Received: 6 March 2024 | Revised: 26 March 2024 | Accepted: 18 May 2024

DOI: 10.1002/pcn5.210

PCN Reports
Psychiatry and Clinical Neurosciences
Open Access

REVIEW ARTICLE

A systematic review of cognitive and behavioral tools to differentiate behavioral variant frontotemporal dementia from other conditions

Joshua Flavell MBBS[1,2,3] | Peter John Nestor PhD[1,2]

[1]The Queensland Brain Institute, The University of Queensland, Brisbane, Australia

[2]The Mater Hospital, Brisbane, Australia

[3]Metro North Hospital and Health Service, Brisbane, Australia

Correspondence
Joshua Flavell, MBBS, The Queensland Brain Institute, QBI Bldg, 79, The University of Queensland, St Lucia, QLD 4072, Australia.
Email: j.flavell@uq.edu.au

Abstract

The behavioral variant of frontotemporal dementia (bvFTD) is thought to be the commonest clinical presentation of frontotemporal lobar degeneration and is predominantly characterized by changes in behavior. In patients lacking unequivocal biomarker evidence of frontotemporal neurodegeneration, the clinical diagnosis of bvFTD is often unstable. In response, we conducted a systematic review and critical appraisal of cognitive and behavioral tools that have sought to differentiate bvFTD from other conditions. A systematic literature review of PubMed, Scopus, and Web of Science was conducted on December 31, 2023 for cognitive and behavioral tools that differentiated bvFTD from other cohorts. Ninety-six studies were included. The quality appraisal of almost all studies was low and introduced a high risk of bias. The few studies that were of high quality had a prospective study design and recruited patients suspected (but not yet confirmed) to have bvFTD. These studies reported that behavioral tools (e.g., the Frontal Behavioral Inventory) and social cognition tests (e.g., the Ekman's Faces Test) had good test performance in differentiating bvFTD from a broad range of psychiatric and neurological conditions. Importantly, the review highlighted the extreme paucity of studies that have evaluated methods where, in Bayesian terms, there is genuine clinical uncertainty regarding a diagnosis of bvFTD. Most studies used healthy controls of typical Alzheimer's disease as comparators—groups that often have negligible pretest probability of bvFTD. In response, we propose a study design checklist for studies seeking to develop diagnostic algorithms in bvFTD research.

KEYWORDS

cognition, diagnostic, frontotemporal dementias, review, tests

INTRODUCTION

The behavioral variant of frontotemporal dementia (bvFTD) is the most common clinical presentation of frontotemporal lobar degeneration (FTLD).[1] It is characterized predominantly by early changes in behavior with cognitive decline emerging later in the disease course.[2] These behavioral and cognitive symptoms often overlap with psychiatric disorders.[3,4] Accordingly, the diagnosis of bvFTD can be challenging, particularly in the absence of unequivocal biomarker evidence of frontotemporal neurodegeneration. Though the current Frontotemporal Dementia Consortium (FTDC) criteria[5] are sensitive, the specificity of a *possible* bvFTD diagnosis can be as low as 27% without supportive biomarker evidence of neurodegeneration.[6]

Difficulty in making an accurate bvFTD diagnosis has been highlighted throughout the literature. Studies reported that between 50% and 70% of people are initially misdiagnosed with a psychiatric disorder,[7,8] and that potentially many years then pass before bvFTD is diagnosed; for instance, one study reported that FTD had the longest delay in diagnosis compared to other forms of dementia with an average latency of 6.1 years from symptom onset.[9] Diagnostic stability is not guaranteed at the point of a bvFTD diagnosis either. At follow up, patients often have their diagnosis changed to a diverse range of conditions, including other dementias, neurological diseases, and primary psychiatric disorders.[10] There is also a group of patients who ultimately are not diagnosed with any clinical disorder[11] that are considered "phenocopies" of bvFTD.[12] The difficulties of accurately diagnosing bvFTD is as applicable to research studies as it is to clinical practice. Instability of diagnosis may adversely affect conclusions in research studies through inclusion of misdiagnosed participants.

In response to these diagnostic challenges, numerous clinical tools employing cognitive, behavioral, and other clinical assessments have been proposed. There is no consensus, however, on which tools best differentiate bvFTD from other conditions. This review addressed this issue by systematically evaluating the literature to provide an overview of the strengths and weaknesses of existing tools, while identifying areas for future research. Furthermore, as with other degenerative dementias, bvFTD is a progressive disease where the diagnostic accuracy of tools may vary according to the disease stage at which they were applied. This review, therefore, also systematically evaluated the diagnostic process and staging used in the studies that were identified.

METHODS

This systematic review was conducted in accordance with the Preferred Reporting Items for a Systematic Review and Meta-Analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) statement.[13] This review was not registered prospectively.

Eligibility criteria

To be included in this review, studies had to meet the following eligibility criteria.

Inclusion criteria:

1. English-language papers reporting on the diagnostic accuracy of clinical tools (e.g., clinical inventories, scales, questionnaires, algorithms, cognitive tests, and batteries, or multimodular cognitive and behavioral instruments) that differentiated bvFTD from other cohorts, including healthy controls.
2. Studies that provided test performance of the evaluated tools with test accuracy, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), negative likelihood ratio (NLR), sensitivity, specificity, or receiver operating characteristic area under the curve (AUC).
3. Observational studies with a minimum sample size of $n = 5$ bvFTD.
4. Studies that included patients diagnosed with bvFTD according to established research or clinical criteria (e.g., the FTDC criteria or the *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edition [DSM-5]). Note: while the included studies stated that study participants met diagnostic criteria, it was noted that many studies did not explain *how* these criteria were met (e.g., whether on clinical grounds alone or if imaging was included).

Exclusion criteria:

1. Studies that were unpublished, inaccessible, or incomplete.
2. Case reports, interventional studies, literature reviews, conference proceedings, book chapters, guidelines, and research protocols.
3. Papers that did not describe the method by which bvFTD was diagnosed.
4. Papers not in English.
5. Duplicate studies.
6. Studies that did not include a comparator and/or lacked outcomes of interest (e.g., did not report test performance).

Information source

The systematic literature review was conducted, without time restriction, until December 31, 2023 by performing an online search in the following databases: PubMed, Scopus, and Web of Science.

Search Strategy

The search strategy included the following terms: ((frontotemporal dementia) OR (behavioral variant frontotemporal dementia) OR

(bvFTD) OR (FTD)) AND ((tool) OR (test) OR (screen*) OR (question*) OR (diagnosis) OR (scale) OR (instrument)).

Study selection

After deleting duplicates, the title and/or abstract of each study were reviewed for reference to clinical tools for bvFTD. If no abstract was available, the full text was reviewed in the first instance. The full text of potentially relevant studies was evaluated to determine if they met the eligibility criteria. Finally, the references of the resulting full texts were searched for further relevant citations.

Data extraction

Data extracted from the identified articles included information on study design (eligibility criteria, sample size, clinical tool); information relevant to the quality appraisal; participant characteristics (diagnosis, sample size); clinical investigations (neuroimaging, genetic testing, pathological assessments); diagnostic criteria applied; and test performance (test accuracy, PPV, NPV, PLR, NLR, sensitivity, specificity, or AUC). To indicate which cohort had the greater impairment on each test, "<" and ">" symbols were used. For this review an AUC of 0.5–0.6 was described as a failure, 0.6–0.7 as poor, 0.7–0.8 as average, 0.8–0.9 as good, and 0.9–1.0 as excellent.

Quality assessment

The Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2)[14] was used to evaluate methodological quality. This tool has four sections and 11 criteria that review the methodological quality in diagnostic accuracy studies. Risk of bias is described for each of the sections: "patient selection," "index test," "reference standard," and "flow and timing." Each study was given a score for each criterion, a total score of 0–11, and a risk of bias of "low," "high," or "unclear" for each section.

RESULTS

Study selection

The systematic literature search generated 19,693 articles. After the removal of duplicate articles, 12,715 articles remained. The titles and abstracts were reviewed, with 633 articles identified that potentially met inclusion criteria. These articles were then reviewed in full, yielding 96 articles that met eligibility criteria[15–110] (Figure 1).

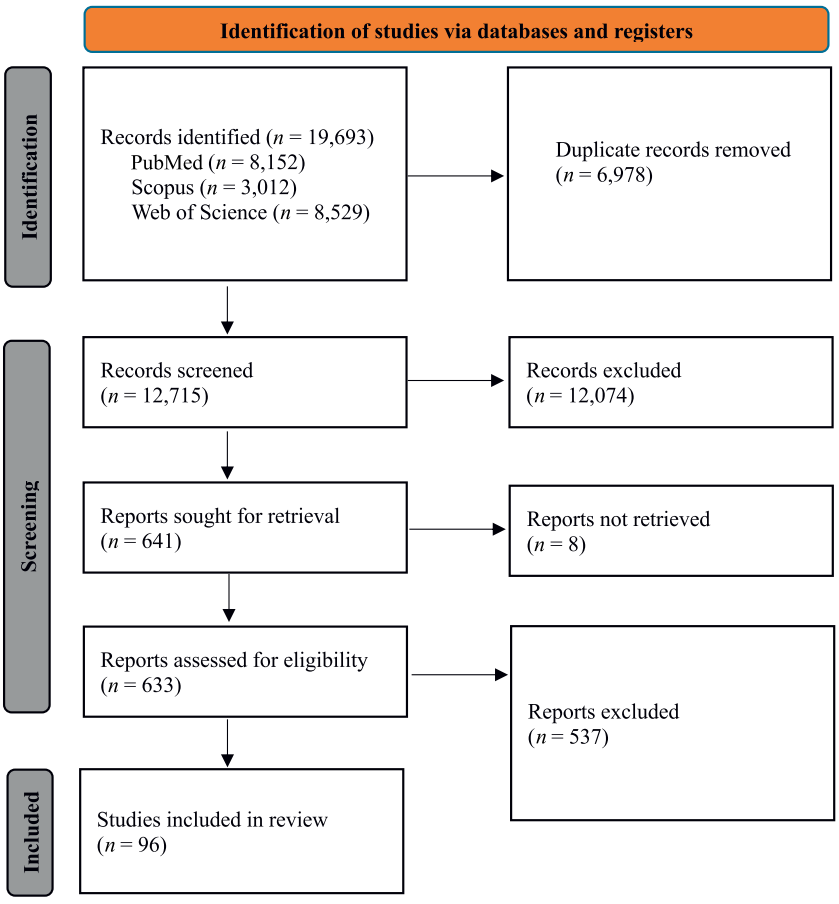

FIGURE 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram. This figure depicts the PRISMA flow diagram of the studies identified, screened, and included.

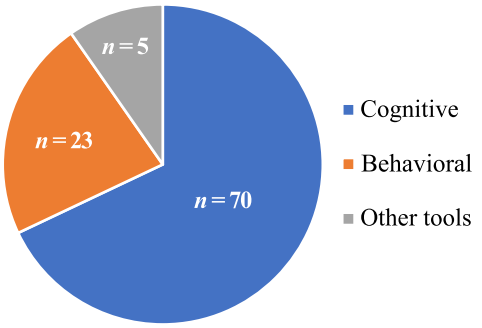

FIGURE 2 Clinical tools. This figure depicts the number of cognitive, behavioral, and other clinical tools identified by this review.

Clinical tools

From the studies included in the review, 98 clinical tools were used to differentiate bvFTD from other conditions and healthy controls, including cognitive tests ($n = 70$), behavioral tools ($n = 23$), and other clinical tools ($n = 5$) (see Supporting Information S1) (Figure 2).

Study characteristics

The studies had retrospective ($n = 8$), prospective ($n = 25$), cross-sectional ($n = 62$), and mixed ($n = 1$) study designs. The comparator cohorts included healthy controls ($n = 25$), Alzheimer's disease (AD) ($n = 59$), or other neurodegenerative disorders ($n = 12$), including progressive supranuclear palsy (PSP); corticobasal syndrome (CBS); primary progressive aphasia (PPA); vascular dementia (VascD); posterior cortical atrophy (PCA); Parkinson's disease (PD); mood disorders ($n = 9$), including major depressive disorder (MDD) and bipolar I disorder (BDI); mixed neurological cohorts ($n = 13$); mixed psychiatric cohorts ($n = 3$); and a combination of both ($n = 5$) (Table 1).

Only five studies described the prospective application of clinical tools to people suspected to have bvFTD.[35,51,53,65,101] Four of these five studies[35,51,65,101] reported data from the Late Onset Frontal Lobe Syndrome (LOF) study.[111] The test results for bvFTD compared to the main comparator cohorts—healthy controls, AD, psychiatric disorders, and mix cohorts—are summarized in Figure 3.

Quality of studies

The QUADAS-2 quality appraisal can be found in the Supporting Information S1: Tables 2 and 3. Only five studies had a high-quality study design with low risk of bias.[35,51,53,65,101] These were the same studies that prospectively applied clinical tools to participants with *suspected* bvFTD in whom the diagnosis had not yet been confirmed (appropriate sampling). The final diagnoses of these studies included a broad range of neurological and psychiatric disorders, likely representing a comprehensive list of the possible diagnostic outcomes for people who develop frontal behavioral change later in life.

All other studies ($n = 91$) were rated as low-quality level of evidence, with an average QUADAS-2 total score of 4.2 ± 1.6 (mean ± standard deviation). The risk of bias for patient selection was high for most studies. For patient sampling, nearly all studies enrolled patients who already had a confirmed diagnosis of bvFTD (appropriate sampling criterion not met). The risk of bias for the index test was high for most studies. Most studies did not state if there was blinding to the diagnosis of bvFTD or index test administration and results, nor did they use a prespecified test cut-off. The risk of bias for the reference test was either high or unclear for most studies. It was unclear for many studies if the diagnosis of bvFTD was supported by neuroimaging or histopathological findings. Finally, the risk of bias for flow and timing was also high for most studies. Most studies included all participants in the results and did not have inappropriate exclusions; however, appropriate timing and universal application of the reference test generally did not occur (Figure 4).

Diagnostic evaluation

The diagnostic work-up for the participants diagnosed with bvFTD can be found in Supporting Information S1: Table 1. The reference standard, diagnostic criteria used to define bvFTD, included the clinical criteria of the FTDC[5] ($n = 53$); Lund and Manchester[112] ($n = 18$); Neary criteria[113] ($n = 26$); the DSM-5[114] ($n = 2$); the Work Group on FTD and Pick's Disease clinical criteria[115] ($n = 1$); and the pathological criteria of Jackson and Lowe[116] ($n = 1$) and Wallin and Brun[117] ($n = 2$). Six studies used multiple diagnostic criteria.[59,62,67,68,90,93]

The diagnosis of bvFTD varied from clinical criteria alone (possible bvFTD) ($n = 2$); clinical criteria with supportive neuroimaging (probable bvFTD) ($n = 41$); clinical criteria with supportive histopathology or genetics (definite bvFTD) ($n = 3$); mixed cohort of possible and probable bvFTD ($n = 8$); mixed cohort of both probable and definite bvFTD ($n = 10$); or it was not specified how a diagnosis was made ($n = 32$). A minority of studies included a follow-up period to ensure stability in bvFTD diagnosis ($n = 23$). For these studies, the follow-up period ranged from 0.5 to 2 years, and about half of these studies had a prospective follow-up period of at least 2 years ($n = 12$) (see Supporting Information S1: Table 1).

Disease severity

Thirty-nine studies (Table 1) described the disease severity of bvFTD and the remaining studies did not. Five studies described the disease severity of the comparator cohort, but not in the cohort with bvFTD.[21,32,50,64,70] When mentioned, studies described bvFTD disease severity as "mild" ($n = 24$) or "mild to moderate" ($n = 15$).

There was considerable variation in the methods used to define disease severity for bvFTD. Eleven studies did not specify how

TABLE 1 Study characteristics.

Study design			Comparator cohort			Disease severity			Rating of severity			Clinical tools		
Cross-sectional cohort	*n* = 62	(i)	Healthy controls	*n* = 25	(v)	Mild	*n* = 24	(xii)	Nil method used	*n* = 11	(xv)	Cognitive tools	*n* = 61	(xx)
Prospective cohort	*n* = 25	(ii)	Alzheimer's disease	*n* = 59	(vi)	Mild to moderate	*n* = 15	(xiii)	CDR or CDR-FTLD	*n* = 18	(xvi)	Behavioral tools	*n* = 30	(xxi)
Retrospective cohort	*n* = 8	(iii)	Mood disorders	*n* = 9	(vii)	Not described	*n* = 57	(xiv)	MMSE	*n* = 4	(xvii)	Miscellaneous tools	*n* = 5	(xxii)
Mixed study design	*n* = 1	(iv)	Other neurodegenerative disorders	*n* = 12	(viii)				Symptom duration	*n* = 3	(xviii)	Combination of tools	*n* = 14	(xxiii)
			Mixed neurological cohort	*n* = 13	(ix)				Other method used	*n* = 3	(xix)			
			Mixed psychiatric cohort	*n* = 3	(x)									
			Mixed neurological and psychiatric cohort	*n* = 5	(xi)									

References

(i)[15–20,25–28,30,36–39,41–44,46,48–50,52,54,55,57–59,62–64,67–70,72–76,78–80,83,87–89,91,95–100,102,103,105–108,110]
(ii)[21,24,29,32–35,40,45,47,51,53,56,61,65,71,77,81,82,84,86,90,93,94,101]
(iii)[22,31,60,66,85,92,104,109]
(iv)[23]
(v)[17,18,21,22,25,37,39,41,44,46,47,52,57,58,60,64,66,67,69,74–76,83,95,97]
(vi)[18,19,21,23–32,36,38,40,42,45,47–49,52,54,55,57–61,63,66,68,70,71,73,75,77–81,84–87,89–92,94–96,98,102,104–106,108,109]
(vii)[33,39,50,56,77,84,96,99,110]
(viii)[20,43,45,47,54,66,75,78,80,84,87,103]
(ix)[15,16,34,45,47,62,72,75,78,88,93,100,107]
(x)[53,65,101]
(xi)[35,51,53,82,84]
(xii)[17,19,25,27,29,30,40,41,43,46,58–61,67–69,72,73,77,86,90,101,110]
(xiii)[18,20,24,31,33,47,62,76,83,94,103–105,107,108]
(xiv)[15,16,21–23,26,28,32,34–39,42,44,45,48–57,63–66,70,71,74,75,78–82,84,85,87–89,91–93,95–100,102,106,109]
(xv)[24,27,40,47,67,69,86,94,101,103,104]
(xvi)[17,20,25,29–31,41,46,58,59,72,73,76,83,105,107,108,110]
(xvii)[43,62,68,90]
(xviii)[19,60,61]
(xix)[18,33,77]
(xx)[15–44,46–76]
(xxi)[22–24,35,37,40,45,77–98]
(xxii)[37,99–102]
(xxiii)[17,29,31,40,54,63,103–110]

Abbreviations: CDR, Clinical Dementia Rating scale; CDR-FTLD, CDR frontotemporal lobar degeneration - (Modified Clinical Dementia Rating scale); MMSE, Mini-Mental State Examination.

disease severity was defined, only stating it was "mild" or "moderate." Of the studies that did use a standardized method, $n = 17$ defined severity with the Clinical Dementia Rating Scale (CDR)[118] and $n = 1$ with the Frontotemporal Lobar Degeneration-modified CDR (CDR-FTLD).[119] Four studies used the Mini-Mental State Examination (MMSE)[120]; however, conflicting definitions of disease severity were offered. Two studies defined mild bvFTD with an MMSE score of >19/30.[43,68] Another study used a cut-off score of >22/30 for mild bvFTD.[90] The final study defined "mild to moderate" bvFTD with an MMSE score of >18/30,[62] which is only one point difference from the definition used for mild bvFTD by others. Three studies used symptom duration. Other methods used were the DSM-5 neurocognitive severity, which is based on impairment in activities of daily living ($n = 1$)[77]; the Mattis Dementia Rating Scale ($n = 1$)[33]; and the Dementia Rating Scale-2 ($n = 1$).[18]

Test performance

Results from high-quality studies

The LOF study[111] produced four QUADAS-2 rated high-quality papers that assessed the diagnostic accuracy of various clinical tools, including neuropsychological tests of language and social cognition, cognitive test batteries, behavioral measures, and other clinical tools, such as a clinical checklist.

The first paper from the LOF study,[35] published in 2015, assessed the test performance of the Stereotypy Rating Inventory (SRI),[121] Frontal Behavioral Inventory (FBI),[122] Frontal Assessment Battery (FAB),[123] and MMSE in a cohort of $n = 55$ bvFTD ($n = 10$ possible and $n = 45$ probable bvFTD) compared to $n = 82$ patients who also presented with late-onset frontal behavioral symptoms but had a final diagnosis of a psychiatric ($n = 51$) or another neurological disorder ($n = 31$). This study reported superior performance of the SRI (AUC = 0.73) and FBI (AUC = 0.68) compared to the FAB (AUC = 0.62) and MMSE (AUC = 0.61). The next study[51] investigated social cognition using the 20-item version of the Faux Pas Test (FAUX)[124] and 60-item version of the Ekman's Faces Test (EFT)[125] to differentiate $n = 22$ bvFTD ($n = 18$ probable and $n = 4$ definite bvFTD) from $n = 57$ patients with either mixed psychiatric ($n = 33$), and other neurodegenerative ($n = 24$), disorders. This study reported that the EFT-60 (AUC = 0.73) outperformed the FAUX-20 (AUC = 0.60). The optimal cut-off for the EFT-60 to differentiate probable/definite bvFTD from this mixed neuropsychiatric cohort was ≤34.5/60, with a sensitivity of 66.7% and specificity of 68.2%. The third study[101] applied the Frontotemporal Dementia versus Primary Psychiatric Disorder (FTDvsPPD) checklist to a prospective cohort of patients presenting with late-onset behavioral change at the Montreal Neurological Institute neuropsychiatry clinic ($n = 20$) and a retrospective analysis of data from the LOF study cohort ($n = 92$). The final sample included $n = 46$ probable bvFTD and $n = 66$ psychiatric disorders. The 17-item version of the FTDvsPPD checklist was reported to have excellent test performance, with an AUC of 0.91. The optimal cut-off for the FTDvsPPD checklist to differentiate probable bvFTD from psychiatric disorders was ≥11/17, with a sensitivity of 93.9% and specificity of 71.1%. The final study[65] applied the 29-item version of the Boston Naming Test (BNT)[126] to differentiate $n = 32$ bvFTD ($n = 28$ probable and $n = 4$ definite) from $n = 53$ with psychiatric disorders. The BNT had good test performance, with an AUC of 0.81. The optimal cut off for the BNT to differentiate probable/definite bvFTD from psychiatric disorders was <72/87, using their scoring system, with a sensitivity of 92.0% and specificity of 64.0%.

The other high-quality study[53] according to QUADAS-2 rating prospectively recruited participants suspected to have bvFTD at Lille University Hospital's memory clinic between October 2006 and February 2014. This study applied the FAUX-20 to differentiate $n = 12$ probable/definite bvFTD from $n = 22$ psychiatric and other neurological disorders. All participants had at least 3 years follow up to ensure diagnostic stability. This study used a z-score from $n = 165$ healthy controls to define the test cut-off for impairment. Based on these results, the FAUX-20 had a sensitivity of 83.0% and specificity of 64.0% to differentiate probable/definite bvFTD from a mixed neuropsychiatric cohort. It is important to note that this study used a broader scoring system, ranging from 0 to 80, that incorporated not only the ability to detect the faux pas but also the explanation of the faux pas. This contrasts to the LOF study,[51] which only scored the faux pas detection, with a 0–10 score range. Accordingly, the study from Lille concluded that the FAUX-20 is more sensitive to detect bvFTD with a broader scoring system.

Results from low-quality studies

The remaining 91 studies were rated low quality according to the QUADAS-2 appraisal.[15–34,36–50,52,54–64,66–100,102–110,127–129] These studies assessed (a) neuropsychological tests, (b) cognitive batteries, (c) behavioral tools, and (d) other clinical tools. Some of these studies also (e) combined different tools in the same cohort.

Neuropsychological tests

The low-quality-rated studies included neuropsychological tests of (i) executive function, (ii) social cognition, (iii) memory, (iv) attention and orientation, (v) language, (vi) visuospatial abilities, and (vii) praxis.

Executive function (Table 2). Thirty studies of low-quality according to the QUADAS-2 appraisal[15–34,36–45] used executive function tests to differentiate bvFTD from healthy controls and other conditions, including AD, PSP, PPA, MDD, and mixed neurological and/or psychiatric cohorts. The prevailing theme of the executive function tests was that most results were based on single study findings; that is, the test performance of each tool in differentiating bvFTD from a specific cohort was supported by only one study. When multiple studies assessed the same test in similar cohorts (e.g., bvFTD vs healthy controls), the results were often inconsistent.

TABLE 2 Executive function tests.

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/ Max score
	Applause sign											
[15]	Total score	bvFTD < PSP/AD	15 vs 39	60.0	56.4	34.6						X/3.0
[16]	Total score	bvFTD < Mx dementia cohort	111 vs 161	09.0	78.9	22.7						<3.0/3.0
	Backward Digit Span Test (BDST)											
[17]	Total score	bvFTD < HC	35 vs 14					0.76				N/A
	Behavioral Dyscontrol Scale (BDS)											
[18]	Total score	bvFTD > HC	21 vs 21	85.7	95.2	94.7	87.0	0.97				≥16.0/19.0
[18]	Total score	bvFTD > AD	21 vs 21	81.0	81.0	81.0	81.0	0.84				≥15.0/19.0
	Brixton Test (BT)											
[19]	Total errors score	bvFTD > AD	11 vs 10					0.57				N/A
[20]	Total errors—scaled score	bvFTD > SvPPA	76 vs 34	53.0	94.0							≤3/10.0
	Ecological Intertemporal Choices Task (EICT)											
[21]	Delay-discounting score	bvFTD < HC	20 vs 20					0.90			82.1	N/A
[21]	Delay-discounting score	bvFTD < AD	20 vs 30					0.70			75.5	N/A
[21]	Delay-discounting score	bvFTD < AD/HC	20 vs 50					0.79				N/A
	Edinburgh Cognitive and Behavioral ALS Screen (ECAS)											
[22]	Executive subscore	bvFTD < HC	16 vs 48	81.3	95.8							≤33.0/48.0
	Executive Interview (EXIT-25)											
[23]	Total score	bvFTD < AD	13 vs 22	84.4	68.8			0.83				<26.0/50.0
[23]	Total score	bvFTD < AD	13 vs 22	87.3	65.8							<24.5/50.0
	Five Digits Test (FDT)											
[24]	Flexibility subscore	bvFTD < AD	27 vs 25	86.4	76.5			0.83				54 s/X
	Frontal Assessment Battery (FAB)											
[25]	Total score	bvFTD < HC	18 vs 15	66.7	66.7			0.72				≤15.5/18.0
[25]	Total score	bvFTD < AD	18 vs 20	55.0	50.0			0.54				≤13.5/18.0
[26]	Total score	bvFTD < AD	26 vs 64	81.0	72.0			0.81				≤11.0/18.0
[26]	Total score	bvFTD (mild) < AD	9 vs 24	77.0	87.0			0.81				≤12.0/18.0
[27]	Total score	bvFTD < AD	18 vs 18	85.0	92.0				12.0			≤10.0/18.0
[28]	Total score	bvFTD < AD	34 vs 25	82.3	48.5			0.73				≤13.0/18.0
[29]	Chinese version—Total score	bvFTD < AD	22 vs 26					0.69				N/A
[19]	Total score	bvFTD < AD	11 vs 10					0.69				N/A
[30]	Total score	bvFTD < AD	25 vs 25	16.0	96.2			0.50				≤10.0/18.0
[31]	Resistance to interference subscore	bvFTD < AD	35 vs 46	97.0	100.0			0.98				X/3.0
[32]	Go/No-go subscore	bvFTD < AD	20 vs 20					0.74			62.5	N/A
[33]	Total score	bvFTD < MDD	37 vs 19					0.71				N/A
[33]	Total score	bvFTD (mod) < MDD	20 vs 19	64.7	52.6			0.57				≤16.0/18.0

(Continues)

TABLE 2 (Continued)

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/ Max score
[34]	Total score	bvFTD < Mx dementia cohort/HC	45 vs 29	94.0	55.0	0.54	0.94	0.70	2.09	0.11	69.0	<12.0/18.0
[35]	Total score	bvFTD < Mx PPD/ Neuro cohort	55 vs 82					0.62				N/A
	FRONTIER Executive Screen (FES)											
[36]	Total score	bvFTD < AD	14 vs 14	86.0	50.0			0.84				≤8.0/15.0
[36]	Total score	bvFTD < AD	14 vs 14	71.0	73.0			0.84				≤7.0/15.0
	Hamasch Five-Point Test (H5PT)											
[37]	Total score	bvFTD < HC	86 vs 43					0.81				N/A
	Hayling Test (HT)											
[19]	Total score	bvFTD < AD	11 vs 10					0.76				N/A
	INECO Frontal Screening (IFS)											
[25]	Total score	bvFTD < HC	18 vs 15	73.3	61.1			0.71				≤20.3/30.0
[38]	Total score	bvFTD < HC	22 vs 26	100	88.0							≤26.0/30.0
[39]	Total score	bvFTD < HC	49 vs 26					0.97				N/A
[25]	Total score	bvFTD < AD	18 vs 20	66.7	60.0			0.59				≤16.7/30.0
[28]	Total score	bvFTD < AD	34 vs 25	94.1	94.2			0.98				≤17.5/30.0
[40]	Total score	bvFTD < AD	18 vs 33	75.8	66.7			0.78	2.27	0.36		≤19.0/30.0
[38]	Total score	bvFTD < AD	22 vs 25	72.0	81.3							≤19.0/30.0
[30]	Total score	bvFTD < AD	25 vs 25	67.7	92.0			0.77				≤21.0/30.0
[39]	Total score	bvFTD < MDD	49 vs 30					0.84				N/A
	Iowa Gambling Task (IGT)											
[17]	Total score	bvFTD < HC	35 vs 14					0.91				N/A
[32]	Total score	bvFTD < AD	20 vs 20					0.67			57.1	N/A
	Letters and Numbers Sequencing Test (LNST)											
[41]	Total score	bvFTD < HC	35 vs 14					0.88				N/A
	Modified Hotel Task (HOT)											
[17]	Total score	bvFTD < HC	35 vs 14					0.80				N/A
	Multiple Errands Test Hospital Version (MET-HV)											
[17]	Rule breaks score	bvFTD > HC	35 vs 14					0.78				N/A
	Novel Verbal Similarity Task (SimiCat)											
[42]	Total differentiation score	bvFTD < AD	40 vs 23	90.0	87.0			0.94				≥1.0/16.0
[42]	Total differentiation score	bvFTD < AD	40 vs 23	80.0	100.0			0.94				≥2.0/16.0
	Phonemic Fluency (PF)											
[17]	P-words—Total score	bvFTD < HC	35 vs 14					0.68				N/A
[37]	S-words—Total score	bvFTD < HC	86 vs 43					0.94				N/A
[22]	P-words + S-words (ECAS fluency subscore)	bvFTD < HC	16 vs 48	87.5	93.7							≤14.0/24.0
[38]	S-words—Total score	bvFTD < AD	22 vs 25					0.49				N/A

TABLE 2 (Continued)

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/ Max score
	Reversal-Learning Test (RLT)											
[32]	Rules reversed score	bvFTD < AD	20 vs 20					0.78			69.7	N/A
[32]	Errors score	bvFTD < AD	20 vs 20					0.54				N/A
	Stroop task (ST)											
[37]	Total score	bvFTD < HC	86 vs 43					0.70				N/A
[43]	Self-corrections subscore	bvFTD < PSP	27 vs 25	82.0	68.0			0.79				<6.0/30.0
	The Test of Practical Judgment (TOP-J)											
[44]	Brazilian version—Total score	bvFTD < HC	15 vs 24	83.3	73.3			0.94				≤18.0/45.0
	Trail-Making Test (TMT)											
[37]	TMT B:A ratio	bvFTD > HC	86 vs 43					0.71				N/A
[17]	TMT–B score (s)	bvFTD > HC	35 vs 14					0.81				N/A
[38]	TMT–B score (s)	bvFTD > AD	22 vs 25					0.46				N/A
[45]	TMT–B score (s)	bvFTD < AD	406 vs 58					0.64				N/A
[45]	TMT–B score (s)	bvFTD < SvPPA	406 vs 61					0.38				N/A
[45]	TMT–B score (s)	bvFTD < Mx dementia cohort	406 vs 385					0.57				N/A
	Tower of London Test (TOL)											
[31]	Two-movement paradigm—Movement number	bvFTD < AD	13 vs 39					0.80				X/X
[31]	Five-movement paradigm—Total time	bvFTD < AD	13 vs 39					0.81				X/X
	Wisconsin Card Sorting Test (WCST)											
[17]	Total score	bvFTD < HC	35 vs 14					0.84				N/A
[38]	Total score	bvFTD < AD	22 vs 25					0.62				N/A
[33]	Total score	bvFTD < MDD	37 vs 19					0.66				N/A
[33]	Total score	bvFTD (mod) < MDD	20 vs 19	17.6	100.0			0.55				<11.0/X

Note: See Supporting Information S1 for test references.

Abbreviations: Acc., accuracy; AD, Alzheimer's disease; AUC, area under the curve; bvFTD, behavioral variant frontotemporal dementia; HC, healthy controls; MDD, major depressive disorder; Mx, mixed; N/A, not applicable; Neuro, neurological; NPV, negative predictive value; NLR, negative likelihood ratio; PLR, positive likelihood ratio; PPD, primary psychiatric disorder; PPV, positive predictive value; PSP, progressive supranuclear disorder; Sens., sensitivity; Spec., specificity; svPPA, semantic variant primary progressive aphasia.

Executive function tests were generally reported as good to excellent at differentiating bvFTD from healthy controls. The most effective tools were the INECO Frontal Screening (IFS),[130] phonemic fluency, and the Test of Practical Judgment (TOP-J).[131] As mentioned, however, these results were either inconsistent or supported by single-study findings. The test performance of the IFS varied greatly across studies, with one study reporting excellent (AUC = 0.97)[39] and yet another reporting only average efficacy (AUC = 0.71).[25] This inconsistency was also observed for Phonemic Fluency, with varying levels of performance reported in different studies. In contrast, though the TOP-J also had excellent test performance, this was only reported by one study.[44]

Differentiating bvFTD from AD, there was considerable variation in performance of executive function tests. Although some tools were reported to have good to excellent test performance, these findings were generally based on single studies. For example, the Novel verbal similarity task (SimiCat)[42] and the FRONTIER Executive Screen[36] were reported to have good to excellent performance in individual studies.[36,42] In contrast, when the same tool was evaluated in multiple studies, inconsistent results were evident. For example,

the IFS had an excellent test performance in one study (AUC = 0.98),[28] average test performance in two other studies (AUC = 0.78–0.77),[30,40] and failed to differentiate bvFTD from AD in a final study (AUC = 0.59).[25] The FAB had good test performance in three studies (AUC = 0.98–0.73),[26,28,31] but performed poorly or failed in four other studies (AUC = 0.69–0.50).[19,25,29,30] The only exceptions were for phonemic fluency and the Trail Making Test—B (TMT-B),[132] which consistently performed poorly in differentiating bvFTD from AD in multiple studies.

Executive function tests also had considerable variation in performance in differentiating bvFTD from MDD.[33,39] These tests consistently performed poorly at differentiating bvFTD from semantic dementia[20,45] or a heterogenous dementia cohort.[16,34,45] There was average test performance in differentiating bvFTD from PSP.[43]

Social cognition (Table 3). Seventeen studies of low-quality according to the QUADAS-2 appraisal[17,19,29,32,33,37,40,46–50,52,54–57] used social cognition tests to differentiate bvFTD from healthy controls and other conditions, including AD, PSP, CBS, PPA, MDD, BDI, presymptomatic FTD genetic carriers and mixed neurological and/or psychiatric cohorts. Social cognition tests were consistently reported as good to excellent at differentiating bvFTD from healthy controls, presymptomatic FTD genetic carriers, AD, and MDD. In contrast, they were consistently average to poor at differentiating bvFTD from PPA, CBS, BDI, and mixed cohorts. In general, each of these results was supported by findings from single studies; however, when multiple studies assessed the same test in similar cohorts the results tended to be consistent.

The most effective social cognition tests were the EFT, FAUX, and the Mini-Social cognition and Emotional Assessment (Mini-SEA),[33] which is a combination of shorter versions of both the EFT and FAUX. These tools consistently demonstrated excellent performance in differentiating bvFTD from healthy controls. Their effectiveness varied, however, when applied to other conditions. The EFT was reported to have average to poor ability in differentiating bvFTD from all other conditions, including mixed cohorts. In contrast, the FAUX exhibited excellent to good performance in differentiating bvFTD from AD, but it was less effective for a mixed neurological and psychiatric cohort. The Mini-SEA was generally reported to have excellent performance for differentiating bvFTD from MDD (AUC = 0.98)[33] and AD (AUC = 0.97–0.87)[29,32,40,48,49]; however, less so for BDI (AUC = 0.63).[50] While the results for differentiating bvFTD from healthy controls and AD were consistent across multiple studies, the findings for other conditions were mostly based on single studies.

Other neuropsychological tests (Table 4).

Memory. Eight studies of low quality according to the QUADAS-2 appraisal[22,29,49,58–62] used memory tests to differentiate bvFTD from healthy controls and other conditions, including AD and a mixed dementia cohort. The primary aim for most studies that used memory tests was to differentiate AD from other conditions, and thus the data for bvFTD was a secondary outcome. Accordingly, the direction of the receiver operating characteristic (ROC) curve analysis was not optimized for bvFTD in some studies.[61,62] The general theme of these tests was that encoding (e.g., immediate recall) and retention (e.g., delayed recall) reported only average performance in differentiating bvFTD from AD; when these verbal memory tests were optimized as composite scores, performance improved. One study[60] calculated an index score for to the Rey Auditory Verbal Learning Test (RAVLT)[133] that combined scores of encoding (trials 1–5), retention (delayed recall), and recognition memory (delayed recognition correct hits and false positives). When using this index score to differentiate bvFTD from AD, the efficacy of verbal memory testing was excellent with an AUC of 0.93. Applying this index score to bvFTD versus healthy controls also had good test performance, with an AUC of 0.80. Other tests of verbal memory, such as the Addenbrooke's Cognitive Examination-Revised (ACE-R)[134] memory subscore reported similar results for differentiating bvFTD from healthy controls, but performed poorly for AD with an AUC of 0.65.[58]

Attention and orientation. Two studies of low quality according to the QUADAS-2 appraisal[58,63] used attention and orientation tests to differentiate bvFTD from healthy controls and AD. In one study,[58] the ACE-R attention and orientation subscore was reported to have average efficacy in differentiating bvFTD from AD with an AUC of 0.71. This study did not optimize the direction of the ROC curve analysis for bvFTD versus healthy controls. Another study[63] that used the orientation subscore on the Virtual Supermarket Task reported an excellent discriminative ability to differentiate bvFTD from AD with an AUC of 0.91.

Language. Seven studies of low quality according to the QUADAS-2 appraisal[22,29,33,37,58,64,66] used language tests to differentiate bvFTD from healthy controls and other conditions, including AD, PPA, MDD, and a mixed psychiatric cohort. In differentiating bvFTD from healthy controls, language tests, including the Screening Linguistics Test,[135] ACE-R language subscore, and Semantic Fluency Test,[136] had good to excellent test performance. However, when language tests were used to differentiate bvFTD from other conditions, such as AD, PPA, and MDD, the efficacy was consistently poor or failed to differentiate these cohorts.

Visuospatial. Four studies of low quality according to the QUADAS-2 appraisal[22,58,67,68] used visuospatial tests to differentiate bvFTD from healthy controls and AD. These tests were consistently reported as poor to average in differentiating bvFTD from healthy controls or AD.

Praxis. Two studies of low quality according to the QUADAS-2 appraisal[69,70] used praxis tests to differentiate bvFTD from healthy controls and AD. The Cologne Apraxia Screening[137] was reported as best at differentiating bvFTD from healthy controls, with an AUC of 0.98.[69] In addition, the Dementia Apraxia Test[70] was reported to be excellent at differentiating bvFTD from AD, with an AUC of 0.90. Both of these results were supported by single-study findings.

TABLE 3 Social cognition tests.

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/Max score
	Chinese Facial Affective Picture System (CFAPS)											
[29]	35-item version—Total score	bvFTD < HC	22 vs 30	90.9	90.0			0.95				≤10.3/15.0
[29]	35-item version—Total score	bvFTD < AD	22 vs 26	77.3	69.2			0.81				≤9.4/15.0
	Ekman's Faces Test (EFT)											
[46]	60-item version—Total score	bvFTD < HC	25 vs 33	94.0	100			0.97			97.0	≤45.0/60.0
[47]	60-item version (Italian)—Total score	bvFTD < HC	32 vs 40	84.0	93.0			0.92			89.0	≤43.1/60.0
[48]	35-item version—Total score	bvFTD < AD	22 vs 20					0.77				N/A
[49]	35-item version—Total score	bvFTD < AD	38 vs 28								76.9	X/15.0
[47]	60-item version (Italian)—Total score	bvFTD < AD	32 vs 26	66.0	85.0			0.78			74.0	≤36.5/60.0
[47]	60-item version (Italian)—Total score	bvFTD < PPA	32 vs 16	75.0	75.0			0.80			77.0	≤42.1/60.0
[47]	60-item version (Italian)—Total score	bvFTD < CBS	32 vs 17					0.66				N/A
[50]	35-item version—Total score	bvFTD < BDI	18 vs 20	67.0	12.0			0.29				X/15.0
[47]	60-item version (Italian)—Total score	bvFTD < AD/PPA/CBS/HC	32 vs 99	66.0	75.0			0.75			71.0	≤36.5/60.0
[51]	60 item version—Total score	bvFTD < Mx PPD/Neuro	22 vs 57	66.7	68.2			0.73				≤34.5/60.0
	Emotion Recognition and Attribution (ERA) Index											
[47]	Total score	bvFTD < HC	32 vs 40	94.0	73.0			0.90			82.0	≤95.8/240.0
[47]	Total score	bvFTD < AD	32 vs 26	63.0	85.0			0.74			72.0	≤65.0/240.0
[47]	Total score	bvFTD < PPA	32 vs 16	63.0	87.0			0.76			71.0	≤65.0/240.0
[47]	Total score	bvFTD < CBS	32 vs 17					0.63				N/A
[47]	Total score	bvFTD < AD/PPA/CBS/HC	32 vs 99	63.0	80.0			0.72			74.0	≤65.0/240.0
	Emotional Recognition Task (ERT)											
[52]	Total score	bvFTD < HC	32 vs 49	95.9	62.5			0.81				≤43.5/96.0
[52]	Total score	bvFTD < AD	32 vs 32					0.52				N/A
[52]	Total score	bvFTD < Pre-sym. carriers	32 vs 47	89.4	78.1			0.83				≤50.5/96.0
	Faux Pas Test (FAUX)											
[29]	10-item version (Chinese)—Total score	bvFTD < HC	22 vs 30	86.4	90.0			0.95				≤9.8/15.0
[17]	20-item version—Total score	bvFTD < HC	35 vs 14					0.92				N/A
[29]	10-item version (Chinese)—Total score	bvFTD < AD	22 vs 26	86.4	84.6			0.89				≤9.8/15.0
[48]	10-item version—Total score	bvFTD < AD	22 vs 20					0.91				N/A
[49]	10-item version—Total score	bvFTD < AD	38 vs 28								89.2	X/15.0
[50]	10-item version—Faux pas detection score	bvFTD < BDI	18 vs 20	72.0	78.0			0.80				X/5.0
[50]	10-item version—Total score	bvFTD < BDI	18 vs 20	78.0	28.0			0.61				X/15.0

(Continues)

TABLE 3 (Continued)

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/Max score
[51]	20-item version—Total score	bvFTD < Mx PPD/ Neuro	22 vs 57					0.60				N/A
[53]	20-item version—Total score	bvFTD < Mx PPD/ Neuro	12 vs 22	83.0	64.0							X/80.0
[53]	20-item version—Total score	bvFTD < Mx PPD cohort	12 vs 15	83.0	81.0							X/80.0
	Mini-SEA											
[29]	Chinese version—Total score	bvFTD < HC	22 vs 30	95.5	93.3			0.99				≤21.4/30.0
[40]	Total score	bvFTD < AD	18 vs 33	100.0	83.3			0.96	6.0	<0.0		≤19.0/30.0
[29]	Chinese version—Total score	bvFTD < AD	22 vs 26	81.8	96.2			0.90				≤18.9/30.0
[32]	Total score	bvFTD < AD	20 vs 20					0.93			82.5	N/A
[48]	Total score	bvFTD < AD	22 vs 20					0.87				N/A
[49]	Total score	bvFTD < AD	38 vs 28					0.95			87.9	X/30.0
[54]	Total score	bvFTD < AD	20 vs 33	75.0	70.0						71.7	X/30.0
[54]	Total score	bvFTD < PD	20 vs 51	93.8	100.0						98.3	X/30.0
[33]	Total score	bvFTD < MDD	37 vs 19	89.2	100.0			0.98				≤22.1/30.0
[33]	Total score	bvFTD (mild) < MDD	17 vs 19	94.1	100.0							≤22.1/30.0
[33]	Total score	bvFTD (mod) < MDD	20 vs 19	85.0	100.0							X/30.0
[50]	Total score	bvFTD < BDI	18 vs 20	78.0	34.0			0.63				X/30.0
	Modified Emotion Hexagon Test (EHT)											
[19]	Total score	bvFTD < AD	11 vs 10					0.79				N/A
	Moral Emotions Assessment Task (MEAT)											
[55]	Total score	bvFTD < AD	22 vs 15	82.0	73.0	90.0	76.0	0.83	3.07	0.25		<37.0/42.0
	Morphed Faces Test (MFT)											
[56]	Total score	bvFTD < MDD	25 vs 21	91.0	76.0			0.91				≤5.2/7.0
[56]	Total negative emotion score	bvFTD < MDD	25 vs 21	91.0	80.0			0.95				≤5.2/7.0
	Reading the Mind in the Eyes Test (RMET)											
[37]	24-item version—Total score	bvFTD < HC	86 vs 43					0.90				N/A
[17]	36-item version—Total score	bvFTD < HC	35 vs 14					0.84				N/A
[19]	36-item version—Total score	bvFTD < AD	11 vs 10					0.86				N/A
	Social Cognition and Emotional Assessment (SEA)											
[57]	Total score	bvFTD < HC	22 vs 30	100.0	100.0			1.00				≤39.4/55.0
[57]	Total score	bvFTD < AD/aMCI	22 vs 22	86.4	95.5			0.96				≤34.6/55.0
[33]	Total score	bvFTD < MDD	37 vs 19	91.9	89.5			0.97				≤37.1/55.0
[33]	Total score	bvFTD (mild) < MDD	17 vs 19	94.1	89.5							≤35.3/55.0
[33]	Total score	bvFTD (mod) < MDD	20 vs 19	90.0	89.5							X/55.0
[57]	Total score	bvFTD < HC/ AD/aMCI	22 vs 52	100	88.5			0.98				≤39.4/55.0

TABLE 3 (Continued)

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/Max score
	Story-based Empathy Task (SET)											
[47]	Total score	bvFTD < HC	32 vs 40	75.0	80.0			0.83			83.0	≤12.3/18.0
[47]	Total score	bvFTD < AD	32 vs 26					0.62				N/A
[47]	Total score	bvFTD < PPA	32 vs 16					0.64				N/A
[47]	Total score	bvFTD < CBS	32 vs 17					0.44				N/A
[47]	Total score	bvFTD < AD/PPA/CBS/HC	32 vs 99					0.57				N/A
	The Awareness of Social Inference Test (TASIT)											
[19]	Social inference-minimal task subscore	bvFTD < AD	11 vs 10					0.88				N/A
[19]	Emotion evaluation task subscore	bvFTD < AD	11 vs 10					0.79				N/A

Note: See supporting Information S1 for test references.

Abbreviations: Acc., accuracy; AD, Alzheimer's disease; aMCI, amnestic mild cognitive impairment; AUC, area under the curve; BDI, Bipolar I disorder; bvFTD, behavioral variant frontotemporal dementia; CBS, corticobasal syndrome; HC, healthy controls; MDD, major depressive disorder; Mx, mixed; N/A, not applicable; Neuro, neurological; NPV, negative predictive value; NLR, negative likelihood ratio; PPA, primary progressive aphasia; PLR, positive likelihood ratio; PPV, positive predictive value; Sens., sensitivity; Spec., specificity.

Cognitive batteries (Table 5)

Fourteen studies of low quality according to the QUADAS-2 appraisal[17,22,23,33,39,40,54,58,71–76] used cognitive batteries to differentiate bvFTD from healthy controls and other conditions, including AD, MDD, mixed dementia, and mixed psychiatric cohorts. These included the Montreal Cognitive Assessment (MoCA),[138] the MMSE, three different versions of the Addenbrooke's Cognitive Examination (ACE,[139] ACE-R, and ACE-III[140]), and the Edinburgh Cognitive and Behavioral ALS screen (ECAS).[141] In differentiating bvFTD from healthy controls, the MoCA, ACE, ACE-R, and ACE-III were reported to have good to excellent test performance. These results, however, were generally supported by only one to two studies. The MMSE performed slightly worse, with an AUC of 0.72–0.88.[17,39,76] In differentiating bvFTD from other conditions, there were conflicting results. The MMSE was reported to have average efficacy to differentiate bvFTD from MDD, with an AUC of 0.78–0.75, in two studies.[33,39] The ACE-III was reported to have good efficacy to differentiate bvFTD from AD, with an AUC of 0.85 in one study,[40] and yet the ACE-R failed in another study.[58] The latter study also investigated the ACE's Verbal-Language/Orientation-Memory (VLOM) ratio, which has been proposed to specifically discriminate frontotemporal lobar degeneration syndromes from AD.[139] The VLOM ratio had an AUC of 0.83 in differentiating bvFTD from AD, compared to an AUC of 0.50 for the total ACE score.

Behavioral tools (Table 6)

Twenty-nine studies of low quality according to the QUADAS-2 appraisal[22–24,35,37,40,45,77–98] used behavioral tools to differentiate bvFTD from healthy controls and other conditions, including AD, PPA, and mixed neurological and/or psychiatric cohorts. The results from the behavioral tools were generally consistent reporting good to excellent test performance in differentiating bvFTD from all cohorts. In addition, the results of two tools were supported by multiple studies—the FBI and Interpersonal Reactivity Index (IRI).[142]

The FBI was reported to have an excellent ability in differentiating bvFTD from AD, with an AUC of 0.99 in one study.[85] Shorter versions of the FBI also had good test performance (though not performing as well as the original version): the mini-FBI[86] had an AUC of 0.81 and the modified-FBI[143] had an AUC of 0.78.[86,87] The FBI also had excellent test performance in differentiating bvFTD from a mixed cohort of AD and vascular dementia.[88] The only exception to these results was one study,[40] which reported that the FBI performed at chance (AUC = 0.50) in differentiating bvFTD from AD. This AUC value may have been incorrectly reported, however, because the study also reported a sensitivity of 83.3%, a specificity of 100% (at a cut-off score of 19), and a statistically significant difference between the mean score of the bvFTD and AD cohorts ($p < 0.001$); thus, making an AUC value of 0.5 implausible.

The IRI was reported to have average to excellent ability in differentiating bvFTD from other cohorts. One study[45] reported that the IRI was average at differentiating bvFTD from a mixed dementia cohort of FTD-motor neuron disease, PPA, AD, PCA, PSP, dementia with Lewy bodies, and CBS. When used to differentiate bvFTD from the isolated AD and semantic dementia cohorts, it still had an average test performance. Another study,[40] however, reported that the perspective taking subscore performed excellently at differentiating

TABLE 4 Other neuropsychological tests.

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/ Max score
	Memory tests											
	Addenbrooke's Cognitive Examination-Revised (ACE-R)											
[58]	Brazilian version—Memory subscore	bvFTD < HC	37 vs 68	75.7	66.2	54.9	83.3	0.76				≤11.0/26.0
[58]	Brazilian version—Memory subscore	bvFTD < AD	37 vs 102	54.1	81.0	71.4	66.7	0.65				≤16.0/26.0
	Edinburgh Cognitive and Behavioral ALS Screen (ECAS)											
[22]	Memory subscore	bvFTD < HC	16 vs 48	75.0	95.8							≤13.0/24.0
	Free and Cued Selective Reminding Test (FCSRT)											
[59]	Immediate + Delayed recall subscore	bvFTD > AD	32 vs 32	71.4	90.9						78.1	X/32.0
[49]	Immediate + Delayed recall subscore	bvFTD > AD	38 vs 28					0.77			69.7	X/32.0
	Modified-World Health Organization/University of California-Los Angeles Auditory Verbal Learning Test (WHO/UCLA AVLT)											
[29]	Delayed recall score	bvFTD > AD	22 vs 26					0.79				N/A
	Rey Auditory Verbal Learning Test (RAVLT)											
[60]	Delay recall score	bvFTD < HC	15 vs 28	100	54.4							≤5.5/15.0
[60]	Delay recall score	bvFTD < HC	11 vs 15	93.3	73.7							≤8.0/15.0
[60]	Learning trials 1–5 score	bvFTD < HC	15 vs 28	69.0	80.0							≤36.5/75.0
[60]	Learning trials 1–5 score	bvFTD < HC	11 vs 15	86.7	91.0							≤44.5/75.0
[60]	Correct hits on delayed recognition score	bvFTD < HC	15 vs 28	86.2	74.3							≤14.5/15.0
[60]	Correct hits on delayed recognition score	bvFTD < HC	11 vs 15	80.0	44.5							≤13.5/15.0
[60]	False positive errors on recognition score	bvFTD > HC	15 vs 28	93.0	40.0							≥0.5/15.0
[60]	False positives errors on recognition score	bvFTD > HC	11 vs 15	87.7	81.8							≥1.5/15.0
[60]	Memory efficiency index score	bvFTD < HC	26 vs 43	81.8	73.1			0.80				≤1.9/N/A
[60]	Delay recall score	bvFTD > AD	15 vs 39	86.7	84.8							≥3.5/15.0
[60]	Delay recall score	bvFTD > AD	11 vs 17	81.8	89.9							≥3.5/15.0
[60]	Learning trials 1–5 score	bvFTD > AD	15 vs 39	80.0	86.5							≥27.5/75.0
[60]	Learning trials 1–5 score	bvFTD > AD	11 vs 17	90.9	50.0							≥26.0/75.0
[60]	Correct hits on delayed recognition score	bvFTD > AD	15 vs 39	86.7	70.0							≥12.5/15.0
[60]	Correct hits on delayed recognition score	bvFTD > AD	11 vs 17	90.9	61.1							≥11.5/15.0
[60]	False positive errors on recognition score	bvFTD < AD	15 vs 39	73.3	70.3							≤1.5/15.0
[60]	False positive errors on recognition score	bvFTD < AD	11 vs 17	55.5	77.8							≤9.5/15.0
[60]	Memory efficiency index score	bvFTD > AD	26 vs 56	84.6	85.0			0.93				≥1.2/N/A

TABLE 4 (Continued)

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/ Max score
	Verbal Learning and Memory Test (VLMT)											
[61]	Recency dominance scores	bvFTD (amnestic) > AD	8 vs 20	75.0	95.0			0.73			0.86	≥−0.15
[61]	Recognition memory score	bvFTD (amnestic) > AD	8 vs 20	87.5	65.0			0.52			0.50	≥5.5/X
[61]	Delayed recall score	bvFTD (amnestic) > AD	8 vs 20	37.5	65.0			0.39			0.64	≥2.5/X
	Word List and Story Recall Test (WLSR)											
[62]	Immediate differential score	bvFTD < AD/ DLB/VasD	20 vs 60					0.42				N/A
[62]	Delay differential score	bvFTD < AD/ DLB/VasD	20 vs 60					0.40				N/A
Attention and orientation tests												
	Addenbrooke's Cognitive Examination-Revised (ACE-R)											
[58]	Brazilian version—Attention and orientation subscore	bvFTD < HC	37 vs 68					0.32				N/A
[58]	Brazilian version—Attention and orientation subscore	bvFTD < AD	37 vs 102	75.7	69.0	68.3	76.3	0.71				≤16.0/18.0
	Virtual Supermarket Task (VST)											
[63]	Orientation score	bvFTD > AD	23 vs 21					0.91				N/A
Language tests												
	Addenbrooke's Cognitive Examination-Revised (ACE-R)											
[58]	Brazilian version—Language subscore	bvFTD < HC	37 vs 68	70.3	79.4	65.0	83.1	0.80				≤24.0/26.0
[58]	Brazilian version—Language subscore	bvFTD < AD	37 vs 102	43.2	78.6	64.0	61.1	0.61				≤20.0/26.0
	Boston Naming Test (BNT)											
[64]	12-item version—Total score	bvFTD < HC	55 vs 46	41.5	87.8			0.72				≤7.0/12.0
[29]	30-item version (Chinese)—Total score	bvFTD < AD	22 vs 26					0.69				N/A
[65]	29-item version—Total score	bvFTD < Mx PPD cohort	32 vs 53	92.0	64.0			0.81				<72.0/87.0
[65]	29-item version—Total score	bvFTD < Mx PPD cohort	32 vs 53	48.0	90.0			0.81				<60.0/87.0
	Edinburgh Cognitive and Behavioral ALS Screen (ECAS)											
[22]	Language subscore	bvFTD < HC	16 vs 48	75.0	75.0							≤26.0/28.0
	Screening Linguistics Test (ScreenLing)											
[66]	Total score	bvFTD < HC	46 vs 35	90.9	91.0			0.93				≤70.0/72.0
[66]	Syntax subscore	bvFTD < HC	46 vs 35	53.3	95.5			0.83				≤23.0/24.0
[66]	Phonology subscore	bvFTD < HC	46 vs 35	50.0	100.0			0.81				≤23.0/24.0
[66]	Semantics subscore	bvFTD < HC	46 vs 35	56.8	90.9			0.84				≤23.0/24.0
[66]	Total score	bvFTD < AD	46 vs 20					0.56				N/A
[66]	Syntax subscore	bvFTD < AD	46 vs 20					0.53				N/A

(Continues)

TABLE 4 (Continued)

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/ Max score
[66]	Phonology subscore	bvFTD < AD	46 vs 20					0.57				N/A
[66]	Semantics subscore	bvFTD < AD	46 vs 20					0.50				N/A
[66]	Total score	bvFTD < svPPA	46 vs 32					0.63				N/A
[66]	Syntax subscore	bvFTD < svPPA	46 vs 32					0.60				N/A
[66]	Phonology subscore	bvFTD < svPPA	46 vs 32					0.52				N/A
[66]	Semantics subscore	bvFTD < svPPA	46 vs 32	59.4	79.5			0.69				≤20.0/24.0
	Semantic fluency (SF)											
[37]	Animals—Total score	bvFTD < HC	86 vs 43					0.95				N/A
	Verbal fluency (VF)											
[58]	P-words + Animals—Total score (ACE-R VF subscore)	bvFTD < HC	37 vs 68	91.9	66.2	59.6	93.8	0.85				≤11.0/14.0
[58]	P-words + Animals—Total score (ACE-R VF subscore)	bvFTD < AD	37 vs 102	43.2	78.6	64.0	61.1	0.61				≤6.0/14.0
[33]	PF (M-words) + SF (animals)—Total score	bvFTD < MDD	37 vs 19					0.71				N/A
[33]	PF (M-words) + SF (animals)—Total score	bvFTD (mod) < MDD	17 vs 19	78.6	50.0			0.61				≤10.0/N/A
Visuospatial tests												
	Addenbrooke's Cognitive Examination-Revised (ACE-R)											
[58]	Brazilian version—Visuospatial subscore	bvFTD < HC	37 vs 68	86.5	54.4	50.8	88.1	0.76				≤16.0/16.0
[58]	Brazilian version—Visuospatial subscore	bvFTD < AD	37 vs 102	86.5	33.3	53.3	73.7	0.57				≤16.0/16.0
	Clock Drawing Test (CDT)											
[67]	Rouleau scoring system—Total score	bvFTD < HC	112 vs 300	55.0	74.0	68.0	62.0	0.69			65.0	X/10.0
[67]	Cahn scoring system—Total score	bvFTD < HC	112 vs 300	59.0	73.0	69.0	64.0	0.71			66.0	X/8.0
[67]	Babins scoring system—Total score	bvFTD < HC	112 vs 300	69.0	65.0	66.0	68.0	0.71			67.0	X/18.0
	Edinburgh Cognitive and Behavioral ALS Screen (ECAS)											
[22]	Visuospatial subscore	bvFTD < HC	16 vs 48	43.8	91.7							≤10.0/12.0
	Rey–Osterrieth Complex Figure copy task (ROCF)											
[68]	Inner details deficits score	bvFTD < AD	15 vs 41	66.7	70.7	45.4	85.3				69.9	X/X
Praxis tests												
	Cologne Apraxia Screening (CAS)											
[69]	Total score	bvFTD < HC	20 vs 20	95.0	85.0			0.98				≤74.0/80.0
	Dementia Apraxia Test (DATE)											
[70]	Limb apraxia minus buccofacial apraxia subscores	bvFTD < AD	24 vs 28	74.0	93.0			0.90				−7.0/−21.0
	Ideomotor Apraxia Test (IAT)											
[69]	Total score	bvFTD < HC	20 vs 20	80.0	70.0			0.82				≤22.0/30.0

TABLE 4 (Continued)

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/ Max score
	Münster Apraxia Items (MI)											
[69]	Total score	bvFTD < HC	20 vs 20	80.0	70.0			0.85				≤21.0/24.0

Note: See Supporting Information S1 for test references.

Abbreviations: Acc., accuracy; AD, Alzheimer's disease; AUC, area under the curve; bvFTD, behavioral variant frontotemporal dementia; DLB, dementia with Lewy bodies; HC, healthy controls; MDD, major depressive disorder; Mx, mixed; N/A, not applicable; NPV, negative predictive value; NLR, negative likelihood ratio; PLR, positive likelihood ratio; PPD, primary psychiatric disorder; PPV, positive predictive value; Sens., sensitivity; Spec., specificity; svPPA, semantic variant primary progressive aphasia; VasD, vascular dementia.

bvFTD from AD with an AUC of 0.97. In contrast, the emotional concern subscore performed slightly worse in the same cohort with an AUC of 0.89.

There were many other behavioral tools created to identify frontal behavioral symptoms, all with reported good to excellent test performance. These included the Disinhibition, Apathy, Perseveration, Hyperorality, Negligence, Empathy loss scale (DAPHNE)[82]; the Behavioral Frontotemporal Lobe Dysfunction Assessment Scale[78]; the Frontal Systems Behavior Scale[144]; the Social Behavior Observer Checklist[97]; the Middelheim Frontality Score[90]; the Pick's Disease Scale[145]; the Informant-Based Questionnaire[146]; the ECAS behavior subscore; and the Mild Behavioral Impairment Checklist (MBI-C).[147] Unfortunately, the results of all these tools were supported by single-study findings of low quality according to the QUADAS-2 appraisal.

Other clinical tools (Table 7)

Four studies of low quality according to the QUADAS-2 appraisal[37,99,100,102] used other tools to differentiate bvFTD from healthy controls and other conditions, including AD, MDD, PPA, and a mixed psychiatric cohort. Two studies focused on the CDR and CDR-FTLD, the first of which[37] reported that both performed excellently at differentiating bvFTD from healthy controls. The second study[100] reported that the behavior subscore on the CDR-FTLD was better at differentiating bvFTD from a mixed cohort of AD, PPA, and healthy controls, compared to the memory and language subscores.

One study[99] used the Alberta Smell Test[148] to explore the utility of odor identification. Good performance was reported in differentiating bvFTD from MDD, with a sensitivity of 91.0% and specificity of 97.0% at a cut-off of 2.0/20.0.

The final study proposed the Screening Instrument for Frontotemporal Dementia (SIFTD)[102] to differentiate bvFTD from AD. This tool included a neurological assessment of primitive reflexes, a social cognition test, and assessment of perseveration. At a cut-off score of 3.0/12.0, the SIFTD had a sensitivity of 83.3% and specificity of 91.7%.

Results of studies that combined different tools (Table 8)

Fourteen studies of low-quality according to the QUADAS-2 appraisal[17,29,31,40,54,63,103–110] used various combinations of tools to differentiate bvFTD from other cohorts.[17,29,31,40,54,63,103–110] Two of these studies included participants with only mild bvFTD.[29,110] These studies reported that the combination of a social cognition test with a test of executive function, memory, or a cognitive battery performed excellently at differentiating bvFTD from healthy controls, AD, or Parkinson's disease (PD).

Two studies combined tools to differentiate bvFTD from healthy controls.[17,54] One study[54] reported that combining the Mini-SEA and MoCA had an excellent test performance with an AUC of 0.95. Another study[17] compared different combinations of executive function and social cognition tests to individual tests. This study reported that combining the Modified Hotel Task (HOT),[149] Multiple Errands Test (MET),[150] Iowa Gambling Task (IGT),[151] and Reading the Mind in the Eyes Test (REMT) had the best efficacy, with an AUC of 0.99, compared to individual tests of executive function, social cognition, or cognitive test batteries, such as the MMSE and ACE. However, combining a single executive function test with a social cognition test also had excellent test performance. For instance, the IGT with the FAUX, or RMET,[152] had an AUC of 0.96.

Nine studies combined tools to differentiate bvFTD from other conditions.[29,31,40,54,63,104–106,108] Four of these[29,104,105,108] combined a memory test with another tool to differentiate bvFTD from AD. Though these studies used different parameters to define test performance, the study that appeared to have the best results[29] reported that the Mini-SEA combined with the Modified-World Health Organization/University of California-Los Angeles Auditory Verbal Learning Test (WHO/UCLA AVLT)[153] had an AUC of 0.95. Another study[63] applied three different memory tests to differentiate bvFTD from AD, with similar results. This study applied the ACE-R memory subscore, RAVLT, and Rey–Osterrieth complex figure,[154] with an AUC of 0.92. Two studies[40,54] combined the Mini-SEA with either the MoCA or ACE-III. Both studies had excellent test performance in differentiating bvFTD from AD and PD. The last two studies[31,106] combined executive function tests with behavioral measures. The first study[31] combined the FAB with an informant-rated questionnaire of frontal behavioral symptoms. This combination had a sensitivity of 100% and specificity of 93.5%. The second study[106] combined phonemic fluency, and an antisaccade task, with their in-house developed, patient-rated Social Norms Questionnaire

TABLE 5 Cognitive test batteries.

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/Max score
	Addenbrooke's Cognitive Examination (ACE)											
[17]	Total score	bvFTD < HC	35 vs 14					0.81				N/A
[71]	Verbal-language/orientation-memory (VLOM) ratio	bvFTD < AD	9 vs 25	11.1	88.0							<2.2
[72]	Japanese version—VLOM ratio	bvFTD < MCI/DLB/VasD	24 vs 115	16.7	96.5							<2.2
	Addenbrooke's Cognitive Examination Revised (ACE-R)											
[58]	Brazilian version—Total score	bvFTD < HC	37 vs 68	73.0	83.8	71.1	85.1	0.85				≤79.0/100.0
[39]	Total score	bvFTD < HC	49 vs 26					0.93				N/A
[58]	VLOM ratio	bvFTD < HC	37 vs 68	32.4	95.6	80.0	72.2	0.56				<2.0
[58]	Brazilian version—Total score	bvFTD < AD	37 vs 102	73.0	35.7	50.0	60.0	0.50				≤79.0/100.0
[73]	VLOM ratio	bvFTD < AD	41 vs 46	79.0	81.0							<3.5
[58]	VLOM ratio	bvFTD < AD	37 vs 102	86.5	71.4	72.7	85.7	0.82				<3.1
[39]	Total score	bvFTD < MDD	49 vs 30					0.79				N/A
	Addenbrooke's Cognitive Examination-III (ACE-III)											
[74]	Spanish version—Total score	bvFTD < HC	31 vs 139	93.6	77.7			0.90				≤87.0/100.0
[75]	Total score	bvFTD < HC	18 vs 28	83.3	96.4							≤88.0/100.0
[22]	Total score	bvFTD < HC	14 vs 48	78.6	97.7							≤82.0/100.0
[22]	Total score	bvFTD < HC	14 vs 48	85.7	86.4							≤88.0/100.0
[40]	Total score	bvFTD < AD	18 vs 33	66.7	94.4			0.85	12.0	0.35		≤70.0/100.0
[75]	Total score	bvFTD < AD	18 vs 31	83.3	96.7							≤88.0/100.0
[22]	Total score	bvFTD < AD	14 vs 25	78.6	20.0							≤82.0/100.0
[22]	Total score	bvFTD < AD	14 vs 25	85.7	08.0							≤88.0/100.0
[75]	Total score	bvFTD < PPA	18 vs 11	83.3	90.9							≤88.0/100.0
[75]	Total score	bvFTD < PCA	18 vs 11	83.3	90.9							≤88.0/100.0
[75]	Total score	bvFTD < AD/PPA/PCA	18 vs 53	83.3	94.3							≤88.0/100.0
	Edinburgh Cognitive and Behavioral ALS screen (ECAS)											
[22]	Total score	bvFTD < HC	16 vs 48	93.8	95.8							≤105.0/136.0
[22]	Anterior composite score (fluency + executive + language score)	bvFTD < HC	16 vs 48	93.8	91.7							≤77.0/100.0
[22]	Posterior composite score (memory + visuospatial score)	bvFTD < HC	16 vs 48	75.0	95.8							≤24.0/36.0
[22]	Total score	bvFTD < AD	16 vs 32	93.8	06.3							≤91.0/136.0
[22]	Anterior composite score	bvFTD < AD	16 vs 32	93.8	15.6							≤62.0/100.0
[22]	Posterior composite score	bvFTD < AD	16 vs 32	75.0	03.1							≤23.0/36.0
	Montreal Cognitive Assessment (MoCA)											
[76]	Total score	bvFTD < HC	50 vs 50	78.0	98.0	98.0	82.0	0.93				<17.0/30.0
[54]	Total score	bvFTD < AD	20 vs 33	68.8	73.7						71.7	X/30.0
	Mini-Mental State Examination (MMSE)											
[76]	Total score	bvFTD < HC	50 vs 50	58.0	88.0	83.0	68.0	0.77				<26.0/30.0

TABLE 5 (Continued)

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/Max score
[17]	Total score	bvFTD < HC	35 vs 14					0.72				N/A
[39]	Total score	bvFTD < HC	49 vs 26					0.88				N/A
[23]	Total score	bvFTD > AD	13 vs 22	88.3	91.7			0.96				≥19.5/30.0
[23]	Total score	bvFTD > AD	13 vs 22	90.5	88.5			0.96				≥21.5/30.0
[33]	Total score	bvFTD < MDD	37 vs 19					0.75				N/A
[33]	Total score	bvFTD (mod) < MDD	20 vs 19	76.5	36.8			0.56				<25.0/30.0
[39]	Total score	bvFTD < MDD	49 vs 30					0.78				N/A
[35]	Total score	bvFTD < Mx PPD/Neuro cohort	55 vs 88					0.61				N/A

Note: See Supporting Information S1 for test references.

Abbreviations: Acc., accuracy; AD, Alzheimer's disease; AUC, area under the curve; bvFTD, behavioral variant frontotemporal dementia; DLB, dementia with Lewy bodies; HC, healthy controls; MCI, mild cognitive impairment; MDD, major depressive disorder; Mx, mixed; N/A, not applicable; Neuro, neurological; NPV, negative predictive value; NLR, negative likelihood ratio; PCA, posterior cortical atrophy; PLR, positive likelihood ratio; PPA, primary progressive aphasia; PPD, primary psychiatric disorder; PPV, positive predictive value; Sens., sensitivity; Spec., specificity; VasD, vascular dementia.

(SNQ) and clinician-rated Behavioral Rating Scale (BRS). The SNQ was a patient-rated questionnaire on socially appropriate behavior and the BRS was a clinician rated scale of observable frontal behavior. This combination did not perform as well, with a sensitivity of 65.0% and specificity of 79.0%.

The comparison of different tools within the same cohort

Seven low-quality studies according to the QUADAS-2 appraisal[17,19,25,29,33,40,83]

compared different tests in the same cohort. All these studies included participants with only mild bvFTD. These studies reported that social cognition and behavioral tools outperformed executive function and other tests.

One study[25] compared two executive function tests, the IFS and FAB, to differentiate mild bvFTD from mild AD, and healthy controls. Both tools were reported to have good test performance in differentiating mild bvFTD from healthy controls; however, both failed to differentiate mild bvFTD from mild AD. Another study[83] compared three behavioral tools, the Neuropsychiatric Inventory (NPI),[155] the FBI, and the MBI-C, to differentiate bvFTD from healthy controls. This study reported that all three tools had similar test performance, with an AUC ranging from 0.96 to 0.91.

Two studies compared executive function and social cognition tests in the same cohort.[17,19] The first study[17] used the IGT, TMT-B, Wisconsin Card Sorting Test (WCST),[156] HOT, MET, Backward Digit Span Test,[157] phonemic fluency, FAUX, and REMT to differentiate mild bvFTD from healthy controls. Of these tools, the IGT and FAUX performed the best—both had excellent test performance with an AUC of 0.92. The second study[19] used the Brixton Test (BT),[158] Hayling Test,[158] Emotion Hexagon Test,[159] REMT, The Awareness of Social Inference Test (TASIT),[160] and the FAB to differentiate mild bvFTD from mild AD. The TASIT performed the best, with an AUC of 0.88, and the BT performed the worst, with an AUC of 0.57.

The last three studies compared cognitive tests of multiple cognitive domains in the same cohort.[29,33,40] One study[29] applied executive function, social cognition, memory, and language tests to differentiate mild bvFTD from mild AD, and from healthy controls. The tests included the FAB, Chinese Facial Affective Picture System (CFAPS),[161] FAUX, Mini-SEA, WHO/UCLA AVLT, and BNT. The social cognition tests outperformed executive function, memory, and language tests in differentiating mild bvFTD from both mild AD and healthy controls. In differentiating mild bvFTD from mild AD, the Mini-SEA performed best (AUC = 0.90), followed by the FAUX (AUC = 0.89), CFAPS (AUC = 0.81), WHO/UCLA AVLT (AUC = 0.79), and, finally, the FAB (AUC = 0.69) and BNT (AUC = 0.69) performed the worst. The next study[40] included the IFS, Mini-SEA, ACE-III, FBI, IRI, and Revised Self-Monitoring Scale (r-SMS).[162] In differentiating bvFTD from AD, the IRI (AUC = 0.97) performed the best, followed by the Mini-SEA (AUC = 0.96), r-SMS (AUC = 0.95), ACE-III (AUC = 0.85), and, finally, the IFS (AUC = 0.78) and FBI performed the worst (AUC = 0.50). Note, however, that the test performance of the FBI in this study seems implausible. The last study[33] included the SEA,[57] Mini-SEA, MMSE, FBI, Verbal Fluency, and WCST. In differentiating bvFTD from MDD, the Mini-SEA performed the best (AUC = 0.98), followed by the SEA (AUC = 0.97), MMSE (AUC = 0.75), FBI (AUC = 0.71), Verbal Fluency (AUC = 0.71), and, finally, the WCST performed the worst (AUC = 0.66).

TABLE 6 Behavioral tools.

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/Max score
	Alzheimer's disease scale (AD scale)											
[23]	Total score	bvFTD < AD	6 vs 28	98.3	82.4			0.96				<5.3/17.0
	Apathy evaluation scale (AES)											
[37]	Self-rated total score	bvFTD > HC	86 vs 43					0.82				N/A
[37]	Informant-rated total score	bvFTD > HC	86 vs 43					0.96				N/A
	Barratt Impulsiveness Scale 11th version (BIS-11)											
[24]	Total score	bvFTD > AD	27 vs 25	68.2	80.0			0.79				68.0/120.0
	Bayer Activities of Daily Living Scale (BADL)											
[37]	Self-rated total score	bvFTD > HC	86 vs 43					0.74				N/A
[37]	Informant-rated total score	bvFTD > HC	86 vs 43					0.96				N/A
	Behavioral Dysfunction Questionnaire (BDQ)											
[77]	Global average score (without time constraint)	bvFTD > AD	34 vs 56	65.0	91.0			0.88				>1.4/5.0
[77]	Global average score (without time constraint)	bvFTD > MDD	34 vs 41	56.0	90.0			0.83				>1.6/5.0
	Behavioral FT Lobe Dysfunction Assessment Scale (BFLDAS)											
[78]	Total score	bvFTD > AD	33 vs 43	100.0	97.0							≥3.0/4.0
[78]	Total score	bvFTD > VasD	33 vs 16	100.0	87.0							≥3.0/4.0
[78]	Total score	bvFTD > AD/ VasD	33 vs 59	100.0	93.0	92.0	100	0.97				≥3.0/4.0
[82]	Total score	bvFTD > AD/ PSP/BDI/HC	36 vs 106	97.0	45.0				1.7			≥3.0/4.0
	Behavioral pathology in AD rating scale (Behave-AD)											
[79]	Aggression— (Affect + Anxiety/phobia) subscore	bvFTD > AD	29 vs 29	82.8	55.2						69.0	X/X
	Cambridge Behavioral Inventory (CBI)											
[80]	Total score	bvFTD > AD	13 vs 37	53.8	86.5							X/X
[80]	Total score	bvFTD > svPPA	13 vs 20	53.8	55.0							X/X
	Disinhibition, Apathy, Preservation, Hyperorality, Negligence, Empathy loss (DAPHNE)											
[81]	DAPHNE-6 (screening score)	bvFTD > fvAD	36 vs 20	92.0	55.0				2.0			≥4.0/6.0
[81]	DAPHNE-40 (total score)	bvFTD > fvAD	36 vs 20	47.0	85.0				3.1			≥16.0/40.0
[81]	DAPHNE "combined" score	bvFTD > fvAD	36 vs 20	92.0	85.0				6.1			X/46.0
[82]	DAPHNE-6 (screening score)	bvFTD > AD/ PSP/BDI/HC	36 vs 106	92.0	57.0				2.1			≥4.0/6.0
[82]	DAPHNE-40 (total score)	bvFTD > AD/ PSP/BDI/HC	36 vs 106	56.0	92.0				7.0			≥15.0/40.0
[82]	DAPHNE "combined" score	bvFTD > AD/ PSP/BDI/HC	36 vs 106	92.0	92.0				11.5			X/46.0
	Edinburgh Cognitive and Behavioral ALS Screen (ECAS)											
[22]	Behavior subscore	bvFTD > AD	15 vs 25	79.0	87.0							≥4.0/10.0
	Frontal Behavioral Inventory (FBI)											
[83]	Total score	bvFTD > HC	52 vs 82	98.0	91.0			0.96				≥8.0/72.0

TABLE 6 (Continued)

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/Max score
[83]	Total score	bvFTD (mild) > HC	52 vs 82	97.0	91.0			0.97				≥8.0/72.0
[83]	Total score	bvFTD (mod-severe) > HC	52 vs 82	100.0	93.0			0.99				≥10.5/72.0
[84]	Total score	bvFTD > AD	26 vs 38	88.5	100.0						100	≥30.0/72.0
[40]	Spanish version—Total score	bvFTD > AD	18 vs 33	83.3	100			0.50		0.17		≥19.0/72.0
[85]	Total score	bvFTD > AD	52 vs 52	90.0	100			0.99				≥27.0/72.0
[85]	Total score	bvFTD > AD	52 vs 52	100.0	62.0			0.99				≥17.0/72.0
[86]	FBI-MINI—Total score	bvFTD > AD	40 vs 33	73.0	76.0			0.81			74.0	≥13.0/50.0
[86]	FBI-MINI—Positive subscale score	bvFTD > AD	40 vs 33	83.0	76.0			0.83			79.0	≥2.0/18.0
[86]	FBI-MINI—Negative subscale score	bvFTD > AD	40 vs 33	83.0	57.0			0.77			71.0	≥9.0/18.0
[87]	FBI-MOD—Total score	bvFTD > AD	26 vs 53	73.1	67.9			0.78				≥16.0/66.0
[87]	FBI-MOD—Total score	bvFTD > MCI	26 vs 50	73.1	78.0							≥16.0/66.0
[84]	Total score	bvFTD > VasD	26 vs 16	88.5	81.2						85.7	≥30.0/72.0
[84]	Total score	bvFTD > PPA	26 vs 11	88.5	100.0						100	≥30.0/72.0
[87]	FBI-MOD—Total score	bvFTD > PPA	26 vs 7	73.1	85.7							≥16.0/66.0
[84]	Total score	bvFTD > Dep	26 vs 17	88.2	92.3						90.7	≥30.0/72.0
[82]	Total score	bvFTD > AD/PSP/BDI/HC	36 vs 106	67.0	91.0				7.4			≥27.0/72.0
[84]	Total score	bvFTD > AD/PPA/VasD/Dep	26 vs 82	88.5	93.9						92.6	≥30.0/72.0
[88]	Total score	bvFTD > AD/VasD	35 vs 37	97.0	95.0			0.99				≥23.0/72.0
[35]	Positive subscale score	bvFTD > Mx PPD/Neuro cohort	55 vs 82					0.68				N/A
[35]	Negative subscale score	bvFTD > Mx PPD/Neuro cohort	55 vs 82					0.67				N/A
	Frontal Systems Behavior Scale (FrSBe)											
[37]	Self-rated frequency subscore	bvFTD > HC	86 vs 43					0.73				N/A
[37]	Informant-rated frequency subscore	bvFTD > HC	86 vs 43					0.97				N/A
[89]	Total score	bvFTD > AD	31 vs 33	90.9	71.0					81.3		X/X
	Informant-Based Questionnaire (IBQ)											
[23]	Total score	bvFTD > AD	18 vs 19	100.0	100.0							≥0.0/17.0
	Interpersonal Reactivity Index (IRI)											
[40]	Emotional concern subscore	bvFTD < AD	18 vs 33	87.9	66.7			0.89	2.64	0.18		≤22.0/35.0
[45]	Emotional concern subscore	bvFTD < AD	406 vs 58	72.0	63.0			0.73				≤23.5/35.0
[40]	Perspective taking subscore	bvFTD < AD	18 vs 33	93.9	88.9			0.97	8.45	0.07		≤16.0/35.0
[45]	Perspective taking subscore	bvFTD < AD	406 vs 58	75.0	61.0			0.76				≤16.5/35.0

(Continues)

TABLE 6 (Continued)

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/Max score
[45]	Emotional concern subscore	bvFTD < svPPA	406 vs 61	66.0	64.0			0.68				≤22.5/35.0
[45]	Perspective taking subscore	bvFTD < svPPA	406 vs 61	64.0	64.0			0.66				≤14.5/35.0
[45]	Emotional concern subscore	bvFTD < Mx dementia cohort	406 vs 385	72.0	63.0			0.73				≤23.5/35.0
[45]	Perspective taking subscore	bvFTD < Mx dementia cohort	406 vs 385	75.0	64.0			0.76				≤16.5/35.0
	Middelheim Frontality Score (MFS)											
[90]	Total score	bvFTD > AD	62 vs 400	88.7	89.0	0.37	0.98					≥5.0/10.0
[90]	Total score	bvFTD (mild) > AD (mild)	24 vs 42	91.7	85.7							≥4.0/10.0
[90]	Total score	bvFTD (mild) > AD (mild)	24 vs 42	79.2	92.9							≥5.0/10.0
	Mild Behavioral Impairment Checklist (MBI-C)											
[83]	Total score	bvFTD > HC	52 vs 82	100.0	83.0			0.96				≥5.5/102.0
[83]	Total score	bvFTD (mild) > HC	52 vs 82	100.0	83.0			0.95				≥5.5/102.0
[83]	Total score	bvFTD (mod-severe) > HC	52 vs 82	95.0	93.0			0.97				≥12.0/102.0
	Neuropsychiatric Inventory (NPI)											
[83]	NPI Questionnaire—Total score	bvFTD > HC	52 vs 82	93.0	79.0			0.91				≥2.5/12.0
[83]	NPI Questionnaire—Total score	bvFTD (mild) > HC	52 vs 82	90.0	79.0			0.89				≥2.5/12.0
[83]	NPI Questionnaire—Total score	bvFTD (mod-severe) > HC	52 vs 82	95.0	79.0			0.93				≥2.5/12.0
[91]	(Apathy + Disinhibition) - Depression subscore	bvFTD > AD	22 vs 30	77.0	77.0							X/X
[92]	Total score	bvFTD > AD	12 vs 12	100.0	100.0							≥13.0/120.0
	Pick's Disease Scale (FTD scale)											
[23]	Total score	bvFTD > AD	6 vs 28	94.9	82.5			0.91				≥4.8/13.0
[93]	Total score	bvFTD > AD/VasD/MixDem	52 vs 138	93.0	92.0	81.0	97.0	0.96				≥6.0/13.0
	Revised Self-Monitoring Scale (r-SMS)											
[40]	Total score	bvFTD < AD	18 vs 33	100	72.2			0.95	3.60	0.0		≤32.0/65.0
	Scale for Emotional Blunting (SEB)											
[94]	Time difference score	bvFTD > AD	13 vs 18	92.0	80.0	84.6	94.4	0.98				≥15.0/32.0
[95]	Total score	bvFTD > AD/HC	12 vs 24	92.0	83.5							>12.0/32.0
	Social Behavior Questionnaire (SBQ)											
[96]	Total score	bvFTD > AD	23 vs 19	78.0	89.0							≥5.0/115.0
[96]	Total score	bvFTD > PSP/PPA	23 vs 14	78.0	79.0							≥5.0/115.0
	Social Behavior Observer Checklist (SBOCL)											
[97]	Disorganized subscale score	bvFTD > HC	135 vs 125			100		0.91				2.0/24.0
[97]	Reactive subscale score	bvFTD > HC	135 vs 125			100		0.67				3.0/15.0

TABLE 6 (Continued)

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/Max score
[97]	Insensitive subscale score	bvFTD > HC	135 vs 125			93.3		0.73				3.0/6.0
	Socioemotional Dysfunction Scale (SDS)											
[98]	Total score	bvFTD > AD	16 vs 18	88.0	83.0			0.92				>105.0/200.0
[98]	Total score	bvFTD > AD	16 vs 18	94.0	72.0			0.92				>89.0/200.0
	Stereotypy Rating Inventory (SRI)											
[35]	Total score	bvFTD > Mx PPD/Neuro cohort	55 vs 82					0.73				N/A

Note: See Supporting Information S1 for test references.

Abbreviations: Acc., accuracy; AD, Alzheimer's disease; AUC, area under the curve; BDI, Bipolar I disorder; bvFTD, behavioral variant frontotemporal dementia; Dep, depression; fvAD, frontal variant Alzheimer's disease; HC, healthy controls; MDD, major depressive disorder; MixDem, mixed dementia; Mx, mixed; N/A, not applicable; Neuro, neurological; NPV, negative predictive value; NLR, negative likelihood ratio; PLR, positive likelihood ratio; PPA, primary progressive aphasia; PPV, positive predictive value; PPD, primary psychiatric disorder; PSP, progressive supranuclear palsy; Sens., sensitivity; Spec., specificity; svPPA, semantic variant primary progressive aphasia; VasD, vascular dementia.

The impact of disease severity

The 24 studies that only included mild bvFTD reported that nearly all the tools had better performance compared to studies where the same tools were used but included participants with a mix of early and more advanced stages of the disease. For example, the studies that differentiated mild bvFTD from mild AD with memory tests had consistently good test performance.[29,61] In contrast, other studies reported a large variation in the ability of memory tests to differentiate bvFTD from AD of mixed disease severity. One study[33] did a subgroup analysis for mild bvFTD compared to a cohort of more advanced bvFTD. This study reported that the WCST, Verbal Fluency, FAB, MMSE, SEA and Mini-SEA all performed worse at differentiating more advanced bvFTD, compared to mild bvFTD, from moderate to severe MDD. Another study[83] did a subgroup analysis for mild bvFTD compared to moderate to severe bvFTD reporting that the FBI, MBI-C, and NPI all performed better at differentiating more advanced bvFTD, compared to mild bvFTD, from healthy controls.

DISCUSSION

While social cognition and behavioral tools appeared most useful, this systematic review highlighted notable inconsistencies within the literature pertaining to the clinical tools used to differentiate bvFTD from other conditions. In general, the robustness of the results was undermined by reliance on data from single studies, many of which were of low quality as evaluated by the QUADAS-2 criteria. In addition, the clinical applicability of reported results was limited due to inappropriate sampling. Most studies used comparator cohorts, such as healthy controls or amnestic presentations of Alzheimer's disease, where the clinical differentiation from bvFTD is rarely a challenge, thus questioning the relevance of results to real-world diagnostic scenarios. There were, however, five studies of high-quality rating that also reported that social cognition and behavioral tools outperformed other tests. Furthermore, despite the poor QUADAS-2 ratings of most studies, it was noted that a small number undertook a comparative analysis of multiple tests within the same cohort, thus design limitations were somewhat cancelled out; these studies reinforced the conclusion that social cognition and behavioral measures were the most effective methods in the identification of bvFTD.

BvFTD is unique among degenerative dementias in that the most salient early features are behavioral disturbances rather than deficits in traditional cognitive domains, such as memory and language. The diagnostic possibilities for behavioral changes with onset in middle age or older are broad, including various dementias, neurological and psychiatric disorders, and bvFTD "phenocopies" with no definable clinical disease.[10–12] To optimize clinical translation, tools should be validated prospectively in cohorts where bvFTD is one of the possible diagnoses out of a range of plausible alternative outcomes. Most studies (61.5%), however, compared bvFTD with AD or healthy controls, thus potentially limiting the clinical relevance (people with typical AD or healthy controls are usually not mistaken for bvFTD). The application of clinical tools to differentiate bvFTD from AD or healthy controls does, nonetheless, offer a proof-of-concept in that if a tool failed to discriminate these groups it could be abandoned. To understand an instrument's true diagnostic value, however, it needs to be validated prospectively in cases of genuine diagnostic uncertainty. Only a small fraction of studies (11.4%) included a cohort of primary psychiatric disorders, although, arguably, this is the main differential for late-onset frontal behavioral change. Notably, very few studies have subjected social

TABLE 7 Other clinical tools

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/ Max score
	Alberta Smell Test (AST)											
[99]	Total score	bvFTD < MDD	21 vs 38	91.0	97.0	79.0	99.0	0.83			94.0	≤2.0/20.0
	Clinical Dementia Rating Scale (CDR)											
[37]	Total score	bvFTD > HC	86 vs 43					1.0				N/A
	FTLD-Modified CDR (CDR-FTLD)											
[37]	Total score	bvFTD > HC	86 vs 43					1.0				N/A
[100]	Spanish version—Memory score	bvFTD > AD/PPA/HC	27 vs 104	85.0	52.0	81.0	58.0	0.68				≥0.8/3.0
[100]	Spanish version—Language score	bvFTD > AD/PPA/HC	27 vs 104	48.0	70.0	80.0	36.0	0.61				≥0.8/3.0
[100]	Spanish version—Behavior score	bvFTD > AD/PPA/HC	27 vs 104	91.0	85.0	94.0	79.0	0.95				≥1.5/3.0
	FTD versus Primary Psychiatric Disorder (FTDvsPPD) Checklist											
[101]	18 item version—Total score	bvFTD > Mx PPD cohort	46 vs 66					0.90				N/A
[101]	17 item version—Total score	bvFTD > Mx PPD cohort	46 vs 66	93.9	71.1	89.2		0.91				≥11.0/17.0
	Screening instrument for frontotemporal dementia (SIFTD)											
[102]	Total score	bvFTD > AD	12 vs 12	83.3	91.7							≥3.0/12.0

Note: See Supporting Information S1 for test references.

Abbreviations: Acc., accuracy; AD, Alzheimer's disease; AUC, area under the curve; bvFTD, behavioral variant frontotemporal dementia; HC, healthy controls; MDD, major depressive disorder; Mx, mixed; N/A, not applicable; NPV, negative predictive value; NLR, negative likelihood ratio; PLR, positive likelihood ratio; PPA, primary progressive aphasia; PPV, positive predictive value; Sens., sensitivity; Spec., specificity.

cognition or behavioral tools to this test. This is problematic, as despite the review's findings favoring social cognition and behavioral tools, most studies assessed these tools against comparator groups where frontal behavioral disturbance, characteristic of bvFTD, are uncommon. When these tools were applied to psychiatric cohorts, the results were far less clear. For example, one study[50] reported that people with bvFTD performed *better* at the Ekman's Faces Test than people with BDI. A key finding of this review, therefore, was that only 5.2% of studies involved appropriate sampling and prospective application of clinical tools in suspected bvFTD cases, covering the broad spectrum of clinical outcomes and thus yielding more clinically meaningful results. The significance of this limitation is exemplified by contrasting the results for specific tests when evaluated in the presence or absence of appropriate sampling and prospective design. In general, when the study design was rated high quality, test performance was found to be weaker. For instance, in the high-quality-rated LOF cohort, the FBI was reported to have a poor AUC of 0.68,[35] whereas other studies with poor quality ratings reported excellent AUC of ≥0.96[83,85,88]; other examples of the same phenomenon include the MMSE (AUC = 0.61 in the LOF cohort[35]; AUC up to 0.96 in studies with poor quality rating[23]) and FAUX-20 (AUC = 0.6 in the LOF cohort,[51] AUC of 0.95 in a study with a poor quality rating[29]).

A further problem in non-prospective studies, though one whose magnitude is impossible to quantify in a review, is that of false positive diagnoses of bvFTD. As mentioned in the introduction, this is a significant issue in bvFTD[6–8] and there is no reason to assume that this would not also be the case in research studies. Only 56.4% of studies mentioned the use of supportive neuroimaging, histopathology, or genetics for diagnosis. People with clear evidence of frontotemporal degeneration on imaging studies, or with known genetic mutations, are arguably less in need of clinical diagnostic tests, although as mentioned already, they can be informative for proof-of-concept studies. Suspected bvFTD with negative markers of neurodegeneration is where clinical tests are most required and yet a mere 12.5% had a follow-up period of at least 2 years, though this is crucial to ensure diagnostic stability for bvFTD.[10] Absence of longitudinal outcome data quite possibly resulted in the inclusion of false positive bvFTD cases, thereby confounding the review's results. This seems especially germane to studies that evaluated behavioral tools. For instance, a participant diagnosed with bvFTD based solely on behavioral symptoms will inherently have high scores on behavioral measures, irrespective of whether they truly have the disease or not.

The results of some studies, even if they reported high accuracy, seemed unlikely to find a role in real-world clinical environments. For

TABLE 8 Combinations of clinical tools.

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/ Max score
[17]	BDST + PF + TMT-B + WCST	bvFTD < HC	35 vs 14					0.07				N/A
	HOT + MET + IGT + REMT + FAUX	bvFTD < HC	35 vs 14					0.98				N/A
	HOT + MET + IGT + REMT	bvFTD < HC	35 vs 14					0.99				N/A
	IGT + HOT	bvFTD < HC	35 vs 14					0.96				N/A
	IGT + FAUX	bvFTD < HC	35 vs 14					0.96				N/A
	IGT + REMT	bvFTD < HC	35 vs 14					0.95				N/A
[29]	MINI-SEA + WHO/UCLA AVLT	bvFTD < AD	22 vs 26					0.95				N/A
[31]	FAB + An in-house bvFTD behavioral questionnaire	bvFTD < AD	35 vs 46	100.0	93.0	94.0	100				97.0	X/X
[40]	ACE-III + IFS	bvFTD < AD	18 vs 33	77.8	90.9			0.91				X/X
	MINI-SEA + ACE-III + IFS	bvFTD < AD	18 vs 33	88.9	100			0.96				X/X
[54]	MINI-SEA + MoCA	bvFTD < AD	20 vs 33	100	93.0			0.99				≤16.7/60.0
	MINI-SEA + MoCA	bvFTD < PD	20 vs 51	100	93.0			0.99				≤15.0/60.0
	MINI-SEA + MoCA	bvFTD < HC	20 vs 29	100	93.0			0.95				≤25.0/60.0
	MINI-SEA + MoCA + IFS	bvFTD < PD	20 vs 51	93.8	100						98.3	X/X
	MINI-SEA + MoCA + IFS	bvFTD < HC	20 vs 29	70.0	93.1						83.7	X/X
	MINI-SEA + IFS	bvFTD < PD	20 vs 51	93.8	100						98.3	X/X
[63]	ACE-R memory + RAVLT + ROCF	bvFTD > AD	23 vs 21					0.92				N/A
	VST orientation + ACE-R memory + RAVLT + ROCF	bvFTD > AD	23 vs 21	91.3	94.4						92.7	X/X
[103]	MMSE + FCSRT free-recall subscore	bvFTD > PPA	35 vs 20	85.0	43.0						69.0	X/X
	MMSE + WSCT	bvFTD > PPA	35 vs 20	91.0	43.0							X/X
	WSCT + FCSRT free-recall subscore	bvFTD > PPA	35 vs 20	85.0	52.0							X/X
	MoCA language subscore + PF	bvFTD > PPA	35 vs 20	79.0	57.0							X/X
	MoCA language subscore + BNT	bvFTD > PPA	35 vs 20	91.0	48.0							X/X
	PF + BNT	bvFTD > PPA	35 vs 20	94.0	33.0							X/X
[104]	Husn Vocabulary Test + Cronholm Verbal Memory Test + Dureman Block Design Test	bvFTD > AD (path dx)	11 vs 17	82.0	94.0						89.0	X/X
		bvFTD > AD (clinic dx)	17 vs 21	76.0	90.0						84.-0	X/X
[105]	ROCF copy + PF + NPI apathy subscore	bvFTD > AD	19 vs 39	73.7	94.7						87.9	X/X
	ROCF copy + PF + NPI apathy subscore	bvFTD > AD	11 vs 23	82.6	81.8						82.3	X/X
[106]	Antisaccade Test + PF + Social Norms Questionnaire + Behavioral Rating Scale	bvFTD > AD	20 vs 24	65.0	79.0						73.0	X/X
[107]	FAB/Perceptual Assessment Battery ratio	bvFTD < AD/ SIVD/HC	23 vs 66	93.0	93.0							<0.83
[108]	PF + ROCF recall + Renzi Apraxia Test + Visual Object and Space Perception battery cube analysis	bvFTD > AD	10 vs 10	70.0	80.0							X/X

(Continues)

TABLE 8 (Continued)

Study	Test	Cohort	Sample size	Sens.	Spec.	PPV	NPV	AUC	PLR	NLR	Acc.	Cut-off/ Max score
[109]	Mattis Dementia Rating Scale memory subscore + PF + Wechsler Intelligence Scale For Children–Revised block design	bvFTD > AD	14 vs 28	76.9	90.9						86.0	X/X
[110]	Penn Emotion Recognition Task + IRI insight	bvFTD < MDD	17 vs 16					0.97				N/A

Note: See Supporting Information S1 for test references.

Abbreviations: Acc., accuracy; ACE-III, Addenbrooke's Cognitive Examination III; ACE-R, Addenbrooke's Cognitive Examination-Revised; AD, Alzheimer's disease; AUC, area under the curve; BDST, backwards digit span test; BNT, Boston Naming Test; bvFTD, behavioral variant frontotemporal dementia; Dx, diagnosis; FAB, Frontal Assessment Battery; FAUX, Faux Pas Test; FCSRT, Free and Cued Selective Reminding Test; FTD, frontotemporal dementia; HC, healthy controls; HOT, Hotel Task; IFS, INECO Frontal Screening; IGT, Iowa Gambling Task; IRI, Interpersonal reactivity index; MET, Multiple Errands Test; MDD, major depressive disorder; MINI-SEA, Mini Social Cognition and Emotional Assessment; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; N/A, not applicable; NLR, negative likelihood ratio; NPI, neuropsychiatric inventory; NPV, negative predictive value; Path, pathological; PD, Parkinson's disease; PF, phonemic fluency; PLR, positive likelihood ratio; PPA, primary progressive aphasia; PPV, positive predictive value; RAVLT, Rey Auditory Verbal Learning Test; REMT, Reading the Mind in the Eyes Test; ROCF, Rey–Osterreith Complex Figure copy task; Sens., sensitivity; SIVD, subcortical ischemic vascular dementia; Spec., specificity; TMT-B, Trail Maker Test Part B; VST, virtual supermarket task; WHO/UCLA AVLT, Modified-World Health Organization/ University of California-Los Angeles Auditory Verbal Learning Test; WSCT, Wisconsin Card Sorting Test.

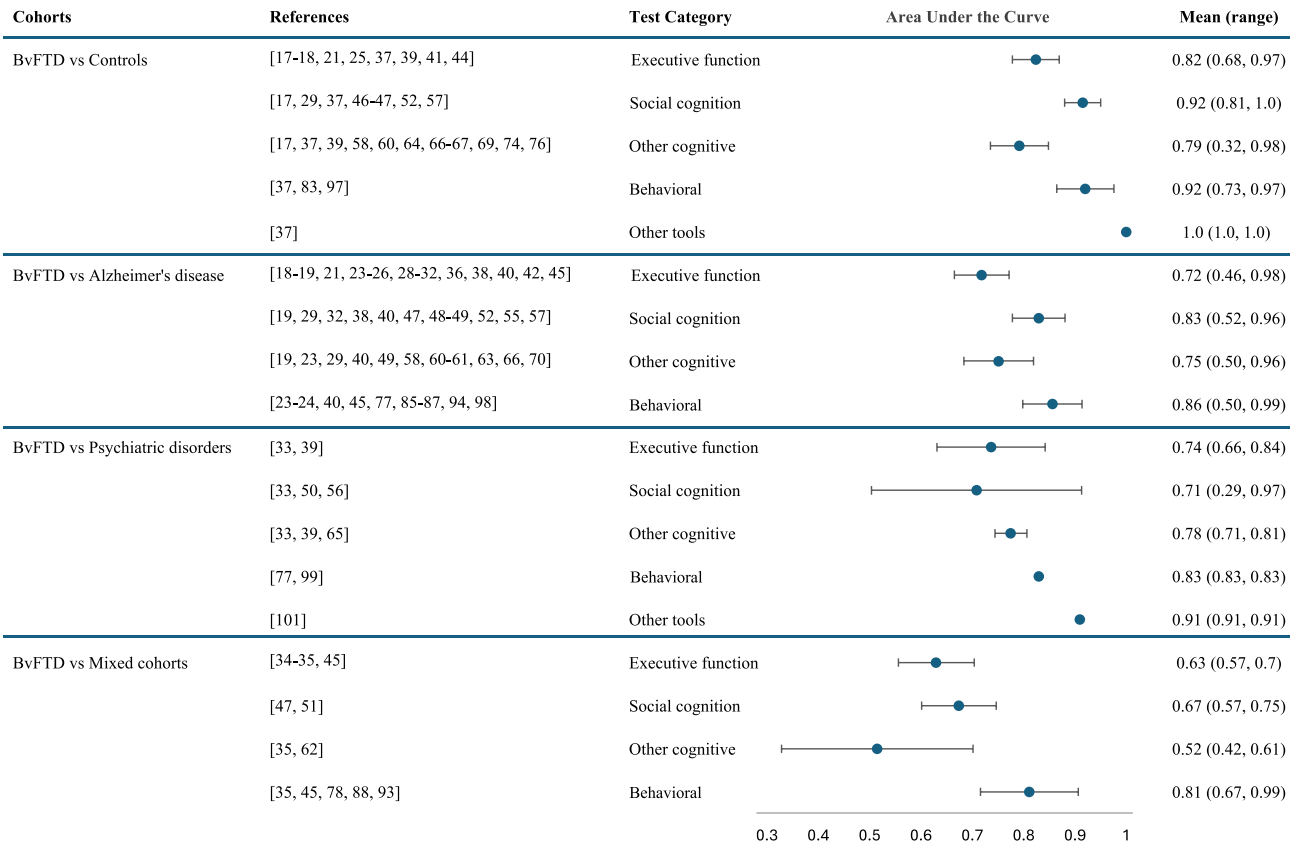

FIGURE 3 Summary of test performance for the main comparator cohorts. This figure depicts the diagnostic accuracy of tests in differentiating behavioral variant frontotemporal dementia (BvFTD) from other groups through Forest plots. Each point represents the area under the curve (AUC) value for tests assessing executive function, social cognition, other cognitive, behavioral, and other clinical tools with the horizontal lines indicating 95% confidence intervals. Mean (range) shows the average and variability of the AUC between studies.

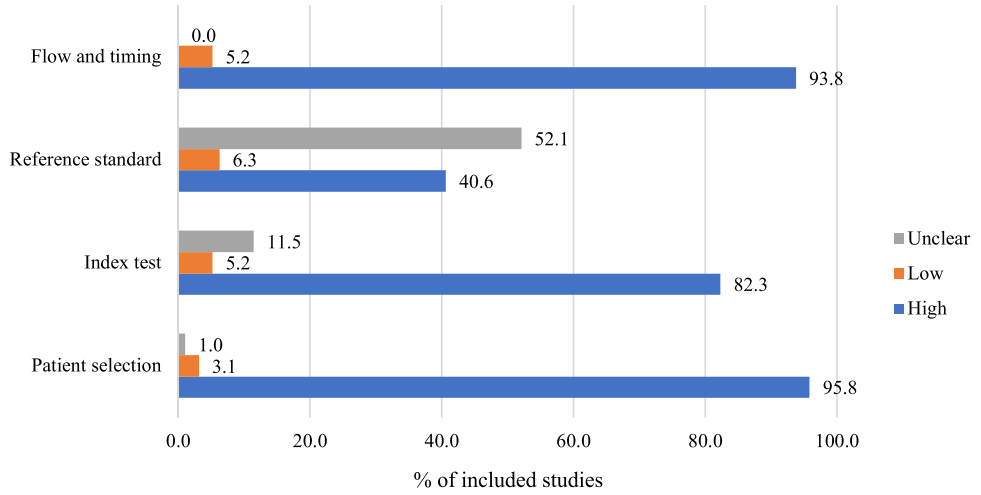

FIGURE 4 Risk of bias. This figure depicts the risk of bias of the included studies according to the Quality Assessment of Diagnostic Accuracy Studies 2 quality appraisal tool.

TABLE 9 Checklist for clinical validation of diagnostic tools in bvFTD

Prospective study design.
Studies in which tests are evaluated and the diagnosis made at the same time are of limited clinical translation as either (i) the diagnosis is already so clear that a new diagnostic test will offer little value or (ii) if not, the diagnosis may be unstable making it an unreliable gold-standard. Cross-sectional studies of this type should be viewed only as proof-of-concept.
Clinical tools should be evaluated as close to first presentation possible.
Inclusion of people with advanced bvFTD where there is little diagnostic uncertainty is of limited value beyond proof-of-concept.
Recruitment should include patients where there is genuine clinical uncertainty but bvFTD is one of the possibilities.
This means including comparator groups where bvFTD is a plausible differential diagnosis; comparing to groups such as amnestic AD or healthy controls where no clinician would ever entertain a diagnosis of bvFTD is of very limited value.
Multiple clinical tools/algorithms should be compared in the same cohort.
There is inevitably a degree of heterogeneity between cohorts making comparison between studies difficult. Comparison of multiple tools in the same cohort offers some mitigation against this problem while also expediting discovery.
Disease severity should be classified, and a subgroup analysis should occur.
To understand heterogeneity in disease severity between cohorts, severity should be assessed by validated measures, such as the CDR-FTLD and symptom duration. Subgroup analyses stratified by severity and illness duration is important to understand where a tool's strengths and weaknesses may lie.
Results should be evaluated in comparison to long-term outcome, with a follow-up of at least 2 years, to ensure diagnostic stability.
This should include imaging evidence of frontotemporal degeneration in the bvFTD cohort and absence thereof in comparator cohort(s).

Abbreviations: AD, Alzheimer's disease; bvFTD, behavioral variant frontotemporal dementia; CDR-FTLD, Frontotemporal Lobar Degeneration-modified Clinical Dementia Rating Scale.

instance, demonstrating that a dementia rating scale (CDR) perfectly discriminated bvFTD from controls[37] seems self-evident and unhelpful when one notes that, by definition, the former had dementia while the latter did not (and thus could not earn a dementia "rating" score). Similarly, reporting that impairment of memory is predictive of bvFTD when compared to controls whereas preservation of memory is predictive of bvFTD compared to AD (see Table 4) seems hard to operationalize. This also raises the important confound of bvFTD being a progressive disease. For example, a patient with bvFTD may transition from preserved to impaired on a test as disease progresses, thus highlighting a major limitation in taking a univariate approach to evaluating test performance. The relationship of disease stage to test accuracy was difficult to assess in this review because of a lack of standardization in defining the former across different studies. Two studies[33,83] did, however, compare mild to more advanced bvFTD. The first study[33] reported that cognitive tools had diminished performance in more advanced bvFTD when the comparator cohort also had a high level of disease severity.

Unsurprisingly, the second study[83] reported that behavioral tools had better performance in more advanced bvFTD when the comparator cohort was behaviorally normal (i.e., healthy controls).

The studies that examined multiple tools within the same cohort[17,19,25,29,33,35,40,51,83] warrant highlighting because limitations in study design were somewhat neutralized—though it does not negate the problem of cohorts being contaminated by false positive bvFTD—by being common to all comparisons. These studies notably indicated that social cognition and behavioral tools were more effective than other clinical tools. A few studies[17,29,31,40,54,63,103–110] further enhanced diagnostic accuracy by integrating various tools into a diagnostic algorithm. The VLOM ratio,[139] as an early application of this multivariate methodology, employed an algorithm that integrated differences across various cognitive domains to differentiate FTD from AD. Five of the studies[17,29,40,54,110] reported that when tests of social cognition were combined with memory, executive function, cognitive batteries, or level of insight, test performance was excellent.

The findings of this review emphasize the importance of robust study design in evaluating clinical tools to differentiate bvFTD from other conditions. It is crucial for these studies to include people with bvFTD whose diagnosis is ultimately confirmed with a high degree of certainty. Moreover, it is paramount that clinical tools undergo evaluation at the initial presentation when diagnostic tests are most needed. This approach minimizes the risk of bias. A clinical tool may pass the hurdle of differentiating bvFTD from healthy controls or AD, but its clinical worth cannot be gauged until it has been tested in a prospective cohort where there is genuine uncertainty about diagnosis. In line with these considerations, we propose a structured checklist for the appraisal of clinical tools in bvFTD trials (Table 9).

CONCLUSION

This systematic review identified cognitive, behavioral, and other clinical tools that have been used to differentiate bvFTD from other conditions. The quality of most studies was low and introduced a high risk of bias, making translation of results into clinical practice extremely challenging. The few studies that were of high quality had a prospective study design, had a follow-up period of 2 years, and applied clinical tools to a cohort where there was genuine diagnostic uncertainty. These studies reported that behavioral tools (e.g., the Frontal Behavioral Inventory) and social cognition tests (e.g., the Ekman's Faces Test) had good test performance in differentiating bvFTD from most cohorts. A striking finding of the review, however, was that the overwhelming majority of test evaluations took a univariate approach. This seems inherently flawed considering that (i) bvFTD is a progressive disease and, as such, will manifest evolving clinical deficits across time; and (ii) the differential diagnosis includes a range of disorders, which themselves exhibit heterogeneity of clinical features. Aside from prioritizing prospective study designs and appropriate sampling, future research should include more emphasis on multivariate data (including algorithms that take account of both impairments and capabilities) and Bayesian approaches to understanding applicability of diagnostic algorithms. In response to these challenges, we have developed a structured checklist for the appraisal of clinical tools in bvFTD research.

AUTHOR CONTRIBUTIONS

Conception and design of the study: Joshua Flavell and Peter John Nestor had equal contribution. *Literature search*: Joshua Flavell and Peter John Nestor jointly developed the search strategy; Joshua Flavell completed the initial article screen; Joshua Flavell and Peter John Nestor reviewed the articles for eligibility criteria; any discrepancies were discussed to meet consensus. *Quality appraisal*: Joshua Flavell completed the initial quality appraisal; Joshua Flavell and Peter John Nestor jointly produced the final quality appraisal. *Data extraction*: Joshua Flavell completed the initial data extraction; Joshua Flavell and Peter John Nestor jointly produced the final results; any discrepancies were discussed to meet consensus. *Drafting the manuscript and figures*: Joshua Flavell and Peter John Nestor had equal contribution. *Revision of the manuscript*: Joshua Flavell and Peter John Nestor had equal contribution.

ACKNOWLEDGMENTS

The authors do not have any acknowledgements. The authors did not receive any funding for this research. Open access publishing facilitated by The University of Queensland, as part of the Wiley - The University of Queensland agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The raw data are available in the Supporting Information S1.

ETHICS APPROVAL STATEMENT

N/A

PATIENT CONSENT STATEMENT

N/A

CLINICAL TRIAL REGISTRATION

N/A

ORCID

Joshua Flavell http://orcid.org/0000-0002-2999-987X

REFERENCES

1. Hogan DB, Jetté N, Fiest KM, Roberts JI, Pearson D, Smith EE, et al. The prevalence and incidence of frontotemporal dementia: a systematic review. Can J Neurol Sci. 2016;43(suppl 1):S96–S109.
2. Ranasinghe KG, Rankin KP, Lobach IV, Kramer JH, Sturm VE, Bettcher BM, et al. Cognition and neuropsychiatry in behavioral variant frontotemporal dementia by disease stage. Neurology. 2016;86(7):600–10.

3. Pose M, Cetkovich M, Gleichgerrcht E, Ibáñez A, Torralva T, Manes F. The overlap of symptomatic dimensions between frontotemporal dementia and several psychiatric disorders that appear in late adulthood. Int Rev Psychiatry. 2013;25(2):159–67.
4. Vijverberg EGB, Schouws S, Meesters PD, Verwijk E, Comijs H, Koene T, et al. Cognitive deficits in patients with neuropsychiatric symptoms: a comparative study between behavioral variant frontotemporal dementia and primary psychiatric disorders. J Clin Psychiatry. 2017;78(8):e940–6.
5. Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain. 2011;134(pt 9):2456–77.
6. Vijverberg EGB, Dols A, Krudop WA, Peters A, Kerssens CJ, van Berckel BNM, et al. Diagnostic accuracy of the frontotemporal dementia consensus criteria in the late-onset frontal lobe syndrome. Dementia Geriatr Cognit Disord. 2016;41(3–4):210–9.
7. Krudop WA, Dols A, Kerssens CJ, Eikelenboom P, Prins ND, Möller C, et al. The pitfall of behavioral variant frontotemporal dementia mimics despite multidisciplinary application of the FTDC criteria. J Alzheimer's Dis. 2017;60(3):959–75. https://doi.org/10.3233/JAD-170608
8. Zapata-Restrepo L, Rivas J, Miranda C, Miller BL, Ibanez A, Allen IE, et al. The psychiatric misdiagnosis of behavioral variant frontotemporal dementia in a Colombian sample. Front Neurol. 2021;12:729381.
9. van Vliet D, de Vugt ME, Bakker C, Pijnenburg YAL, Vernooij-Dassen MJFJ, Koopmans RTCM, et al. Time to diagnosis in young-onset dementia as compared with late-onset dementia. Psychol Med. 2013; 43(2):423–32. https://doi.org/10.1017/S0033291712001122
10. de Boer SCM, Gossinkt F, Krudop W, Vijverberg E, Schouws S, Reus LM, et al. Diagnostic instability over time in the late-onset frontal lobe syndrome: when can we say it's FTD? Am J Geriatr Psychiatry. 2023;31(9):679–90.
11. Perry DC, Datta S, Miller ZA, Rankin KP, Gorno-Tempini ML, Kramer JH, et al. Factors that predict diagnostic stability in neurodegenerative dementia. J Neurol. 2019;266(8):1998–2009.
12. Power C, Lawlor BA. The behavioral variant frontotemporal dementia phenocopy syndrome: a review. J Geriatr Psychiatry Neurol. 2021;34(3):196–208.
13. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA. 2018;319(4):388–96.
14. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36.
15. Luzzi S, Fabi K, Pesallaccia M, Silvestrini M, Provinciali L. Applause sign: is it really specific for Parkinsonian disorders? Evidence from cortical dementias. J NeurolNeurosurg Psychiatry. 2011;82(8):830–3.
16. Schönecker S, Hell F, Bötzel K, Wlasich E, Ackl N, Süßmair C, et al. The applause sign in frontotemporal lobar degeneration and related conditions. J Neurol. 2019;266(2):330–8.
17. Gleichgerrcht E, Torralva T, Roca M, Manes F. Utility of an abbreviated version of the executive and social cognition battery in the detection of executive deficits in early behavioral variant frontotemporal dementia patients. J Int Neuropsychol Soc. 2010;16(4):687–94.
18. Alsemari A, Malloy PF. The Behavioral Dyscontrol Scale in the differential diagnosis of behavioral variant of frontotemporal dementia and Alzheimer disease. Clin Neuropsychol. 2021;35(3):633–42.
19. Buhl C, Stokholm J, Gade A. Clinical utility of short social cognitive tests in early differentiation of behavioral variant frontotemporal dementia from Alzheimer's disease. Dement Geriatr Cogn Dis Extra. 2013;3(1):376–85.
20. Rao SP, Nandi R, Dutt A, Kapur N, Harris JM, Thompson JC, et al. Distinct performance profiles on the Brixton test in frontotemporal dementia. J Neuropsychol. 2021;15(2):162–85.
21. Bertoux M, de Souza LC, Zamith P, Dubois B, Bourgeois-Gironde S. Discounting of future rewards in behavioural variant frontotemporal dementia and Alzheimer's disease. Neuropsychology. 2015;29(6):933–9.
22. De Icaza Valenzuela MM, Bak TH, Thompson HE, Colville S, Pal S, Abrahams S. Validation of The Edinburgh Cognitive and Behavioural ALS screen (ECAS) in behavioural variant frontotemporal dementia and Alzheimer's disease. Int J Geriatr Psychiatry. 2021;36(10):1576–87.
23. Hooten WM, Lyketsos CG. Differentiating Alzheimer's disease and frontotemporal dementia: receiver operator characteristic curve analysis of four rating scales. Dementia Geriatr Cognit Disord. 1998;9(3):164–74.
24. Mariano LI, O'Callaghan C, Guimarães HC, Gambogi LB, da Silva TBL, Yassuda MS, et al. Disinhibition in frontotemporal dementia and Alzheimer's disease: a neuropsychological and behavioural investigation. J Int Neuropsychol Soc. 2020;26(2):163–71.
25. Bahia VS, Cecchini MA, Cassimiro L, Viana R, Lima-Silva TB, de Souza LC, et al. The accuracy of INECO frontal screening in the diagnosis of executive dysfunction in frontotemporal dementia and alzheimer disease. Alzheimer Dis Assoc Disord. 2018;32(4):314–9.
26. Slachevsky A, Villalpando JM, Sarazin M, Hahn-Barma V, Pillon B, Dubois B. Frontal Assessment Battery and differential diagnosis of frontotemporal dementia and Alzheimer disease. Arch Neurol. 2004;61(7):1104–7.
27. Nakaaki S, Murata Y, Sato J, Shinagawa Y, Matsui T, Tatsumi H, et al. Reliability and validity of the Japanese version of the Frontal Assessment Battery in patients with the frontal variant of frontotemporal dementia. Psychiatry Clin Neurosci. 2007;61(1):78–83.
28. Custodio N, Herrera-Perez E, Lira D, Roca M, Manes F, Báez S, et al. Evaluation of the INECO Frontal Screening and the Frontal Assessment Battery in Peruvian patients with Alzheimer's disease and behavioral variant Frontotemporal dementia. eNeurologicalSci. 2016;5:25–9.
29. Wang F, Zhou A, Wei C, Zuo X, Ma X, Zhao L, et al. Good performance of the Chinese version of Mini Social Cognition and Emotional Assessment in the early diagnosis of behavioral variant frontotemporal dementia. Front Neurol. 2022;13:827945.
30. Gleichgerrcht E, Roca M, Manes F, Torralva T. Comparing the clinical usefulness of the Institute of Cognitive Neurology (INECO) Frontal Screening (IFS) and the Frontal Assessment Battery (FAB) in frontotemporal dementia. J Clin Exp Neuropsychol. 2011;33(9):997–1004.
31. Valverde AH, Jimenez-Escrig A, Gobernado J, Barón M. A short neuropsychologic and cognitive evaluation of frontotemporal dementia. Clin Neurol Neurosurg. 2009;111(3):251–5.
32. Bertoux M, Funkiewiez A, O'Callaghan C, Dubois B, Hornberger M. Sensitivity and specificity of ventromedial prefrontal cortex tests in behavioral variant frontotemporal dementia. Alzheimer's Dementia. 2013;9(5 suppl):S84–94.
33. Bertoux M, Delavest M, de Souza LC, Funkiewiez A, Lépine JP, Fossati P, et al. Social Cognition and Emotional Assessment differentiates frontotemporal dementia from depression. J Neurology Neurosurg Psychiatry. 2012;83(4):411–6.

34. Larner A. Can the Frontal Assessment Battery (FAB) help in the diagnosis of behavioural variant frontotemporal dementia? A pragmatic study. Int J Geriatr Psychiatry. 2013;28(1):106–7.
35. Krudop WA, Kerssens CJ, Dols A, Prins ND, Möller C, Schouws S, et al. Identifying bvFTD within the wide spectrum of late onset frontal lobe syndrome: a clinical approach. Am J Geriatr Psychiatry. 2015;23(10):1056–66.
36. Leslie FVC, Foxe D, Daveson N, Flannagan E, Hodges JR, Piguet O. FRONTIER Executive Screen: a brief executive battery to differentiate frontotemporal dementia and Alzheimer's disease. J Neurol Neurosurg Psychiatry. 2016;87(8):831–5.
37. Schroeter ML, Pawelke S, Bisenius S, Kynast J, Schuemberg K, Polyakova M, et al. A modified reading the mind in the eyes test predicts behavioral variant frontotemporal dementia better than executive function tests. Front Aging Neurosci. 2018;10:11.
38. Torralva T, Roca M, Gleichgerrcht E, López P, Manes F. INECO Frontal Screening (IFS): a brief, sensitive, and specific tool to assess executive functions in dementia—ERRATUM. J Int Neuropsychol Soc. 2010;16(5):737–47.
39. Fiorentino N, Gleichgerrcht E, Roca M, Cetkovich M, Manes F, Torralva T. The INECO Frontal Screening tool differentiates behavioral variant—frontotemporal dementia (bv-FTD) from major depression. Dement Neuropsychol. 2013;7(1):33–9.
40. Custodio N, Montesinos R, Cruzado L, Herrera-Perez E, Failoc-Rojas VE, Pintado-Caipa M, et al. Social cognition and behavioral assessments improve the diagnosis of behavioral variant of frontotemporal dementia in older peruvians with low educational levels. Front Neurol. 2021;12:704109.
41. Torralva T, Roca M, Gleichgerrcht E, Bekinschtein T, Manes F. A neuropsychological battery to detect specific executive and social cognitive impairments in early frontotemporal dementia. Brain. 2009;132(pt 5):1299–309.
42. Garcin B, Volle E, Funkiewiez A, Miller BL, Dubois B, Levy R. A mosquito bites and a butterfly flies: a specific response type of frontal patients in a similarity task. Neuropsychologia. 2018;117:371–8.
43. Plutino A, Camerucci E, Ranaldi V, Baldinelli S, Fiori C, Silvestrini M, et al. Insight in frontotemporal dementia and progressive supranuclear palsy. Neurol Sci. 2020;41(8):2135–42.
44. Vale-Britto PHF, Rabin L, Spindola L, Nitrini R, Brucki SMD. Assessment of judgment ability in a Brazilian sample of patients with mild cognitive impairment and dementia. Dement Neuropsychol. 2021;15(2):200–9.
45. Lane CM, Lee M, Lowe J, Bolton C, Pyykkonen BA. Utility of empathy informant report in FTD differential diagnosis. J Int Neuropsychol Soc. 2022;29(7):670–6.
46. Diehl-Schmid J, Pohl C, Ruprecht C, Wagenpfeil S, Foerstl H, Kurz A. The Ekman 60 Faces Test as a diagnostic instrument in frontotemporal dementia. Arch Clin Neuropsychol. 2007;22(4):459–64.
47. Dodich A, Crespi C, Santi GC, Luzzi S, Ranaldi V, Iannaccone S, et al. diagnostic accuracy of affective social tasks in the clinical classification between the behavioral variant of frontotemporal dementia and other neurodegenerative disease. J Alzheimer's Dis. 2021;80(4):1401–11.
48. Mariano LI, Caramelli P, Guimarães HC, Gambogi LB, Moura MVB, Yassuda MS, et al. Can social cognition measurements differentiate behavioral variant frontotemporal dementia from Alzheimer's disease regardless of apathy? J Alzheimer's Dis. 2020;74(3):817–27.
49. Bertoux M, de Souza LC, O'Callaghan C, Greve A, Sarazin M, Dubois B, et al. Social cognition deficits: the key to discriminate behavioral variant frontotemporal dementia from Alzheimer's disease regardless of amnesia? J Alzheimer's Dis. 2016;49(4):1065–74.
50. Barbosa IG, Leite FMC, Bertoux M, Guimarães HC, Mariano LI, Gambogi LB, et al. Social cognition across bipolar disorder and behavioral variant frontotemporal dementia: an exploratory study. Braz J Psychiatry. 2023;11(2):132–6.
51. Gossink F, Schouws S, Krudop W, Scheltens P, Stek M, Pijnenburg Y, et al. Social Cognition Differentiates Behavioral Variant Frontotemporal Dementia From Other Neurodegenerative Diseases and Psychiatric Disorders. Am J Geriatr Psychiatry. 2018;26(5):569–79.
52. Jiskoot LC, Poos JM, Vollebergh ME, Franzen S, van Hemmen J, Papma JM, et al. Emotion recognition of morphed facial expressions in presymptomatic and symptomatic frontotemporal dementia, and Alzheimer's dementia. J Neurol. 2021;268(1):102–13.
53. Delbeuck X, Pollet M, Pasquier F, Bombois S, Moroni C. The clinical value of the Faux Pas Test for diagnosing behavioral-variant frontotemporal dementia. J Geriatr Psychiatry Neurol. 2022;35(1):62–5.
54. Ibañez A, Fittipaldi S, Trujillo C, Jaramillo T, Torres A, Cardona JF, et al. Predicting and characterizing neurodegenerative subtypes with multimodal neurocognitive signatures of social and cognitive processes. J Alzheimer's Dis. 2021;83(1):227–48.
55. Teichmann M, Daigmorte C, Funkiewiez A, Sanches C, Camus M, Mauras T, et al. Moral emotions in frontotemporal dementia. J Alzheimer's Dis. 2019;69(3):887–96.
56. Chiu I, Piguet O, Diehl-Schmid J, Riedl L, Beck J, Leyhe T, et al. Dissociation in rating negative facial emotions between behavioral variant frontotemporal dementia and major depressive disorder. Am J Geriatr Psychiatry. 2016;24(11):1017–27.
57. Funkiewiez A, Bertoux M, de Souza LC, Lévy R, Dubois B. The SEA (Social Cognition and Emotional Assessment): a clinical neuropsychological tool for early diagnosis of frontal variant of frontotemporal lobar degeneration. Neuropsychology. 2012;26(1):81–90.
58. Amaral-Carvalho V, Lima-Silva TB, Mariano LI, de Souza LC, Guimarães HC, Bahia VS, et al. Brazilian version of Addenbrooke's Cognitive Examination-Revised in the differential diagnosis of Alzheimer'S disease and behavioral variant frontotemporal dementia. Arch Clin Neuropsychol. 2022;37(2):437–48.
59. Lemos R, Duro D, Simoes MR, Santana I. The free and cued selective reminding test distinguishes frontotemporal dementia from Alzheimer's disease. Arch Clin Neuropsychol. 2014;29(7):670–9.
60. Ricci M, Graef S, Blundo C, Miller LA. Using the Rey Auditory Verbal Learning Test (RAVLT) to differentiate alzheimer's dementia and behavioural variant fronto-temporal dementia. Clin Neuropsychol. 2012;26(6):926–41.
61. Kloth N, Lemke J, Wiendl H, Meuth SG, Duning T, Johnen A. Serial position effects rapidly distinguish Alzheimer's from frontotemporal dementia. J Neurol. 2020;267(4):975–83.
62. Perri R, Fadda L, Caltagirone C, Carlesimo GA. Word list and story recall elicit different patterns of memory deficit in patients with Alzheimer's disease, frontotemporal dementia, subcortical ischemic vascular disease, and Lewy body dementia. J Alzheimer's Dis. 2013;37(1):99–107.
63. Tu S, Wong S, Hodges JR, Irish M, Piguet O, Hornberger M. Lost in spatial translation—a novel tool to objectively assess spatial disorientation in Alzheimer's disease and frontotemporal dementia. Cortex. 2015;67:83–94.
64. Graviotto HG, Sorbara MG, Rodriguez CMT, Serrano C. 12-item version of Boston Naming Test: usefulness in the diagnosis of primary progressive aphasia, frontotemporal dementia, and Alzheimer's disease. Dement Neuropsychol. 2022;16(2):181–6.
65. Overbeek JM, Korten N, Gossink F, Fieldhouse J, van de Beek M, Reus L, et al. The value of neuropsychological assessment in the differentiation between behavioral variant frontotemporal

dementia and late-onset psychiatric disorders. J Clin Psychiatry. 2020;81(1):19m12811. https://www.psychiatrist.com/toc/pa-volume/81/2020/1/jcp/

66. Jiskoot LC, Poos JM, van Boven K, de Boer L, Giannini LAA, Satoer DD, et al. The ScreeLing: detecting semantic, phonological, and syntactic deficits in the clinical subtypes of frontotemporal and Alzheimer's dementia. Assessment. 2023;30:2545–59.
67. Duro D, Tábuas-Pereira M, Freitas S, Santiago B, Botelho MA, Santana I. Validity and clinical utility of Different Clock Drawing Test Scoring Systems in multiple forms of dementia. J Geriatr Psychiatry Neurol. 2018;31(3):114–22.
68. Gasparini M, Masciarelli G, Vanacore N, Ottaviani D, Salati E, Talarico G, et al. A descriptive study on constructional impairment in frontotemporal dementia and Alzheimer's disease. Eur J Neurol. 2008;15(6):589–97.
69. Johnen A, Tokaj A, Kirschner A, Wiendl H, Lueg G, Duning T, et al. Apraxia profile differentiates behavioural variant frontotemporal from Alzheimer's dementia in mild disease stages. J Neurol Neurosurg Psychiatry. 2015;86(7):809–15.
70. Johnen A, Frommeyer J, Modes F, Wiendl H, Duning T, Lohmann H. Dementia Apraxia Test (DATE): a brief tool to differentiate behavioral variant frontotemporal dementia from Alzheimer's dementia based on apraxia profiles. J Alzheimer's Dis. 2015;49(3):593–605.
71. Bier JC, Ventura M, Donckels V, Van Eyll E, Claes T, Slama H, et al. Is the Addenbrooke's cognitive examination effective to detect frontotemporal dementia? J Neurol. 2004;251(4):428–31.
72. Yoshida H, Terada S, Honda H, Ata T, Takeda N, Kishimoto Y, et al. Validation of Addenbrooke's cognitive examination for detecting early dementia in a Japanese population. Psychiatry Res. 2011;185(1–2):211–4.
73. Torralva T, Roca M, Gleichgerrcht E, Bonifacio A, Raimondi C, Manes F. Validación de la versión en español del Addenbrooke's Cognitive Examination-Revisado (ACE-R). Neurología. 2011;26(6):351–6.
74. Bruno D, Slachevsky A, Fiorentino N, Rueda DS, Bruno G, Tagle AR, et al. Validación argentino-chilena de la versión en español del test Addenbrooke's Cognitive Examination III para el diagnóstico de demencia. Neurología. 2020;35(2):82–8.
75. Elamin M, Holloway G, Bak TH, Pal S. The utility of the Addenbrooke's Cognitive Examination Version Three in early-onset dementia. Dementia Geriatr Cognit Disord. 2016;41(1–2):9–15.
76. Freitas S, Simões MR, Alves L, Duro D, Santana I. Montreal Cognitive Assessment (MoCA): validation study for frontotemporal dementia. J Geriatr Psychiatry Neurol. 2012;25(3):146–54.
77. Semenkova A, Piguet O, Johnen A, Schroeter ML, Godulla J, Linnemann C, et al. The Behavioural Dysfunction Questionnaire discriminates behavioural variant frontotemporal dementia from Alzheimer's disease dementia and major depressive disorder. J Neurol. 2023;270:3433–41.
78. Lebert F, Pasquier F, Souliez L, Petit H. Frontotemporal behavioral scale. Alzheimer Dis Assoc Disord. 1998;12(4):335–9.
79. Mendez MF, Perryman KM, Miller BL, Cummings JL. Behavioral differences between frontotemporal dementia and Alzheimer's disease: a comparison on the BEHAVE-AD rating scale. Int Psychogeriatr. 1998;10(2):155–62.
80. Bozeat S, Gregory CA, Ralph MA, Hodges JR. Which neuropsychiatric and behavioural features distinguish frontal and temporal variants of frontotemporal dementia from Alzheimer's disease? J Neurol Neurosurg Psychiatry. 2000;69(2):178–86.
81. Lehingue E, Gueniat J, Jourdaa S, Hardouin JB, Pallardy A, Courtemanche H, et al. Improving the diagnosis of the frontal variant of Alzheimer's disease with the DAPHNE scale. J Alzheimer's Dis. 2021;79(4):1735–45.
82. Boutoleau-Bretonnière C, Evrard C, Hardouin JB, Rocher L, Charriau T, Etcharry-Bouyx F, et al. DAPHNE: a new tool for the assessment of the behavioral variant of frontotemporal dementia. Dement Geriatr Cogn Dis Extra. 2015;5(3):503–16.
83. Cui Y, Liu L, Xie K, Chen Z, Nan H, Kong Y et al. Application of the mild behavioral impairment checklist in Chinese patients with the behavioral variant of frontotemporal dementia. Neurol Sci. 2023;45:557–64.
84. Kertesz A, Nadkarni N, Davidson W, Thomas AW. The Frontal Behavioral Inventory in the differential diagnosis of frontotemporal dementia. J Int Neuropsychol Soc. 2000;6(4):460–8.
85. Kertesz A, Davidson W, McCabe P, Munoz D. Behavioral quantitation is more sensitive than cognitive testing in frontotemporal dementia. Alzheimer Dis Assoc Disord. 2003;17(4):223–9.
86. Cerami C, Perdixi E, Meli C, Marcone A, Zamboni M, Iannaccone S, et al. Early identification of different behavioral phenotypes in the behavioral variant of frontotemporal dementia with the aid of the Mini-Frontal Behavioral Inventory (mini-FBI). J Alzheimer's Dis. 2022;89(1):299–308.
87. Suhonen NM, Hallikainen I, Hänninen T, Jokelainen J, Krüger J, Hall A, et al. The Modified Frontal Behavioral Inventory (FBI-mod) for patients with frontotemporal lobar degeneration, alzheimer's disease, and mild cognitive impairment. J Alzheimer's Dis. 2017;56(4):1241–51.
88. Milan G, Lamenza F, Iavarone A, Galeone F, Lorè E, de Falco C, et al. Frontal Behavioural Inventory in the differential diagnosis of dementia. Acta Neurol Scand. 2008;117(4):260–5.
89. Malloy P, Tremont G, Grace J, Frakey L. The Frontal Systems Behavior Scale discriminates frontotemporal dementia from Alzheimer's disease. Alzheimer's Dementia. 2007;3(3):200–3.
90. De Deyn PP, Engelborghs S, Saerens J, Goeman J, Mariën P, Maertens K, et al. The Middelheim Frontality Score: a behavioural assessment scale that discriminates frontotemporal dementia from Alzheimer's disease: The MFS: A Behavioural Assessment Scale That Discriminates FTD from AD. Int J Geriatr Psychiatry. 2005;20(1):70–9.
91. Levy ML, Miller BL, Cummings JL, Fairbanks LA, Craig A. Alzheimer disease and frontotemporal dementias. Behavioral distinctions. Arch Neurol. 1996;53(7):687–90.
92. Bahia VS, Viana R. Accuracy of neuropsychological tests and the Neuropsychiatric Inventory in differential diagnosis between frontotemporal dementia and Alzheimer's disease. Dement Neuropsychol. 2009;3(4):332–6.
93. Gustafson L, Englund E, Brunnström H, Brun A, Erikson C, Warkentin S, et al. The accuracy of short clinical rating scales in neuropathologically diagnosed dementia. Am J Geriatr Psychiatry. 2010;18(9):810–20.
94. Karve S, Barsuglia JP, Mather MJ, Jimenez EE, Shapira J, Mendez MF. Evaluation of emotional blunting in behavioral variant frontotemporal dementia compared to Alzheimer's disease. Dementia Geriatr Cognit Disord. 2014;38(1–2):79–88.
95. Mendez MF, McMurtray A, Licht E, Shapira JS, Saul RE, Miller BL. The scale for emotional blunting in patients with frontotemporal dementia. Neurocase. 2006;12(4):242–6.
96. Phan TX, Reeder JE, Keener LC, Considine CM, Zald DH, Claassen DO, et al. Measuring antisocial behaviors in behavioral variant frontotemporal dementia with a novel informant-based questionnaire. J Neuropsychiatry Clin Neurosci. 2023;35(4):374–84.
97. Rankin KP, Toller G, Gavron L, La Joie R, Wu T, Shany-Ur T, et al. Social behavior observer checklist: patterns of spontaneous behaviors differentiate patients with neurodegenerative disease from healthy older adults. Front Neurol. 2021;12:683162.
98. Barsuglia JP, Kaiser NC, Wilkins SS, Karve SJ, Barrows RJ, Paholpak P, et al. A scale of socioemotional dysfunction in

frontotemporal dementia. Arch Clin Neuropsychol. 2014;29(8):793–805.
99. Heyanka DJ, Golden CJ, McCue RB, Scarisbrick DM, Linck JF, Zlatkin NI. Olfactory deficits in frontotemporal dementia as measured by the Alberta Smell Test. Appl Neuropsychol Adult. 2014;21(3):176–82.
100. Russo G, Russo MJ, Buyatti D, Chrem P, Bagnati P, Suarez MF, et al. Utility of the Spanish version of the FTLD-modified CDR in the diagnosis and staging in frontotemporal lobar degeneration. J Neurol Sci. 2014;344(1–2):63–8.
101. Ducharme S, Pearl-Dowler L, Gossink F, McCarthy J, Lai J, Dickerson BC, et al. The Frontotemporal Dementia versus Primary Psychiatric Disorder (FTD versus PPD) checklist: a bedside clinical tool to identify behavioral variant FTD in patients with late-onset behavioral changes. J Alzheimer's Dis. 2019;67(1):113–24.
102. Gregory CA, Orrell M, Sahakian B, Hodges JR. Can frontotemporal dementia and Alzheimer's disease be differentiated using a brief battery of tests? Int J Geriatr Psychiatry. 1997;12(3):375–83.
103. Velilla L, Hernández J, Giraldo-Chica M, Guzmán-Vélez E, Quiroz Y, Lopera F. A Spanish neuropsychological battery discriminates between the behavioral variant of frontotemporal dementia and primary progressive aphasia in a colombian sample. Front Neurol. 2021;12:656478.
104. Elfgren C, Brun A, Gustafson L, Johanson A, Minthon L, Passant U, et al. Neuropsychological tests as discriminators between dementia of alzheimer-type and frontotemporal dementia. Int J Geriatr Psychiatry. 1994;9(8):635–42.
105. Perri R, Koch G, Carlesimo GA, Serra L, Fadda L, Pasqualetti P, et al. Alzheimer's disease and frontal variant of frontotemporal dementia—a very brief battery for cognitive and behavioural distinction. J Neurol. 2005;252(10):1238–44.
106. Possin KL, Feigenbaum D, Rankin KP, Smith GE, Boxer AL, Wood K, et al. Dissociable executive functions in behavioral variant frontotemporal and Alzheimer dementias. Neurology. 2013;80(24):2180–5.
107. Mendez MF, McMurtray AM, Licht EA, Saul RE. Frontal-executive versus posterior-perceptual mental status deficits in early-onset dementias. Am J Alzheimer's Dis Other Demen. 2009;24(3):220–7.
108. Siri S, Benaglio I, Frigerio A, Binetti G, Cappa SF. A brief neuropsychological assessment for the differential diagnosis between frontotemporal dementia and Alzheimer's disease. Eur J Neurol. 2001;8(2):125–32.
109. Rascovsky K, Salmon DP, Ho GJ, Galasko D, Peavy GM, Hansen LA, et al. Cognitive profiles differ in autopsy-confirmed frontotemporal dementia and AD. Neurology. 2002;58(12):1801–8.
110. Lichtenstein ML, Stewart PV, Kirchner HL, Fenny G, Feldman HH. Exploring social cognition tests to differentiate frontotemporal dementia from depression: a two-step pilot study. Alzheimer Dis Assoc Disord. 2022;37(2):145–8.
111. Krudop WA, Kerssens CJ, Dols A, Prins ND, Möller C, Schouws S, et al. Building a new paradigm for the early recognition of behavioral variant frontotemporal dementia: late onset Frontal Lobe Syndrome study. Am J Geriatr Psychiatry. 2014;22(7):735–40.
112. The Lund and Manchester Groups. Clinical and neuropathological criteria for frontotemporal dementia. J Neurol Neurosurg Psychiatry. 1994;57(4):416–8.
113. Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology. 1998;51(6):1546–54.
114. American Psychiatric Association and D.S.M.T.F. Diagnostic and statistical manual of mental disorders, DSM-5. 5th ed. Arlington, VA: American Psychiatric Association; 2013.
115. McKhann GM, Albert MS, Grossman M, Miller B, Dickson D, Black S, et al. Clinical and pathological diagnosis of frontotemporal dementia: report of the Work Group on Frontotemporal Dementia and Pick's Disease. Arch Neurol. 2001;58(11):1803–9.
116. Jackson M, Lowe J. The new neuropathology of degenerative frontotemporal dementias. Acta Neuropathol. 1996;91(2):127–34.
117. Wallin A, Brun A, Gustafson L. Swedish consensus on dementia diseases. Acta Neurol Scand Suppl. 1994;157:1–31.
118. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology. 1993;43(11):2412–4.
119. Knopman DS, Kramer JH, Boeve BF, Caselli RJ, Graff-Radford NR, Mendez MF, et al. Development of methodology for conducting clinical trials in frontotemporal lobar degeneration. Brain. 2008;131(pt 11):2957–68.
120. Folstein MF, Folstein SE, McHugh PR. Mini-mental state. J Psychiatr Res. 1975;12(3):189–98.
121. Shigenobu K, Ikeda M, Fukuhara R, Maki N, Hokoishi K, Nebu A, et al. The Stereotypy Rating Inventory for frontotemporal lobar degeneration. Psychiatry Res. 2002;110(2):175–87.
122. Kertesz A, Davidson W, Fox H. Frontal Behavioral Inventory: diagnostic criteria for frontal lobe dementia. Can J Neurol Sci. 1997;24(1):29–36.
123. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a Frontal Assessment Battery at bedside. Neurology. 2000;55(11):1621–6.
124. Stone VE, Baron-Cohen S, Knight RT. Frontal lobe contributions to theory of mind. J Cogn Neurosci. 1998;10(5):640–56.
125. Ekman P, Friesen WV. Pictures of facial affect. Palo Alto, CA: Consulting Psychologists Press; 1976.
126. Kaplan E, Goodglass H, Weintraub S. Boston naming test. 2nd ed. Philadelphia: Lea & Febiger; 2001.
127. Reul S, Lohmann H, Wiendl H, Duning T, Johnen A. Can cognitive assessment really discriminate early stages of Alzheimer's and behavioural variant frontotemporal dementia at initial clinical presentation? Alzheimer's Res Ther. 2017;9(1):61.
128. Engelborghs S, Maertens K, Mariën P, Vloeberghs E, Somers N, Nagels G, et al. Behavioural and neuropsychological correlates of frontal lobe features in dementia. Psychol Med. 2006;36(8):1173–82.
129. McLaughlin N, Westervelt H. Odor identification deficits in frontotemporal dementia: a preliminary study. Arch Clin Neuropsychol. 2008;23(1):119–23.
130. Torralva T, Roca M, Gleichgerrcht E, López P, Manes F. INECO Frontal Screening (IFS): a brief, sensitive, and specific tool to assess executive functions in dementia—corrected version. J Int Neuropsychol Soc. 2009;15(5):777–86.
131. Rabin LA, Borgos MJ, Saykin AJ, Wishart HA, Crane PK, Nutter-Upham KE, et al. Judgment in older adults: development and psychometric evaluation of the Test of Practical Judgment (TOP-J). J Clin Exp Neuropsychol. 2007;29(7):752–67.
132. Partington JE, Leiter RG. Partington's Pathway Test. Psychol Center Bull. 1949;1:9–20.
133. Rey A. L'examen psychologique dans les cas d'encéphalopathie traumatique.(Les problems.). Arch psychol. 1941;28:215–85.
134. Carvalho VA, Caramelli P. Brazilian adaptation of the Addenbrooke's Cognitive Examination-Revised (ACE-R). Dement Neuropsychol. 2007;1(2):212–6.
135. Doesborgh SJC, van de Sandt-Koenderman WME, Dippel DWJ, van Harskamp F, Koudstaal PJ, Visch-Brink EG. Linguistic deficits in the acute phase of stroke. J Neurol. 2003;250(8):977–82.
136. Crawford JH, Henry JD, Assessment of executive dysfunction. In: Halligan PW, Wade DT, editors. The effectiveness of rehabilitation for cognitive deficits. Oxford: Oxford University Press; 2005;233–46.
137. Weiss PH, Kalbe E, Kessler J, Fink GR. Kölner Apraxie Screening. Göttingen: Hogrefe Verlag; 2013.
138. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment,

MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53(4):695–9.
139. Mathuranath PS, Nestor PJ, Berrios GE, Rakowicz W, Hodges JR. A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. Neurology. 2000;55(11):1613–20.
140. Hsieh S, Schubert S, Hoon C, Mioshi E, Hodges JR. Validation of the Addenbrooke's Cognitive Examination III in frontotemporal dementia and Alzheimer's disease. Dementia Geriatr Cognit Disord. 2013;36(3–4):242–50.
141. Abrahams S, Newton J, Niven E, Foley J, Bak TH. Screening for cognition and behaviour changes in ALS. Amyotroph Lateral Scler Frontotemporal Degener. 2014;15(1–2):9–14.
142. Davis MH. A multidimensional approach to individual differences in empathy. JSAS Catalog Sel. Doc. Psychol. 1980;10:85.
143. Heidler-Gary J, Gottesman R, Newhart M, Chang S, Ken L, Hillis AE. Utility of behavioral versus cognitive measures in differentiating between subtypes of frontotemporal lobar degeneration and Alzheimer's disease. Dementia Geriatr Cognit Disord. 2007;23(3):184–93.
144. Grace J, Malloy PH, Frontal systems behavior scale (FrSBe): Professional manual. Lutz, FL: PAR; 2001.
145. Gustafson L, Nilsson L. Differential diagnosis of presenile dementia on clinical grounds. Acta Psychiatr Scand. 1982;65(3):194–209.
146. Barber R, Snowden JS, Craufurd D. Frontotemporal dementia and Alzheimer's disease: retrospective differentiation using information from informants. J Neurol Neurosurg Psychiatry. 1995;59(1):61–70.
147. Ismail Z, Agüera-Ortiz L, Brodaty H, Cieslak A, Cummings J, Fischer CE, et al. The Mild Behavioral Impairment Checklist (MBI-C): a rating scale for neuropsychiatric symptoms in pre-dementia populations. J Alzheimer's Dis. 2017;56(3):929–38.
148. Green P. Alberta Smell Test: instructions, record forms, clinical data. Edmonton, AB: Green's Publishing; 1989.
149. Manly T, Hawkins K, Evans J, Woldt K, Robertson IH. Rehabilitation of executive function: facilitation of effective goal management on complex tasks using periodic auditory alerts. Neuropsychologia. 2002;40(3):271–81.
150. Knight C, Alderman N, Burgess PW. Development of a simplified version of the multiple errands test for use in hospital settings. Neuropsychol Rehabil. 2002;12(3):231–55.
151. Bechara A, Damasio AR, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. Cognition. 1994;50(1–3):7–15.
152. Baron-Cohen S, Jolliffe T, Mortimore C, Robertson M. Another advanced test of theory of mind: evidence from very high functioning adults with autism or asperger syndrome. J Child Psychol Psychiatry. 1997;38(7):813–22.
153. Jin HM, Li D, Yu YY, Zhang M, Liu XN, Li FY, et al. [Features of a modified WHO/UCLA AVLT performance in amnestic mild cognitive impairment and mild Alzheimer's disease]. Zhonghua Yi Xue Za Zhi. 2019;99(31):2423–8.
154. Lezak MD. Neuropsychological assessment. USA: Oxford University Press; 2004.
155. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. Neurology. 1994;44(12):2308.
156. Nelson HE. A Modified Card Sorting Test sensitive to frontal lobe defects. Cortex. 1976;12(4):313–24.
157. Wechsler D. Wechsler Intelligence Scales for children. 3rd ed. San Antonio, TX: The Psychological Corporation; 1991.
158. Burgess PW, Shallice T. The Hayling and Brixton Tests. Bury, St Edmunds: Thames Valley Test Company; 1997.
159. Sprengelmeyer R, Young AW, Calder AJ, Karnat A, Lange H, Hömberg V, et al. Loss of disgust. Perception of faces and emotions in Huntington's disease. Brain. 1996;119(pt 5):1647–65.
160. McDonald S, Flanagan S, Rollins J, Kinch J. TASIT: a new clinical tool for assessing social perception after traumatic brain injury. J Head Trauma Rehabil. 2003;18(3):219–38.
161. Gong X, Huang YX, Wang Y, Luo YJ. Revision of the Chinese facial affective picture system. Chinese Mental Health J. 2011;25:40–6.
162. Lennox RD, Wolfe RN. Revision of the self-monitoring scale. J Pers Soc Psychol. 1984;46(6):1349–64.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Flavell J, Nestor PJ. A systematic review of cognitive and behavioral tools to differentiate behavioral variant frontotemporal dementia from other conditions. Psychiatry Clin Neurosci Rep. 2024;3:e210. https://doi.org/10.1002/pcn5.210