: a comparative study of k-means and fuzzy c-means clustering. *Pattern Analysis and Applications*, *23*(1), 455–466. https://doi.org/10.1007/s10044-019-00783-6

Zhutovsky, P., Vijverberg, E. G. B., Bruin, W. B., Thomas, R. M., Wattjes, M. P., Pijnenburg, Y. A. L., van Wingen, G. A., & Dols, A. (2019). Individual Prediction of Behavioral Variant Frontotemporal Dementia Development Using Multivariate Pattern Analysis of Magnetic Resonance Imaging Data. *Journal of Alzheimer's Disease : JAD*, *68*(3), 1229–1241. https://doi.org/10.3233/JAD-181004

Ziegler, S., Maier, C., & Reichenbach, A. (2020). Stratification of patients with Alzheimer's disease based on longitudinal neuropsychological tests. *2020 IEEE International Conference on Healthcare Informatics (ICHI)*, 1–7. https://doi.org/10.1109/ICHI48887.2020.9374343

Zimmerer, V. C., Hardy, C. J. D. D., Eastman, J., Dutta, S., Varnet, L., Bond, R. L., Russell, L., Rohrer, J. D., Warren, J. D., & Varley, R. A. (2020). Automated profiling of spontaneous speech in primary progressive aphasia and behavioral-variant frontotemporal dementia: An approach based on usage-frequency. *Cortex*, *133*, 103–119. https://doi.org/10.1016/j.cortex.2020.08.027

Appendix

The complete code of the analysis can be found in the following github repository: https://github.com/MSontgerath/Masterthesis_FTLD_Clustering

All results of the analysis can be downloaded until 29th September 2022 from the following link: https://we.tl/t-yv7p0Pet6Q. In case this link should not be active anymore, please send an e-mail to the following e-mail address for further information: marie.soentgerath@studenti.unipd.it.