Plasma and cerebrospinal fluid ABeta42 for the differential diagnosis of Alzheimer's disease dementia in participants diagnosed with any dementia subtype in a specialist care setting

Cochrane Library

Cochrane Database of Systematic Reviews

Plasma and cerebrospinal fluid ABeta42 for the differential diagnosis of Alzheimer's disease dementia in participants diagnosed with any dementia subtype in a specialist care setting (Review)

Kokkinou M, Beishon LC, Smailagic N, Noel-Storr AH, Hyde C, Ukoumunne O, Worrall RE, Hayen A, Desai M, Ashok AH, Paul EJ, Georgopoulou A, Casoli T, Quinn TJ, Ritchie CW

Kokkinou M, Beishon LC, Smailagic N, Noel-Storr AH, Hyde C, Ukoumunne O, Worrall RE, Hayen A, Desai M, Ashok AH, Paul EJ, Georgopoulou A, Casoli T, Quinn TJ, Ritchie CW.
Plasma and cerebrospinal fluid ABeta42 for the differential diagnosis of Alzheimer's disease dementia in participants diagnosed with any dementia subtype in a specialist care setting.
Cochrane Database of Systematic Reviews 2021, Issue 2. Art. No.: CD010945.
DOI: 10.1002/14651858.CD010945.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Plasma and cerebrospinal fluid ABeta42 for the differential diagnosis of Alzheimer's disease dementia in participants diagnosed with any dementia subtype in a specialist care setting

Michelle Kokkinou[1], Lucy C Beishon[2], Nadja Smailagic[3], Anna H Noel-Storr[4], Chris Hyde[5], Obioha Ukoumunne[6], Rosemary E Worrall[7], Anja Hayen[8], Meera Desai[9], Abhishekh Hulegar Ashok[1,10], Eleanor J Paul[1,11], Aikaterini Georgopoulou[12], Tiziana Casoli[13], Terry J Quinn[14], Craig W Ritchie[15]

[1]MRC London Institute of Medical Sciences, Imperial College London, London, UK. [2]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK. [3]Institute of Public Health, University of Cambridge , Cambridge, UK. [4]Radcliffe Department of Medicine, University of Oxford , Oxford, UK. [5]Exeter Test Group, College of Medicine and Health, University of Exeter Medical School, University of Exeter, Exeter , UK. [6]NIHR CLAHRC South West Peninsula (PenCLAHRC), University of Exeter Medical School, Exeter, UK. [7]University of Oxford Medical School, Oxford, UK. [8]Department of Psychology and Clinical Language Sciences, University of Reading, Reading, UK. [9]Department of Experimental Psychology, University of Oxford, Oxford, UK. [10]Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College , London, UK. [11]Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, London, UK. [12]Department of Surgery and Cancer, Imperial College London, London, UK. [13]Center for Neurobiology of Aging, IRCCS INRCA, Ancona, Italy. [14]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. [15]Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

Contact address: Lucy C Beishon, lb330@le.ac.uk, lucy.beishon@nhs.net.

Editorial group: Cochrane Dementia and Cognitive Improvement Group.
Publication status and date: New, published in Issue 2, 2021.

Citation: Kokkinou M, Beishon LC, Smailagic N, Noel-Storr AH, Hyde C, Ukoumunne O, Worrall RE, Hayen A, Desai M, Ashok AH, Paul EJ, Georgopoulou A, Casoli T, Quinn TJ, Ritchie CW. Plasma and cerebrospinal fluid ABeta42 for the differential diagnosis of Alzheimer's disease dementia in participants diagnosed with any dementia subtype in a specialist care setting. *Cochrane Database of Systematic Reviews* 2021, Issue 2. Art. No.: CD010945. DOI: 10.1002/14651858.CD010945.pub2.

ABSTRACT

Background

Dementia is a syndrome that comprises many differing pathologies, including Alzheimer's disease dementia (ADD), vascular dementia (VaD) and frontotemporal dementia (FTD). People may benefit from knowing the type of dementia they live with, as this could inform prognosis and may allow for tailored treatment. Beta-amyloid (1-42) (ABeta42) is a protein which decreases in both the plasma and cerebrospinal fluid (CSF) of people living with ADD, when compared to people with no dementia. However, it is not clear if changes in ABeta42 are specific to ADD or if they are also seen in other types of dementia. It is possible that ABeta42 could help differentiate ADD from other dementia subtypes.

Objectives

To determine the accuracy of plasma and CSF ABeta42 for distinguishing ADD from other dementia subtypes in people who meet the criteria for a dementia syndrome.

Search methods

We searched MEDLINE, and nine other databases up to 18 February 2020. We checked reference lists of any relevant systematic reviews to identify additional studies.

Selection criteria

We considered cross-sectional studies that differentiated people with ADD from other dementia subtypes. Eligible studies required measurement of participant plasma or CSF ABeta42 levels and clinical assessment for dementia subtype.

Data collection and analysis

Seven review authors working independently screened the titles and abstracts generated by the searches. We collected data on study characteristics and test accuracy. We used the second version of the 'Quality Assessment of Diagnostic Accuracy Studies' (QUADAS-2) tool to assess internal and external validity of results. We extracted data into 2 x 2 tables, cross-tabulating index test results (ABeta42) with the reference standard (diagnostic criteria for each dementia subtype). We performed meta-analyses using bivariate, random-effects models. We calculated pooled estimates of sensitivity, specificity, positive predictive values, positive and negative likelihood ratios, and corresponding 95% confidence intervals (CIs).

In the primary analysis, we assessed accuracy of plasma or CSF ABeta42 for distinguishing ADD from other mixed dementia types (non-ADD). We then assessed accuracy of ABeta42 for differentiating ADD from specific dementia types: VaD, FTD, dementia with Lewy bodies (DLB), alcohol-related cognitive disorder (ARCD), Creutzfeldt-Jakob disease (CJD) and normal pressure hydrocephalus (NPH). To determine test-positive cases, we used the ABeta42 thresholds employed in the respective primary studies. We then performed sensitivity analyses restricted to those studies that used common thresholds for ABeta42.

Main results

We identified 39 studies (5000 participants) that used CSF ABeta42 levels to differentiate ADD from other subtypes of dementia. No studies of plasma ABeta42 met the inclusion criteria. No studies were rated as low risk of bias across all QUADAS-2 domains. High risk of bias was found predominantly in the domains of patient selection (28 studies) and index test (25 studies).

The pooled estimates for differentiating ADD from other dementia subtypes were as follows: ADD from non-ADD: sensitivity 79% (95% CI 0.73 to 0.85), specificity 60% (95% CI 0.52 to 0.67), 13 studies, 1704 participants, 880 participants with ADD; ADD from VaD: sensitivity 79% (95% CI 0.75 to 0.83), specificity 69% (95% CI 0.55 to 0.81), 11 studies, 1151 participants, 941 participants with ADD; ADD from FTD: sensitivity 85% (95% CI 0.79 to 0.89), specificity 72% (95% CI 0.55 to 0.84), 17 studies, 1948 participants, 1371 participants with ADD; ADD from DLB: sensitivity 76% (95% CI 0.69 to 0.82), specificity 67% (95% CI 0.52 to 0.79), nine studies, 1929 participants, 1521 participants with ADD. Across all dementia subtypes, sensitivity was greater than specificity, and the balance of sensitivity and specificity was dependent on the threshold used to define test positivity.

Authors' conclusions

Our review indicates that measuring ABeta42 levels in CSF may help differentiate ADD from other dementia subtypes, but the test is imperfect and tends to misdiagnose those with non-ADD as having ADD. We would caution against the use of CSF ABeta42 alone for dementia classification. However, ABeta42 may have value as an adjunct to a full clinical assessment, to aid dementia diagnosis.

PLAIN LANGUAGE SUMMARY

How accurate is the ABeta42 test for distinguishing Alzheimer's disease from other types of dementia in patients seen in a specialist clinic?

Why is improving dementia diagnosis important?

Dementia is a condition characterised by progressive problems with memory and thinking. Dementia can be caused be a number of different conditions (for example, by Alzheimer's disease), and the best treatments depend on the underlying cause. Levels of the protein ABeta42 in blood or spinal fluid may determine the underlying cause of dementia. This could help clinicians choose the best treatments.

What is the aim of this review?

The aim of this review was to find out how accurate are the levels of ABeta42 in blood or spinal fluid for determining the cause of dementia.

What was studied in the review?

We included studies that examined the levels of ABeta42 taken from samples of blood or spinal fluid. At present, this test is only used in specialist clinics. Levels of ABeta42 may be lower in persons with Alzheimer's dementia compared to those with other types of dementia.

What are the main results of this review?

We included 39 studies with a total of 5000 participants. All studies used spinal fluid tests of ABeta42. None of the included studies used a blood test of ABeta42.

In theory, the results of these studies indicate that if ABeta42 were to be used in a specialist clinic in a group of 1000 people, where 520 (52%) have Alzheimer's dementia, an estimated 602 would have an ABeta42 result. This would indicate that Alzheimer's dementia is present. Of

these, 192 (32%) would be incorrectly classified as having Alzheimer's disease. Of the 398 people with a result indicating that Alzheimer's disease is not present, 110 (28%) would be incorrectly classified as not having Alzheimer's disease. The included studies used different levels of ABeta42 to make the diagnosis of Alzheimer's disease, and the accuracy of the test depended on the level of ABeta42 used.

How reliable are the results of the studies in this review?

In most of the included studies, the diagnosis of Alzheimer's dementia was made by assessing all participants with standard diagnostic criteria.This is likely to have been a reliable method for deciding whether patients really had Alzheimer's disease. However, there were some problems with how the studies were conducted. This may result in ABeta42 appearing more accurate than it really is.

To whom do the results of this review apply?

The results apply to patients undergoing dementia assessment in a specialist setting.

What are the implications of this review?

Measuring levels of ABeta42 in spinal fluid may help distinguish Alzheimer’s disease from other types of dementia, but the test is not perfect. ABeta42 is unlikely to be used in isolation for making a diagnosis, and may have greatest value when used in addition to the other assessments and tests that are undertaken to make a diagnosis of dementia.

How up-to-date is the review?

The review authors searched for and included studies published up to February 2020.

SUMMARY OF FINDINGS

Summary of findings 1. Summary of findings table

Patient population	People with a clinical diagnosis of dementia.
Review question	How accurate is CSF ABeta42 test for distinguishing Alzheimer's disease dementia (ADD) from other types of dementia?
Index test	Cerebrospinal fluid (CSF) ABeta42 test.
Reference standard	Clinical diagnostic criteria for dementia pathological subtypes (Appendix 1).
Target condition	ADD vs other dementia subtypes
Included studies	39 studies (5000 participants).
Quality concerns	The majority of studies were classified at high or unclear risk of bias, particularly for patient selection (n = 28), and the index test (n = 25). The majority of studies were at low risk for applicability concerns (n = 33 to 36 for each domain). Studies were mainly at unclear risk of bias due to inadequate reporting. Few studies pre-specified the test threshold and used optimal cut-offs calculated using the study data.
Heterogeneity	The majority of studies in this review were conducted in specialist secondary care settings. The majority of studies conducted the index test in a similar manner. Sources of heterogeneity were: patient population and dementia subtype enrolled, test threshold used, and the diagnostic criteria and definition of ADD and dementia subtypes.

Differential Diagnosis	Number of participants	Number of studies	Number of participants with ADD	Pooled sensitivity (95% confidence interval)	Pooled specificity (95% confidence interval)	Pooled false positive rate (95% confidence interval)	Pooled positive likelihood ratio (95% confidence interval)	Pooled negative likelihood ratio (95% confidence interval)
ADD vs non-ADD	1704	13	880	0.79 (0.73, 0.85)	0.60 (0.52, 0.67)	0.40 (0.33, 0.48)	1.98 (1.58, 2.47)	0.34 (0.24, 0.49)
ADD vs VaD	1151	11	941	0.79 (0.75, 0.83)	0.69 (0.55, 0.81)	0.31 (0.20, 0.45)	2.58 (1.75, 3.81)	0.30 (0.25, 0.36)
ADD vs FTD	1948	17	1371	0.85 (0.79, 0.89)	0.72 (0.55, 0.84)	0.28 (0.16, 0.45)	3.00 (1.81, 5.00)	0.21 (0.16, 0.28)
ADD vs DLB	1929	9	1521	0.77 (0.70, 0.83)	0.66 (0.51, 0.78)	0.34 (0.22, 0.49)	2.27 (1.57, 3.28)	0.35 (0.28, 0.45)

ADD vs NPH	336	4	258	0.84 (0.79, 0.88)	0.42 (0.26, 0.60)	0.58 (0.40, 0.74)	1.45 (1.07, 1.97)	0.38 (0.23, 0.63)
ADD vs CJD	382	3	321	0.82 (0.77, 0.86)	0.46 (0.34, 0.58)	0.54 (0.42, 0.66)	1.51 (1.15, 1.87)	0.40 (0.26, 0.54)
ADD vs ARC-D[a]	53	1	33	0.80	0.85	-	-	-
Conclusions	Our results suggest that ABeta42 could be useful in improving differential diagnosis of the dementia syndrome, but the test is imperfect. It is unlikely that the ABeta42 biomarker would be used in isolation in clinical practice and ideally it should be used to support the diagnosis alongside full clinical, radiological, and neuropsychological assessment. Our review does not help answer questions around the added value of the test over routine diagnostics.							
Implications	The test accuracy demonstrated does lend some support to the concept of using biomarkers to differentiate dementia type for tailored therapy. Clinical trials of anti-amyloid interventions could consider using quantification of ABeta42 for patient selection. The biomarker does not guarantee an exclusively ADD population but it may help select those people most likely to benefit from the intervention.							

ADD: Alzheimer's disease dementia; ARCD: Alcohol-related cognitive disorder; CJD: Creutzfeldt-Jakob disease; DLB: Dementia with Lewy bodies; FTD: Frontotemporal dementia; LR: Likelihood ratio; NPH: Normal pressure hydrocephalus; Sens: sensitivity; Spec: specificity; VaD: Vascular dementia
[a]Note that there was only one study for the ADD vs ARCD comparison; therefore, data presented are from a single study

Summary of findings 2. Summary of subgroup analyses

Differential diagnosis	Number of participants	Number of studies	Number of participants with ADD	Pooled sensitivity (95% confidence interval)	Pooled specificity (95% confidence interval)	Pooled false positive rate (95% confidence interval)	Pooled positive likelihood ratio (95% confidence interval)	Pooled negative likelihood ratio (95% confidence interval)
Effect of test threshold								
ADD vs non-ADD (threshold ≤ 500 pg/ml)	1160	7	519	0.79 (0.68, 0.86)	0.58 (0.45, 0.70)	0.42 (0.30, 0.55)	1.87 (1.26, 2.77)	0.37 (0.20, 0.67)
ADD vs non-ADD (threshold > 500 pg/ml)	406	5	292	0.78 (0.70, 0.84)	0.62 (0.50, 0.73)	0.38 (0.27, 0.50)	2.04 (1.54, 2.71)	0.36 (0.27, 0.49)
ADD vs VaD (threshold ≤ 500 pg/ml)	809	7	697	0.79 (0.74, 0.82)	0.68 (0.51, 0.82)	0.32 (0.19, 0.49)	2.47 (1.54, 3.95)	0.31 (0.25, 0.39)
ADD vs VaD (threshold > 500 pg/ml)	194	3	133	0.86 (0.74, 0.93)	0.65 (0.37, 0.85)	0.35 (0.15, 0.63)	2.43 (1.25, 4.74)	0.22 (0.14, 0.36)

ADD vs FTD (threshold ≤ 500 pg/ml)	1033	8	753	0.87 (0.80, 0.92)	0.51 (0.21, 0.80)	0.49 (0.20, 0.79)	1.77 (0.92, 3.41)	0.25 (0.14, 0.44)
ADD vs FTD (threshold >500 pg/ml)	513	5	345	0.81 (0.73, 0.88)	0.84 (0.72, 0.91)	0.16 (0.09, 0.29)	5.02 (2.66, 9.48)	0.22 (0.14, 0.35)
ADD vs bvFTD (all thresholds)	898	8	651	0.85 (0.80, 0.89)	0.68 (0.51, 0.81)	0.32 (0.19, 0.49)	2.68 (1.65, 4.36)	0.22 (0.15, 0.32)
ADD vs PPA (all thresholds)	192	3	171	0.94 (0.50, 1.00)	0.23 (0.00, 0.98)	0.77 (0.03, 1.00)	1.22 (0.45, 3.34)	0.27 (0.03, 2.71)
ADD vs DLB (threshold ≤ 500 pg/ml)	751	6	563	0.79 (0.69, 0.86)	0.68 (0.46, 0.85)	0.32 (0.15, 0.54)	2.49 (1.37, 4.50)	0.31 (0.22, 0.43)
Effect of age								
ADD vs non-ADD (older participants)	1555	10	779	0.80 (0.76, 0.84)	0.62 (0.52, 0.70)	0.39 (0.30, 0.48)	2.08 (1.66, 2.61)	0.32 (0.26, 0.40)
ADD vs non-ADD (younger participants)	149	3	105	0.71 (0.47, 0.87)	0.51 (0.32, 0.69)	0.49 (0.31, 0.68)	1.44 (0.78, 2.65)	0.58 (0.22, 1.54)
ADD vs VaD (older participants)	1067	9	881	0.80 (0.75, 0.84)	0.68 (0.53, 0.80)	0.32 (0.20, 0.48)	2.49 (1.65, 3.74)	0.30 (0.25, 0.37)
ADD vs FTD (older participants)	1788	14	1220	0.85 (0.79, 0.90)	0.68 (0.47, 0.84)	0.32 (0.16, 0.53)	2.67 (1.52, 4.69)	0.22 (0.16, 0.30)
ADD vs FTD (younger participants)	160	3	95	0.82 (0.69, 0.91)	0.86 (0.76, 0.93)	0.14 (0.07, 0.25)	6.01 (3.24, 11.14)	0.20 (0.11, 0.38)
Effect of studies with high drop-out rates removed								
ADD vs VaD	896	9	712	0.79 (0.74, 0.84)	0.70 (0.53, 0.83)	0.30 (0.17, 0.47)	2.64 (1,65, 4.24)	0.30 (0.24, 0.36)
ADD vs FTD	1480	14	1023	0.81 (0.76, 0.85)	0.75 (0.62, 0.85)	0.25 (0.15, 0.39)	3.24 (2.05, 5.13)	0.25 (0.20, 0.32)
ADD vs DLB	1929	9	1521	0.745 (0.66, 0.83)	0.68 (0.48, 0.83)	0.33 (0.17, 0.53)	2.32 (1.43, 3.76)	0.37 (0.29, 0.46)
ADD vs NPH	137	3	93	0.86 (0.72, 0.94)	0.49 (0.32, 0.67)	0.51 (0.33, 0.68)	1.70 (1.13, 2.57)	0.28 (0.11, 0.73)
Effect of studies without pre-specified thresholds removed								

ADD vs non-ADD	566	5	366	0.79 (0.73, 0.84)	0.60 (0.49, 0.71)	0.40 (0.29, 0.51)	1.98 (1.50, 2.62)	0.35 (0.26, 0.46)
ADD vs VaD	265	3	175	0.80 (0.73, 0.85)	0.73 (0.61, 0.82)	0.28 (0.18, 0.39)	1.36 (1.01, 1.71)	0.97 (0.44, 1.50)
ADD vs FTD	870	7	615	0.84 (0.71, 0.92)	0.63 (0.21, 0.91)	0.37 (0.09, 0.79)	2.27 (0.79, 6.57)	0.25 (0.90, 2.47)
ADD vs DLB	214	3	129	0.70 (0.62, 0.76)	0.70 (0.54, 0.82)	0.30 (0.18, 0.46)	2.31 (1.43, 3.75)	0.44 (0.32, 0.59)

ADD: probable or possible Alzheimer's disease dementia; ARCD: alcohol-related cognitive disorder; bvFTD: behavioral variant frontotemporal dementia; DLB: dementia with Lewy bodies; FTD: frontotemporal dementia; non-ADD: two or more other subtype dementias; NPH: normal pressure hydrocephalus; PPA: primary progressive aphasia; VaD: vascular dementia

BACKGROUND

Dementia is a syndrome of chronic decline in cognitive abilities severe enough to impair function in everyday activities (Robinson 2015). The ageing population will lead to an increased prevalence of neurodegenerative diseases such as dementia, with substantial implications for economies and society. Dementia has an annual estimated cost of over USD 818 billion worldwide (Prince 2015).

Dementia is a clinical syndrome that may have multiple aetiologies (DeTure 2019). Alzheimer's disease dementia (ADD) is the most common dementia subtype, affecting 6% of individuals over the age of 65 and 20% over the age of 80 (Knapp 2007). In terms of prevalence, it is followed by vascular dementia (VaD), mixed ADD/VaD, dementia with Lewy bodies (DLB), alcohol-related dementia and frontotemporal dementia (FTD) (Lopes 2010). In practice and in research, it can be difficult to differentiate between dementia subtypes (Karantzoulis 2011; Ryan 2018). There is often considerable overlap in the presentation with many common clinical features across the dementia subtypes (Karantzoulis 2011). Clinical diagnosis of dementia subtype is imperfect and diagnosis of ADD and other related disorders based on clinical criteria alone does not always align with the diagnosis made on neuropathology at autopsy (Beach 2012). However, differentiating subtypes is important for clinical practice. A pathological diagnosis of dementia type can guide personalised treatments and inform discussions around prognosis (Karantzoulis 2011). Medications approved for symptomatic treatment of dementia, such as cholinesterase inhibitors, are only recommended in certain dementia types. It is also possible that new treatments under development may have differential efficacy across dementia types (Karantzoulis 2011).

In Alzheimer's disease, amyloid beta peptides (ABeta) are produced via sequential cleavage, involving the action of beta and gamma secretases (De Strooper 2010). The most prevalent ABeta species produced during amyloid precursor protein processing are ABeta40 and ABeta42 (Murphy 2010). Amyloid deposition in the brain is a hallmark of Alzheimer's disease. The amyloid hypothesis of Alzheimer's disease describes a pathological cascade process resulting in the aggregation of soluble ABeta42 into insoluble oligomers and then plaques (Takami 2009). Measuring ABeta has been proposed as a diagnostic biomarker, as these proteins may reflect the underlying pathology of Alzheimer's disease (Hansson 2019). ABeta42 in cerebrospinal fluid (CSF) is a biomarker that is entering research and practice, and is said to reflect amyloid plaque burden in the brain (Hansson 2019). There is increasing evidence to suggest that the neurobiology underlying ADD is associated with reductions in ABeta42 levels in CSF (Hansson 2019). Although CSF ABeta42 reductions have been clearly associated with ADD, it is not yet clear if these changes are specific to ADD, or are a marker of other neurodegenerative processes (Hansson 2019). While most amyloid beta research has used CSF, it has been recently demonstrated that plasma markers of ABeta42 may have utility (Nakamura 2018).

Use of ABeta42 is increasing in clinical research of agents that target specific components of the amyloid neuropathological cascade. However, the association between ABeta42 levels and clinical dementia is not fully understood. People can have substantial cortical amyloid without developing clinical symptoms (Jansen WJ 2015) and individuals display variation in their resilience to the presence of cortical amyloid. Amyloid beta itself may not be the pathological entity and amyloidosis triggers downstream pathological processes that drive neurodegeneration and neuronal dysfunction, e.g. tau aggregation (Blurton-Jones 2006). It has also been postulated that amyloidosis may need the co-occurrence of another insult, e.g. cerebrovascular disease, to mediate clinical symptomatology (DeTure 2019; Klohs 2019).

While previous Cochrane reviews have sought to understand the value of abnormal levels of cortical amyloid to predict decline from a prodromal to a dementia phase of Alzheimer's disease (Ritchie 2014; Ritchie 2017), this review focussed on the ability of ABeta42 measures to differentiate between ADD from other dementia types.

Target condition being diagnosed

In this review we considered ADD and other pathological subtypes of dementia. We considered non-ADD subtypes as a group, and then considered separate pathological diagnoses within that group.

1) ADD

Alzheimer's disease is thought to underlie ADD. Alzheimer's disease is a clinical syndrome that manifests as progressive memory decline, with impairment in at least one other domain of cognitive function, which impacts on the person's function and behaviour (Karantzoulis 2011; Ryan 2018). Alzheimer's pathology affects the limbic system (primarily the hippocampus) and other mesiotemporal structures (DeTure 2019). The pathology also extends to other regions of the neocortex, including the frontal and parietal lobes, generating executive dysfunction and problems with praxis respectively (DeTure 2019; Karantzoulis 2011). Over time, the patient will develop worsening functional impairment as a consequence of their cognitive symptoms (Wilkosz 2010). Criteria such as those of the National Institute of Neurological and Communicative Diseases and Stroke, and the Alzheimer's Disease and Related Disorders Association (the NINCDS-ADRDA Alzheimer's Criteria 1984) and the Diagnostic and Statistical Manual of Mental Disorders (DSM) are currently used for the differential diagnosis of other dementia subtypes from ADD (Dubois 2007) (Appendix 1).

2) VaD

VaD is caused by underlying cerebrovascular disease (Burns 2005). Vascular dementia tends to follow a stepwise deterioration that is unpredictable in both speed of progression and clinical features (Iadecola 2019). The diagnosis for probable vascular dementia is based on criteria such as those of National Institute of Neurological Disorders and Stroke and the Association Internationale pour la Recherché et l'Enseignement en Neurosciences (the NINDS-AIREN criteria) (Roman 1993). These criteria have 58% sensitivity and 80% specificity for differentiating VaD from other dementias (Appendix 1).

3) FTD

FTD is the second most common form of dementia in people below the age of 65 years. FTD is associated with progressive change in personality, behaviour and language (Young 2018). Frontotemporal dementias tend to affect planning, judgement, personality and language early (Karantzoulis 2011; Young 2018). Memory impairment is not a prominent feature but by late stage, multiple cognitive domains may be affected (Karantzoulis 2011; Young 2018). The mean sensitivity and specificity for the Lund and

Manchester criteria for differentiating FTD from other dementia subtypes were both 97% (Lopez 1999) (Appendix 1). Within the FTD classification, there are subgroups of disease with differing risk factors, pathology and presentation.

4) DLB

In DLB, the characteristic pathology responsible for neurodegeneration in vulnerable neuronal populations is the presence of alpha-synuclein and ubiquitin aggregates within intraneuronal inclusion bodies, known as Lewy bodies (Outeiro 2019). These consist of a dense granular core, surrounded by a halo of radiating filaments (Beyer 2009). DLB principally leads to impairment in attention, with prominent, early neuropsychiatric symptoms (Outeiro 2019). According to Braak's and McKeith's staging/categorisation systems, the pathology correlates with clinical symptoms such that brainstem pathology is responsible for the extrapyramidal effects, whereas dementia results from neocortical pathology (Parkkinen 2008). The sensitivity and specificity of McKeith's 1996 clinical diagnostic criteria for differentiating DLB from other dementias was 60% and 94% respectively, while McKeith's 2005 criteria give sensitivity and specificity of 91% and 67% respectively (Rizzo 2018). Thus, clinical diagnostic criteria have become more sensitive and less specific over time (Appendix 1).

5) Dementia caused by alcohol-related cognitive disorder (ARCD)

Dementia originating primarily from chronic alcohol abuse or secondarily by alcohol-related syndromes, such as Wernicke's encephalopathy, is a common form of dementia in older individuals (Thomas 2001). The similarities between ADD and ethanol-related neurodegeneration, in addition to the higher prevalence of ADD in older patients, and the reluctance to admit alcohol excess, makes differentiating the two problematic (Kril 1999). The clinical diagnosis of 'alcohol induced persisting dementia' (Kapaki 2005) is based on the criteria set out in the DSM, 4th edition (DSM-IV) (APA 2000) (Appendix 1).

6) Dementia caused by CJD

Sporadic CJD and Alzheimer's disease share some clinical features, although the former is characterised by rapidly progressive dementia (Otto 2000). The International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) (WHO 1993) clinical criteria, such as clinical symptoms and characteristic electroencephalography (EEG), are used for diagnosis of CJD, including the presence of 14-3-3 protein in CSF, with 84% sensitivity and 92% specificity (Van Everbroeck 1999) (Appendix 1).

7) Dementia caused by NPH

NPH is classically characterized by the triad of symptoms, namely gait disturbance, dementia and urinary incontinence, and is associated with brain ventricular enlargement (Hakim 1965). NPH is one of the few known treatable causes of dementia. Thus, the discrimination of patients with dementia caused by NPH from patients with ADD or VaD is important, as dementia in early stage NPH is considered surgically reversible (Kapaki 2007).

Index test(s)

Our index test is a quantitative measure of ABeta42, measured in either CSF or blood. The assays commonly used to measure ABeta42 levels are the Innogenetics INNOTEST beta-amyloid 1-42 kit and the Athena Diagnostics test.

Clinical pathway

Dementia symptoms can develop slowly and only become obvious when there is marked cognitive impairment. Early assessment of cognitive issues would usually be in primary care or a generalist setting, with referral to a specialist dementia service as needed. The differentiation of dementia subtype would usually be performed in specialist, secondary care services. If CSF samples were to be used, this would necessarily be the reserve of the specialist clinic (NICE 2018), due to the invasive nature of these samples. Thus, our question relates to later stages in the clinical pathway, when people are already diagnosed with suspected, but undifferentiated, dementia. The potential use of the ABeta42 biomarkers that we consider in this review would be to differentiate dementia subtype, allowing individualised treatment (Khoury 2019).

ABeta42 testing is not standard practice in clinical settings. Using measures of amyloid in people with suspected neurodegenerative disease has been the subject of a substantial amount of research (Fantoni 2018; Ossenkoppele 2015; Ritchie 2014; Ritchie 2017) and debate within the dementia community. To date, the low specificity of abnormal ABeta42 levels in CSF has limited the clinical uptake of this biomarker (O'Brien 2017). The situation is different in research, and use of amyloid beta biomarkers to identify participants for anti-amyloid therapies is now obligatory in certain disease-modifying ADD trials (Cummings 2019). Even in this context, ABeta42 in isolation is imperfect as a case-mix adjuster or method for ensuring a pure ADD population (Hansson 2019; Niemantsverdriet 2017; Ritchie 2014).

Alternative test(s)

There are other methods for quantifying amyloid burden in the brain, e.g. neuroimaging using positron emission tomography (PET). For the purposes of this review, we focused only on CSF or blood testing of ABeta42 (Rabinovici 2019).

Rationale

Research criteria for defining the pathological process of Alzheimer's disease incorporate and promote use of biomarkers that can quantify amyloid burden. In clinical trials, ABeta42 is used to select potential participants. The use of CSF biomarkers, while not routine, is increasing in clinical practice (Albert 2011; Dubois 2010; McKhann 2011). However, before we incorporate biomarkers into practice or research it is crucial that we understand their diagnostic accuracy.

In this review, we considered ABeta42 as a tool for differentiating dementia subtypes. If a test could classify people with dementia based on the underlying pathology, this could have utility in clinical practice. It would allow tailored treatment (for example cholinesterase inhibitors work well in ADD but less well in VaD) and could be used to inform discussions around prognosis. A tool to classify dementia subtype would also have utility in research. Treatments are being developed that are specific to certain pathological processes, and tools such as ABeta42 could

help ensure that the participants enrolled in trials are those with the pathology most likely to benefit from the intervention.

OBJECTIVES

- To determine the diagnostic accuracy of plasma and CSF ABeta42 for distinguishing ADD from other forms of dementia in people who meet the general diagnostic criteria for a dementia syndrome in a specialist care setting

Secondary objectives

- To determine the diagnostic accuracy of plasma and CSF ABeta42 for distinguishing Alzheimer's disease dementia from specific forms of dementia (VaD, FTD, DLB, ARCD, CJD, NPH) in people who meet the general diagnostic criteria for a dementia syndrome in a specialist care setting.
- To investigate the effect of ABeta42 thresholds used to define test positivity on the test accuracy reported

METHODS

Criteria for considering studies for this review

Types of studies

We considered cross-sectional studies and noted the timeframe between the clinical diagnostic criteria and the ABeta42 measurement. In line with our review question, we only considered studies in which people with ADD were differentiated from patients with other dementia subtypes and not from cognitively healthy controls. In some studies, the final diagnosis was only confirmed after one to two years of follow-up, where CSF samples taken at the initial assessment were retrospectively analyzed. We considered these delayed verification studies eligible for inclusion in the review. We limited our inclusion to English-language studies.

Participants

We included all participants with a clinical diagnosis of any form of dementia, made using the standard clinical diagnostic criteria (Appendix 1) for the respective dementia subtype. We did not include participants with mild cognitive impairment. The setting of interest was specialist dementia services, whether serving outpatients or inpatients.

Index tests

Our index test is a quantitative measure of ABeta42, measured in either CSF or blood. There is currently no consensus on the threshold value that should signify test positivity for plasma or CSF ABeta42 tests. For our analyses, we did not pre-specify the positivity threshold, but used the thresholds that informed the primary analyses in the respective individual studies. We classified participants assessed by ABeta42 biomarkers as either test-positive (below study-specific threshold) or test-negative (above study-specific threshold) at baseline. We accepted any assay used to quantify the ABeta42.

Target conditions

Target conditions in this review are as follows.

- ADD and non-ADD, considered in aggregate and then considered by specific diagnoses:
  - VaD
  - FTD
  - DLB
  - Dementia caused by ARCD
  - Dementia caused by CJD
  - Dementia caused by NPH

Reference standards

For the purpose of this review, we accepted any validated clinical criteria-based definition of dementia, including iterations of DSM and ICD (APA 1987; APA 1994; WHO 1993) (Appendix 1). For ADD, we also accepted the NINCDS-ADRDA criteria (McKhann 1984).

Diagnostic criteria used to establish the other dementia subtypes in those participants with non-ADD were as follows:

- for VaD: the NINDS-ARIEN criteria (Roman 1993), the Alzheimer's Disease Diagnostic and Treatment Centers (ADDTC) criteria (Chui 1992), DSM-III-R criteria, DSM-IV criteria or ICD criteria;
- for FTD: the Lund criteria (Lund Manchester Groups 1994), Neary 1998 criteria or Boxer 2005 criteria;
- for DLB: the reference standard is the McKeith criteria (McKeith 1996, McKeith 2002 or McKeith 2005);
- for ARCD: the diagnostic criteria should follow DSM-III-R or DSM-IV;
- for dementia in CJD: the ICD-10 clinical criteria and characteristic EEG should be used;
- for dementia caused by NPH: we accepted ICD or DSM criteria.

Search methods for identification of studies

We used a variety of information sources to ensure all relevant studies are included. The Information Specialist of the Cochrane Dementia and Cognitive Improvement Group devised the search strategies for electronic database searching.

Electronic searches

The most recent searches for this review were performed on 18 February 2020. We searched the following databases.

- MEDLINE (OvidSP); earliest records to 18 February 2020
- Embase (OvidSP); earliest records to 18 February 2020
- BIOSIS Previews (Thomson Reuters Web of Science); earliest records to 18 February 2020
- Web of Science Core Collection, including Conference Proceedings Citation Index (Thomson Reuters Web of Science); earliest records to 18 February 2020
- PsycINFO (OvidSP); earliest records to 18 February 2020
- LILACS (Latin American and Caribbean Health Science Information database); earliest records to 18 February 2020

See Appendix 2 for details of the sources searched, the search strategies used, and the number of records that were retrieved.

We did not apply any language or date restrictions to the electronic searches. We did not use methodological search filters (i.e. collections of terms aimed at reducing the number needed to screen by filtering out irrelevant records and retaining only those

that are relevant) that were designed to retrieve diagnostic test accuracy studies, because available filters have not yet proved sensitive enough for systematic review searches (Beynon 2013).

Searching other resources

We also conducted searches in the following databases for other related systematic diagnostic accuracy reviews.

- Meta-analyses van Diagnostisch Onderzoek (MEDION) (www.mediondatabase.nl)
- Database of Abstracts of Reviews of Effects (DARE) (www.york.ac.uk/inst/crd/crddatabases.htm#DARE),
- Health Technology Assessments Database (HTA Database) (www.york.ac.uk/inst/crd/crddatabases.htm#HTA)
- Aggressive Research Intelligence Facility (ARIF) database (www.arif.bham.ac.uk)

We searched for systematic reviews of diagnostic studies from the International Federation of Clinical Chemistry and Laboratory Medicine Committee for Evidence-based Laboratory Medicine database (C-EBLM). We checked reference lists of any relevant systematic reviews for additional studies.

Data collection and analysis

Selection of studies

One review author (ANS) screened all titles and abstracts generated by electronic database searches for relevance, and excluded duplicate records. Following de-duplication, second assessment of the search results was divided among seven review authors (MK, RW, AH, MD, AG, EP, AA, LB, and TQ). Pairs of review authors (from among LB, MK, NS, and TQ) independently assessed full manuscripts against the inclusion criteria. Where necessary, a third review author (CR) resolved disagreements. The search was updated on 18 February 2020. When the same dataset was presented in more than one paper, we included the primary paper, which was the paper with the largest number of patients or with the most informative data.

Data extraction and management

We extracted data on study characteristics into a pre-standardised data extraction form, including data for the assessment of study quality and data for investigation of heterogeneity, as described in Appendix 3. We also extracted data for creating 2 x 2 tables (cross-relating index test results to the reference standards). Data extraction was performed independently by four blinded review authors (MK, NS, LB, TQ). Disagreement in data extraction was resolved by discussion, involving a third review author (CR) as arbitrator when necessary. Where a study did not present all relevant data for creating a 2 x 2 table, we contacted the study authors directly to request further information.

Assessment of methodological quality

We assessed the methodological quality of each study using the second version of the 'Quality Assessment of Diagnostic Accuracy Studies' (QUADAS-2) tool (Whiting 2011). The tool is made up of four domains: patient selection; index test; reference standard; and patient flow. Four independent raters (MK, NS, LB, TQ), blinded to each other's scores, performed QUADAS-2 assessments. Disagreement was resolved by further review and discussion with potential to involve a third review author (CR) as arbitrator if necessary. We assessed each domain in terms of risk of bias, with the first three domains also considered in terms of applicability. The components of each of these domains, and a rubric that details how judgements concerning risk of bias are made, are detailed in Appendix 4 and Appendix 5. We produced a narrative summary, describing numbers of studies that were found to have high, low, or unclear risk of bias, as well as describing our concerns regarding applicability.

Statistical analysis and data synthesis

We extracted the data from each study into a 2 x 2 table, showing the binary test results cross-classified with the binary reference standard. We organised test data so that the reference standard was always ADD and thus accuracy data were around differentiating ADD from other dementias. We entered true positive (TP), false negative (FN), false positive (FP) and true negative (TN) data from the included studies into RevMan 5.4 (Cochrane 2020) to calculate sensitivity and specificity and their 95% confidence intervals. We performed summary analyses using bivariate random-effects models, based on pairs of sensitivity and specificity, to calculate pooled estimates of sensitivity, specificity, positive predictive values, positive likelihood ratios and negative likelihood ratios, all with their associated 95% confidence intervals.

We used version 1.2 of the MetaDTA diagnostic test accuracy meta-analytic software (Freeman 2019; Patel 2020) in our analyses.

We presented summary analyses as forest plots and in receiver operating characteristic (ROC) space by plotting estimates of sensitivity and specificity with the associated 95% confidence interval of the pooled estimate. We only performed meta-analyses where there were sufficient studies (three or more studies).

Investigations of heterogeneity

We described the following factors:

- Index test: i) thresholds used; ii) method used to measure ABeta42 levels;
- Target disorder: i) reference standard used, e.g. NINCDS-ADRDA criteria versus DSM criteria versus ICD-10 criteria for ADD; ii) criteria used for the definition of a dementia syndrome: e.g. individual, clinician, algorithm, or consensus group
- Target population: i) spectrum of patients: age, sex, education, sampling strategy, Mini-Mental State Examination (MMSE) score and Apolipoprotein E (APOE) status of study participants; ii) clinical setting: outpatients versus inpatients versus participants in residential care.

Sensitivity analyses

We performed sensitivity analyses to assess the effect of differing ABeta42 test thresholds. In comparisons of ADD versus non-ADD, ADD versus VaD, ADD versus FTD, and ADD versus DLB, we grouped studies by similar thresholds as follows: those using thresholds less than or equal to 500 pg/ml, and those using thresholds over 500 pg/ml. We performed sensitivity analyses only where there were sufficient studies (three or more studies) to do so.

In addition, we performed sensitivity analyses for studies with younger populations of ADD participants: those where the mean age was under 66 years or who specifically enrolled participants with early-onset ADD.

We performed subgroup analyses on FTD variants: behavioural variant (bvFTD), and primary progressive aphasia (PPA).

Finally, we performed sensitivity analyses for studies with high drop out rates (greater than 30% of participants), and those which not pre-specify the test threshold.

Assessment of reporting bias

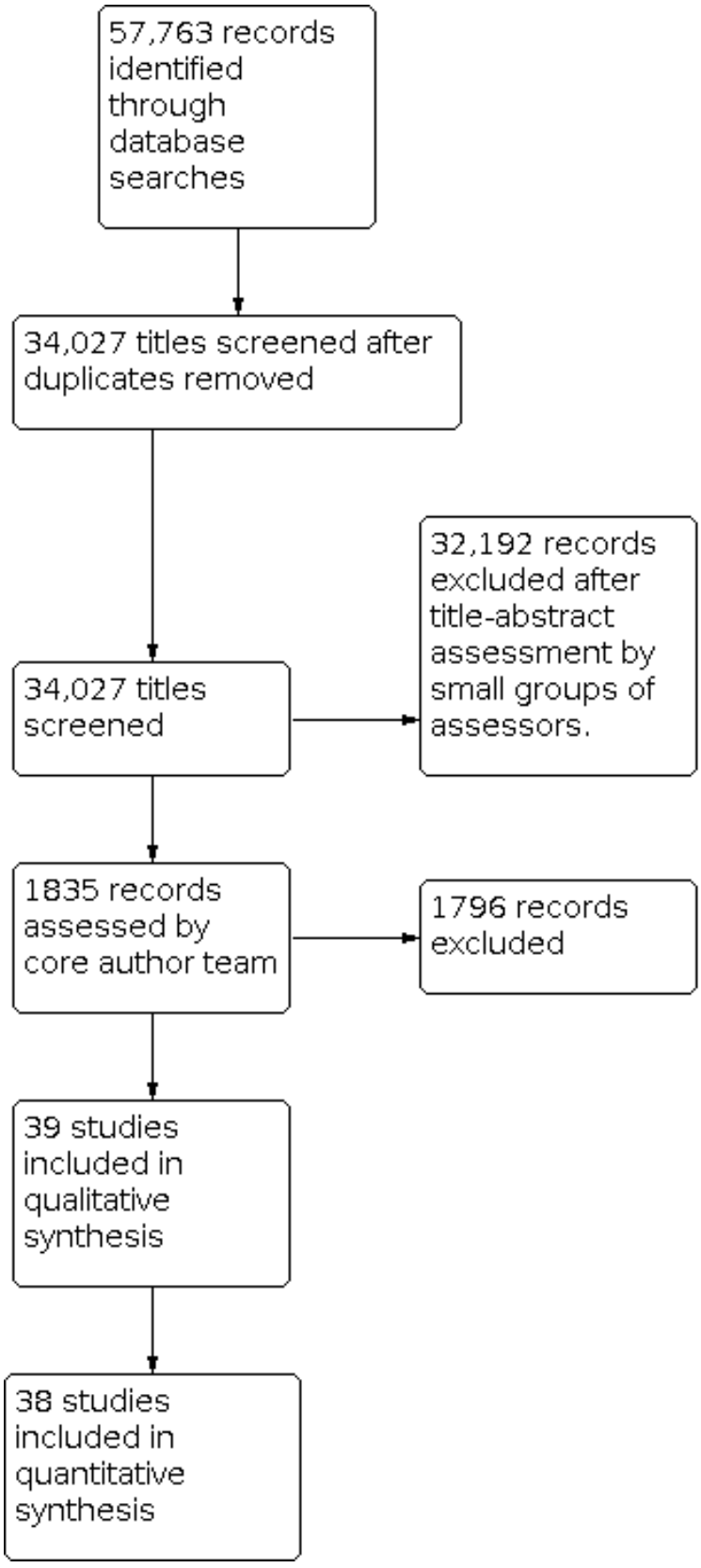
We did not investigate reporting bias because of current uncertainty about how it operates in test accuracy studies, and concerns about the interpretation of existing analytical tools, such as funnel plots.

RESULTS

Results of the search

We identified 57,763 titles after the electronic searches (Figure 1). After de-duplication and screening of titles for relevance, we screened 34,027 abstracts. We assessed 1835 full papers for eligibility and included 39 papers in the review.

Figure 1. Study flow diagram through the screening process.



We contacted seven authors for additional information about their studies but did not obtain usable data (Brandt 2008; Carandini 2019; Hampel 2018; Smach 2008a; Toledo 2012; van Steenoven 2018; van Steenoven 2019).

Summary of included studies

The Characteristics of included studies table lists the characteristics of the 39 included studies, comprising a total of 7246 participants. All studies were published between 2000 and 2020. Thirty-five studies were conducted in Europe. Three studies (Montine 2001; Shi 2018; Taricciotti 2018) were conducted in the USA and one study (Smach 2008) was conducted in Tunisia.

Index test

For the method used to measure ABeta42 levels (Table 1), 31 studies used the Innogenetics ELISA kit. Two studies used INNOTEST β-AMYLOID (1-42) ELISA kits from Fujirebio Inc. (Casoli 2019; Marchegiani 2019). One study used Athena Diagnostics (Montine 2001), one study used the ADmark ELISA kit (Taricciotti 2018) and one study used the ABeta-SDS-Page Immunoblot (Wiltfang 2003). Two studies did not report the ELISA kit they used (Lombardi 2018; Schirinzi 2015).

Three studies did not report thresholds used (Bibl 2007; Shi 2018; Spies 2010). Eleven studies pre-specified the thresholds used (Bousiges 2016; Bousiges 2018; Falgas 2020; Khoonsari 2019; Knapskog 2018; Lombardi 2018; Montine 2001; Perani 2016; Santangelo 2017; Sjogren 2000; Taricciotti 2018).

Target disorder

The majority of studies (n = 31) used the NINCDS-ADRDA criteria alone or in combination to define ADD. Two studies (Abu-Rumeileh 2018; Casoli 2019) used the International Working Group-2 criteria (Dubois 2014), six studies (Bibl 2006; Bibl 2007; Khoonsari 2019; Stefani 2005; Wiltfang 2003) used the DSM-IV, seven (Baldeiras 2015; Bousiges 2016; Casoli 2019; Falgas 2020; Lombardi 2018; Marchegiani 2019; Shi 2018) used the National Institute on Aging and Alzheimer's Association criteria (McKhann 2011), two (Bousiges 2016; Bousiges 2018) used Dubois (Dubois 2007), and one (Bousiges 2018) used Albert's (Albert 2011). Only one study did not report the criteria used to diagnose ADD (Knapskog 2018). The majority of studies (n = 26) did not report whether the diagnosis was made by a single clinician or consensus opinion. Of the studies that did report the diagnostic process, eight (Aerts 2011; Bibl 2006; Bibl 2007; de Rino 2012; Herbert 2014; Knapskog 2018; Perani 2016; Smach 2008) were by consensus amongst clinicians or multi-disciplinary team members, and a single clinician provided the diagnosis in five studies (Bousiges 2016; Bousiges 2018; de Jong 2006; Lombardi 2018; Taricciotti 2018).

Spectrum of participants

The sample sizes of the included studies ranged from 27 participants to 937 participants. Most (n = 32) studies enrolled late-onset ADD participants, or an older (mean age greater than 65 years) sample of participants with ADD. Three studies specifically enrolled participants with early-onset ADD (Falgas 2020; Rosler 2001; Sjogren 2000). Four studies enrolled participants with a mean age equal to or under 65 years (Bibl 2007; Kapaki 2005; Knapskog 2018; Montine 2001), but did not specifically investigate early-onset ADD.

Most studies enrolled more females than males, and the median proportion of males across studies was 42% (range 20% to 76%). In three studies, less than 30% of the sample was male (Herbert 2014; Lewczuk 2004; Wiltfang 2003). In two studies, more than 60% of the sample was male (Aerts 2011; Smach 2008). One study did not report the distribution of sex within the sample (Montine 2001).

Only seven studies (Abu-Rumeileh 2018; Baldeiras 2015; Lombardi 2018; Montine 2001; Santangelo 2017; Smach 2008; Taricciotti 2018) reported the education level of participants (range 6.2 years to 15.4 years).

Most studies (n = 24) did not clearly report the sampling strategy for included participants. Of those that did report sampling strategies, nine were retrospective analyses (Abu-Rumeileh 2018; Aerts 2011; Bousiges 2018; de Jong 2006; Herbert 2014; Lins 2004; Lombardi 2018; Smach 2008; Spies 2010; Taricciotti 2018), and five were consecutive samples (Bibl 2006; de Rino 2012; Marchegiani 2019; Sjogren 2000; Stefani 2005).

Eleven studies did not report the baseline MMSE scores for included participants (Brettschneider 2006; de Jong 2006; Kapaki 2001; Lins 2004; Santangelo 2017; Schirinzi 2015; Shi 2018; Sjogren 2000; Spies 2010; Taricciotti 2018; Wiltfang 2003). The median MMSE score across all studies was 18.4 (range 14 to 23.6), indicating the majority of participants had mild to moderate dementia severity. Only two studies reported the APOE4 status of participants, with 51% of ADD participants positive (Baldeiras 2015), and a mean level 14 amongst ADD participants (Rosler 2001).

Clinical setting

Memory clinics in specialist services or research centres recruited the majority of participants. Seventeen studies enrolled outpatients (Aerts 2011; Baldeiras 2015; Bibl 2006; Bousiges 2016; Bousiges 2018; Brettschneider 2006; de Jong 2006; de Rino 2012; Falgas 2020; Herbert 2014; Kapaki 2003; Knapskog 2018; Lombardi 2018; Maddalena 2003; Perani 2016; Santangelo 2017; Stefani 2005), three studies enrolled patients from mixed settings (inpatients and outpatients) (Bibl 2007; Kapaki 2005; Taricciotti 2018) and the remaining 19 studies did not report whether they included inpatients or outpatients. Three studies (Abu-Rumeileh 2018; Khoonsari 2019; Rosler 2001) did not report the sources of recruitment.

Methodological quality of included studies

We assessed methodological quality using the QUADAS-2 tool and at item level and provide aggregate scores in Figure 2, and Figure 3. We did not rate any studies as being at low risk of bias across all domains, with risk of bias predominantly resulting from patient selection and application of the index test.

Figure 2. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study

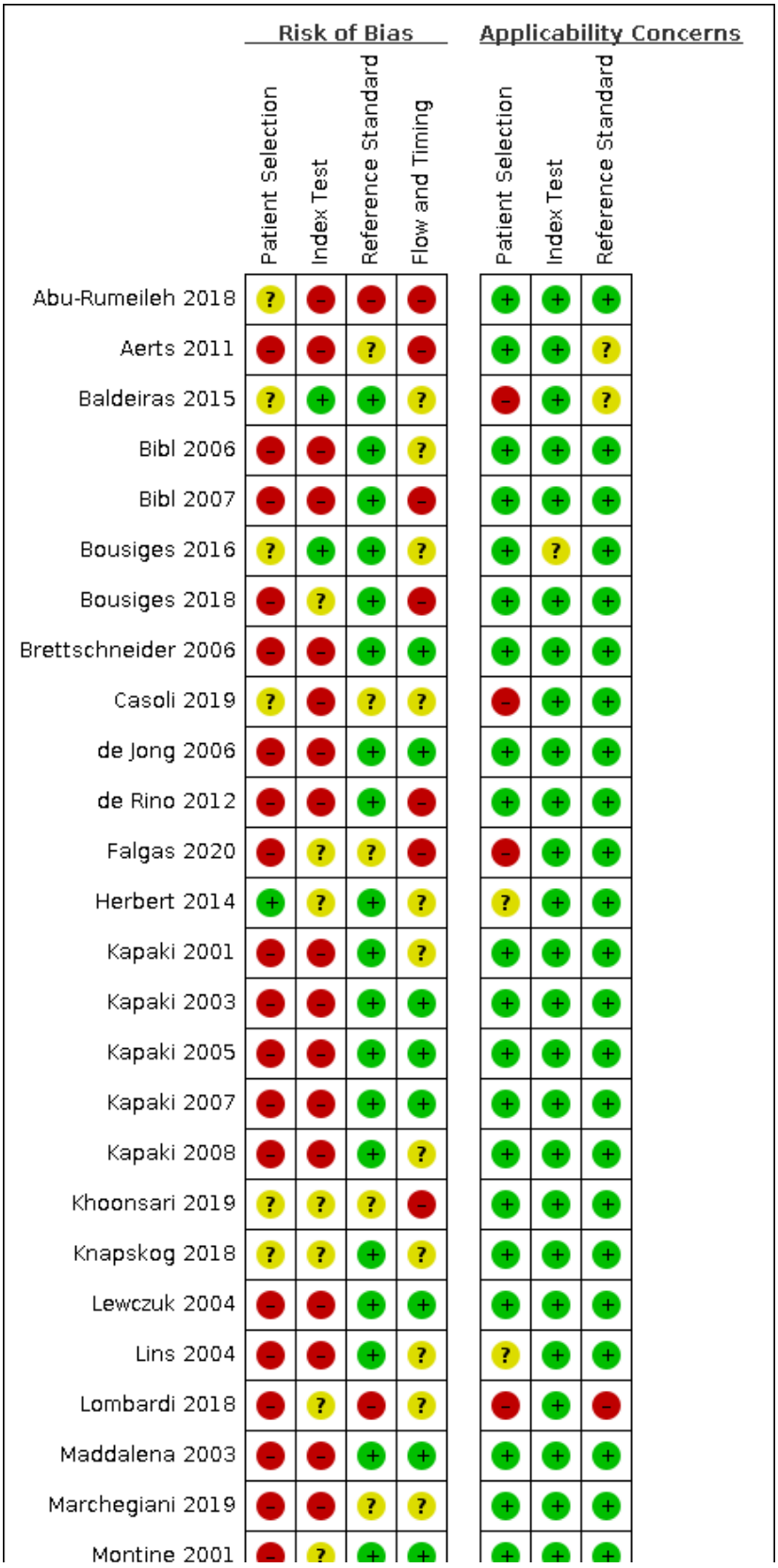


Figure 2. (Continued)

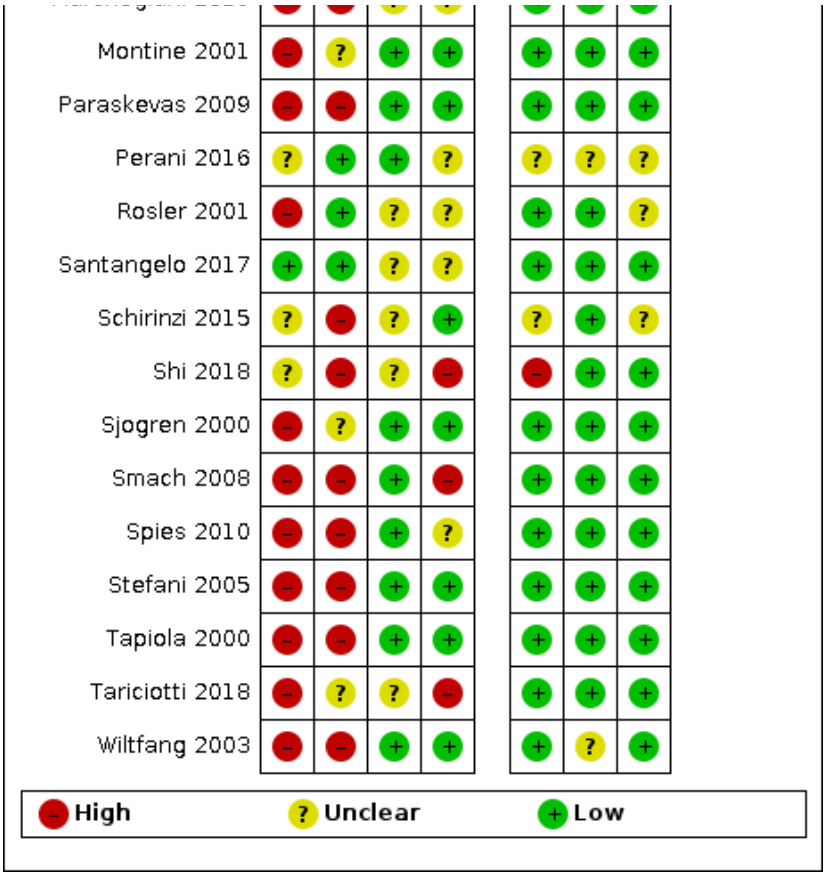
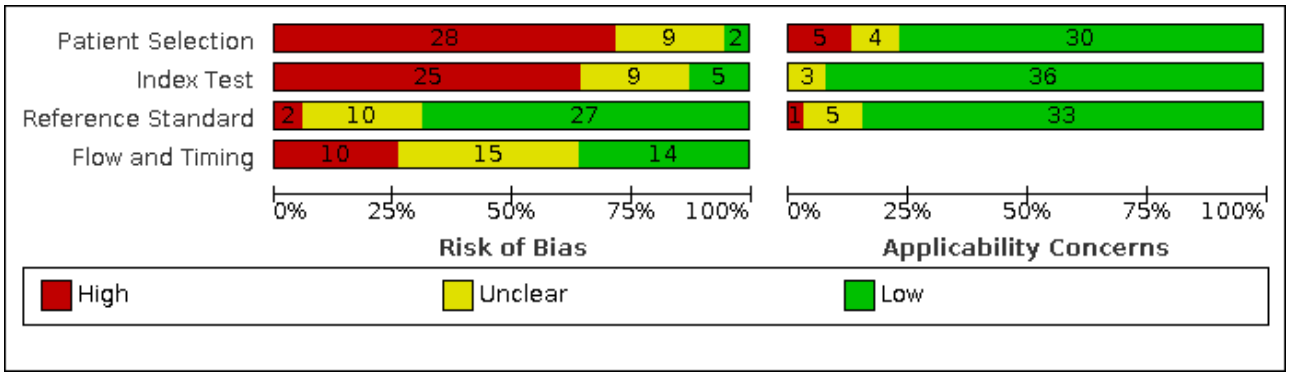


Figure 3. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies



We considered 28 studies to be at high risk of bias due to selective patient inclusion (for example, selective inclusion or enriching the population with a certain dementia type). We scored a further nine studies to be at unclear risk of bias in this domain, due to poor reporting.

In the index test domain, we considered 25 studies to be at high risk of bias because the ABeta42 threshold used was not pre-specified. Only eleven studies reported and used a pre-specified threshold. However, we judged nine of those studies to be at unclear risk of bias because they did not report whether investigators interpreted the ABeta42 data without knowledge of the dementia classification. Three studies did not report the threshold for the values of sensitivity and specificity they presented.

In the reference standard domain, we considered two studies to be at high risk of bias, because investigators made the dementia assessment with the knowledge of the ABeta42 result. We judged ten studies to be at unclear risk of bias because they did not report whether the investigator, who interpreted the results of reference

standard, conducted the assessment without the knowledge of the ABeta42 data.

In the flow and timing domain, we judged 10 studies to be at high risk of bias because the final clinical diagnosis was established (reassessed) 12 months or longer after CSF sampling. We considered fifteen studies to be at unclear risk of bias because not all patients were included in the analysis and/or studies did not report the interval between index test and reference standard.

For assessment of applicability concerns, we rated only five studies to be high risk. Many of the studies recruited from specialist, tertiary referral services and had access to assessments that may not be routine across all international dementia services. However, we did not consider this a major concern, as only specialist settings use the ABeta42 test at present.

Findings

We included a total of 39 studies (5000 participants) (Table 1). We present summary results of test accuracy for undifferentiated non-ADD and for specific dementia subgroups (Summary of findings 1).

CSF ABeta42 for differentiating ADD from non-ADD

The accuracy of ABeta42 to differentiate ADD from a mixed population of non-ADD subtypes was evaluated in a total of 13 studies (1704 participants, 880 with ADD). The pooled sensitivity at all thresholds was 79% (95% CI 73% to 85%), and the pooled specificity was 60% (95% CI 52% to 67%) (Figure 4 Figure 5).

Figure 4. Summary ROC Plot of CSF ABeta42 for differentiating ADD from non-ADD (all studies). Summary statistics: sensitivity: 79% (95% CI 73%-85%), specificity: 60% (95% CI 52%-67%).

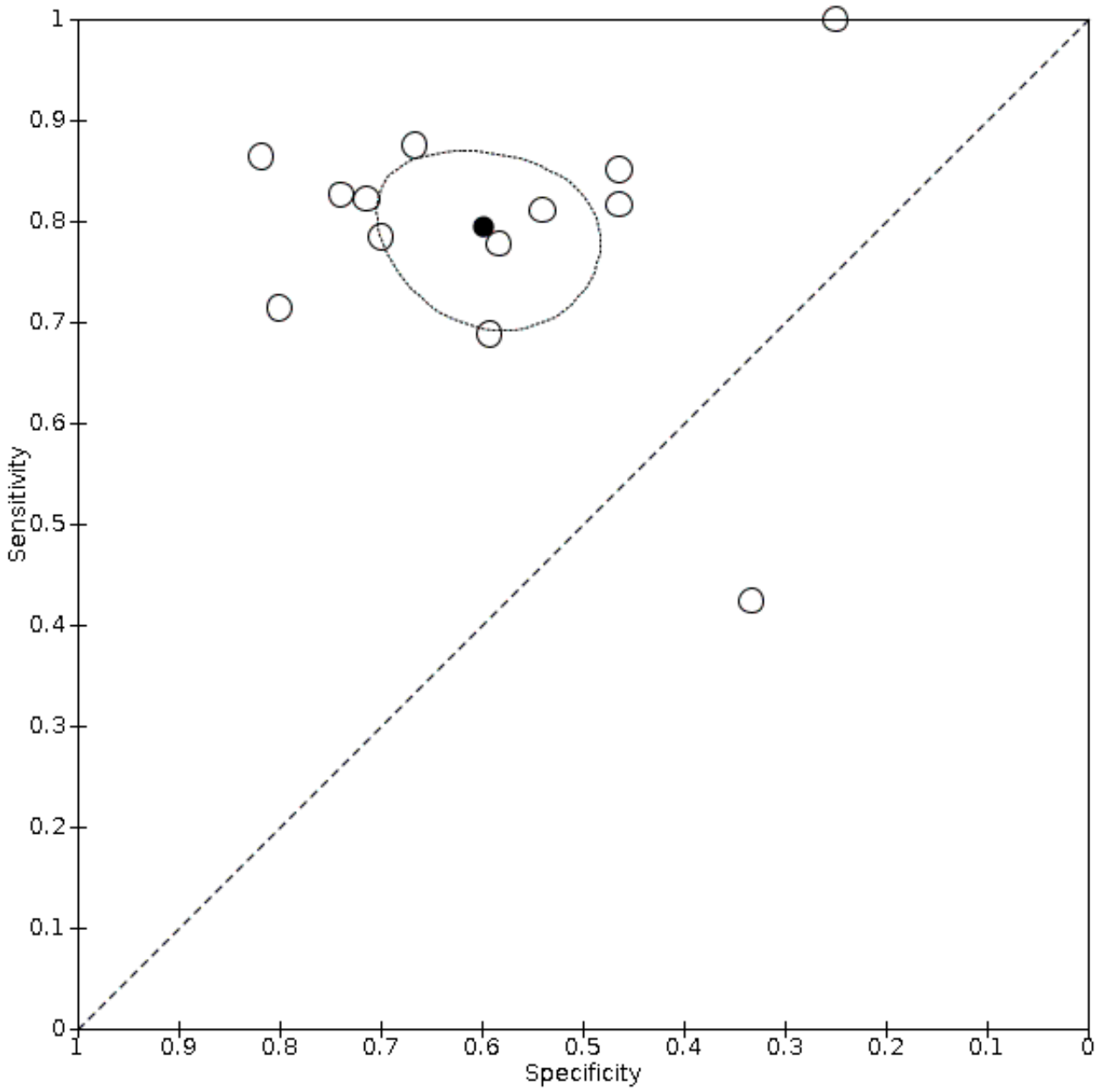
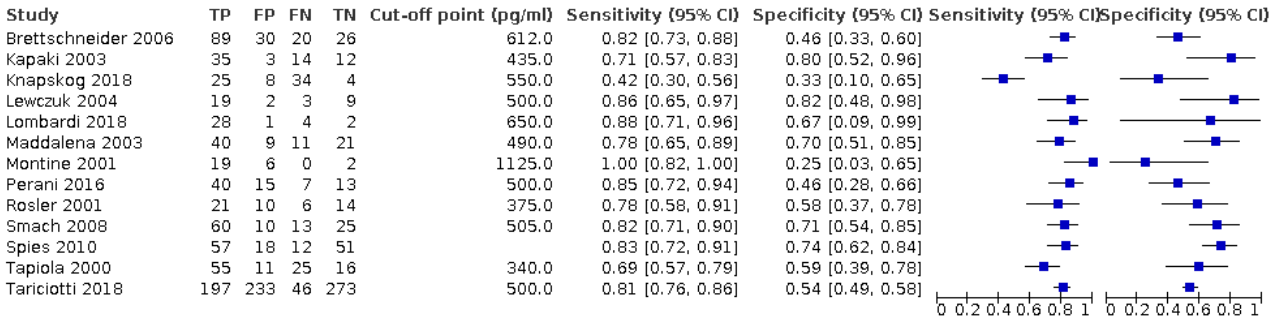


Figure 5. Forest plot of CSF ABeta42 for differentiating ADD from non-ADD (all studies)



In subgroup analysis, studies were separated into those using a threshold less than or equal to 500 pg/ml (seven studies, 1160 participants, 519 with ADD Figure 6; Figure 7), and those using a threshold above 500 pg/ml (five studies, 406 participants, 292 with ADD, Figure 8; Figure 9). The pooled sensitivity for studies using a threshold less than or equal to 500 pg/ml was 79% (95% CI 73% to 86%), and the pooled specificity was 58% (95% CI 45% to 70%). For studies using a threshold above 500 pg/ml, the pooled sensitivity was 78% (95% CI 70% to 84%), and the pooled specificity was 62% (95% CI 50% to 73%). We excluded one study (Spies 2010) that did not report a test threshold from the subgroup analyses. One study (Knapskog 2018) used two thresholds (550 pg/ml and 700 pg/ml); we included their 550pg/ml data in the subgroup analysis.

Figure 6. Summary ROC Plot of CSF ABeta42 for differentiating ADD from non-ADD (threshold ≤ 500 pg/ml). Summary statistics: sensitivity: 77% (95% CI 68%-86%), specificity: 58% (95% CI 45%-70%).

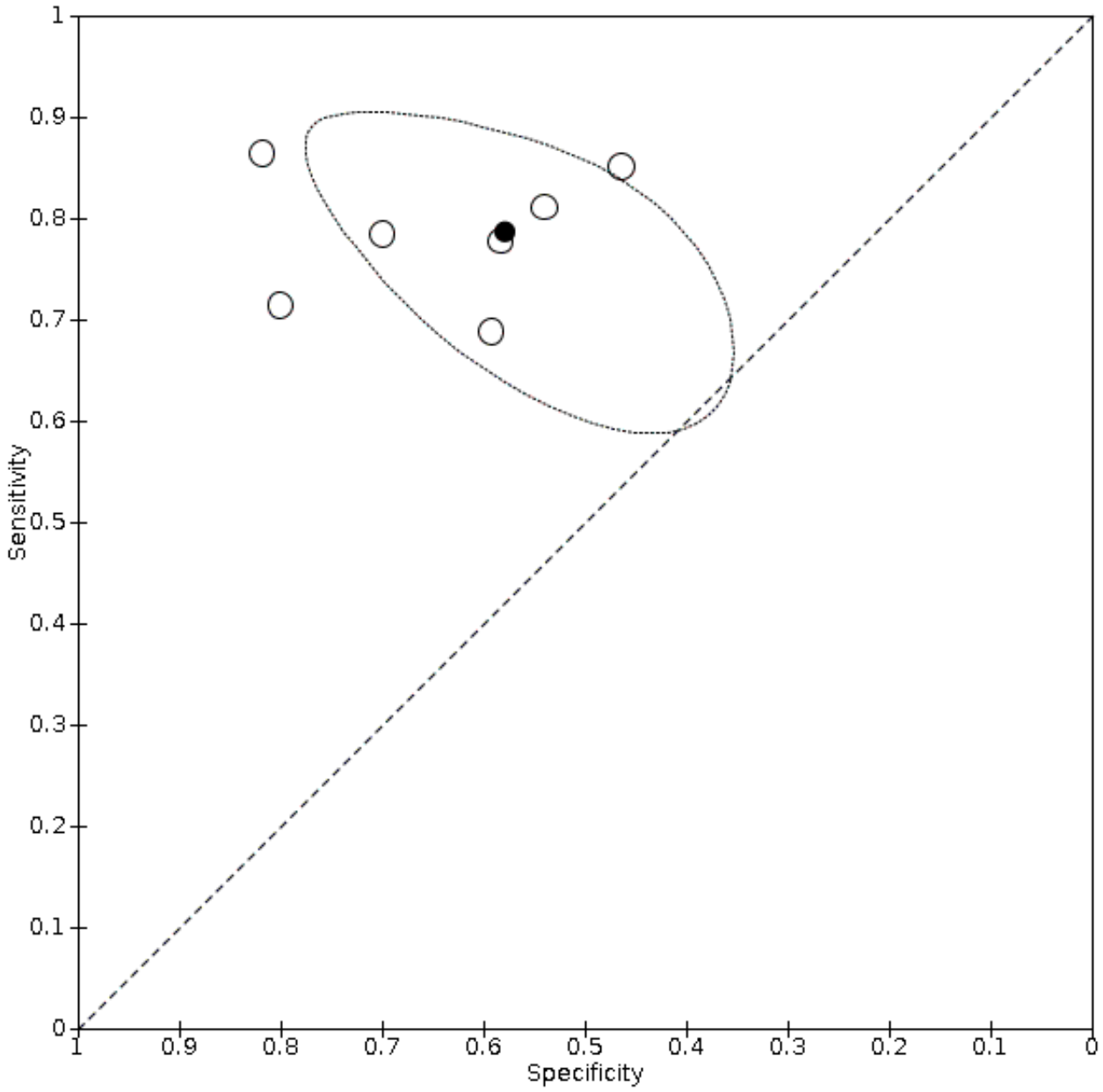


Figure 7. Forest plot of CSF ABeta42 for differentiating ADD from non-ADD (threshold ≤ 500 pg/ml).

Study	TP	FP	FN	TN	Cut-off point (pg/ml)	Sensitivity (95% CI)	Specificity (95% CI)
Kapaki 2003	35	3	14	12	435.0	0.71 [0.57, 0.83]	0.80 [0.52, 0.96]
Lewczuk 2004	19	2	3	9	500.0	0.86 [0.65, 0.97]	0.82 [0.48, 0.98]
Maddalena 2003	40	9	11	21	490.0	0.78 [0.65, 0.89]	0.70 [0.51, 0.85]
Perani 2016	40	15	7	13	500.0	0.85 [0.72, 0.94]	0.46 [0.28, 0.66]
Rosler 2001	21	10	6	14	375.0	0.78 [0.58, 0.91]	0.58 [0.37, 0.78]
Tapiola 2000	55	11	25	16	340.0	0.69 [0.57, 0.79]	0.59 [0.39, 0.78]
Tariciotti 2018	197	233	46	273	500.0	0.81 [0.76, 0.86]	0.54 [0.49, 0.58]

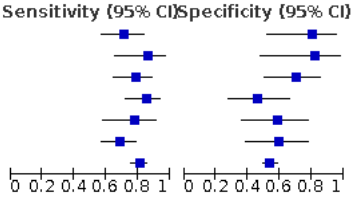


Figure 8. Summary ROC Plot of CSF ABeta42 for differentiating ADD from non-ADD (threshold > 500 pg/ml). Summary statistics: sensitivity: 78% (95% CI 70%-84%), specificity: 62% (95% CI 50%-73%).

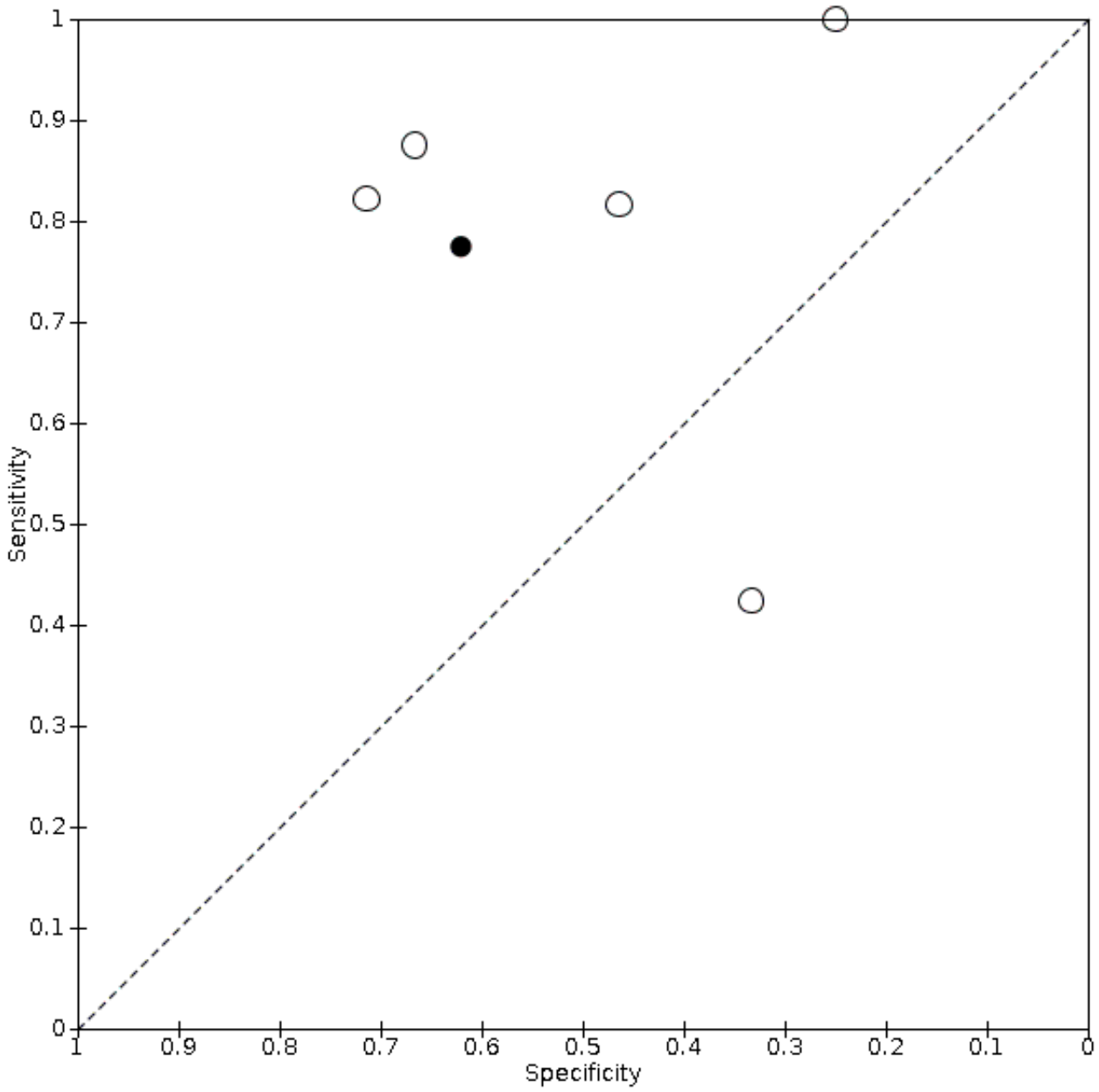
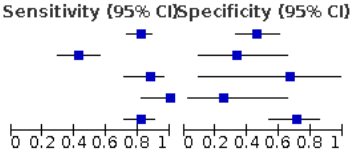


Figure 9. Forest plot of CSF ABeta42 for differentiating ADD from non-ADD (threshold > 500 pg/ml).

Study	TP	FP	FN	TN	Cut-off point (pg/ml)	Sensitivity (95% CI)	Specificity (95% CI)
Brettschneider 2006	89	30	20	26	612.0	0.82 [0.73, 0.88]	0.46 [0.33, 0.60]
Knapskog 2018	25	8	34	4	550.0	0.42 [0.30, 0.56]	0.33 [0.10, 0.65]
Lombardi 2018	28	1	4	2	650.0	0.88 [0.71, 0.96]	0.67 [0.09, 0.99]
Montine 2001	19	6	0	2	1125.0	1.00 [0.82, 1.00]	0.25 [0.03, 0.65]
Smach 2008	60	10	13	25	505.0	0.82 [0.71, 0.90]	0.71 [0.54, 0.85]



CSF ABeta42 for differentiating ADD from VaD

The accuracy of ABeta42 to differentiate ADD from VaD subtypes was evaluated in a total of 11 studies (1151 participants, 830 with ADD). The pooled sensitivity at all reported thresholds was 79% (95% CI 75% to 83%), and the pooled specificity was 69% (95% CI 55% to 81%) (Figure 10 Figure 11).

Figure 10. Summary ROC Plot of CSF ABeta42 for differentiating ADD from VaD (all studies). Summary statistics: sensitivity: 79% (95% CI 75%-83%), specificity: 69% (95% CI 55%-81%).

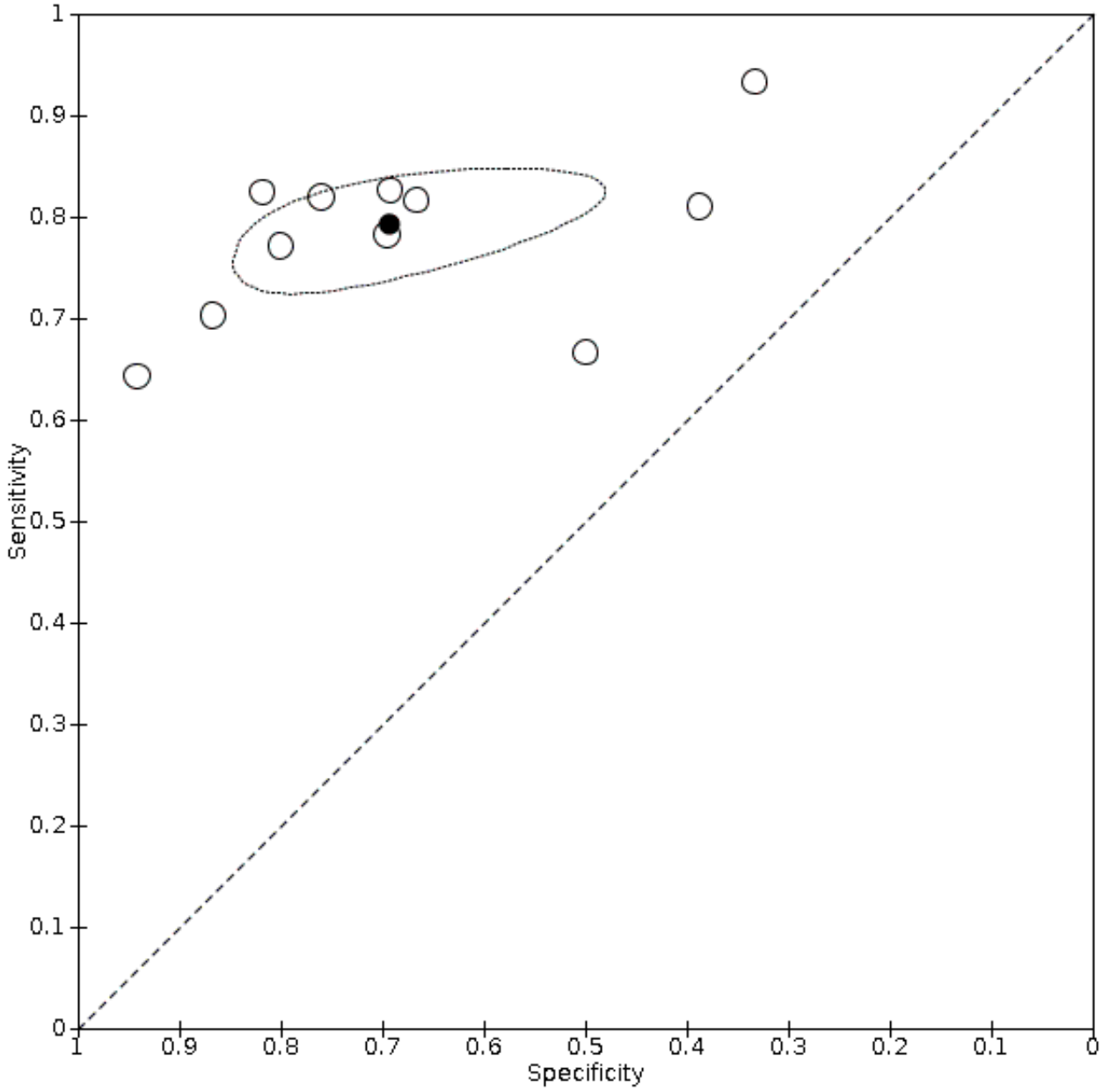
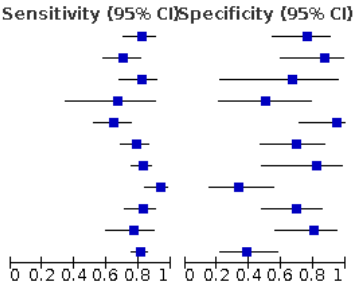


Figure 11. Forest plot of CSF ABeta42 for differentiating ADD from VaD (all studies).

Study	TP	FP	FN	TN	Cut-off point (pg/ml)	Sensitivity (95% CI)	Specificity (95% CI)
de Jong 2006	50	6	11	19	520.0	0.82 [0.70, 0.91]	0.76 [0.55, 0.91]
Herbert 2014	45	2	19	13	500.0	0.70 [0.58, 0.81]	0.87 [0.60, 0.98]
Kapaki 2003	40	2	9	4	435.0	0.82 [0.68, 0.91]	0.67 [0.22, 0.96]
Lins 2004	8	6	4	6	562.0	0.67 [0.35, 0.90]	0.50 [0.21, 0.79]
Marchegiani 2019	45	1	25	16	431.0	0.64 [0.52, 0.75]	0.94 [0.71, 1.00]
Paraskevas 2009	72	7	20	16	461.0	0.78 [0.68, 0.86]	0.70 [0.47, 0.87]
Santangelo 2017	136	2	29	9	500.0	0.82 [0.76, 0.88]	0.82 [0.48, 0.98]
Sjogren 2000	56	16	4	8	537.0	0.93 [0.84, 0.98]	0.33 [0.16, 0.55]
Spies 2010	57	8	12	18		0.83 [0.72, 0.91]	0.69 [0.48, 0.86]
Stefani 2005	27	4	8	16	493.0	0.77 [0.60, 0.90]	0.80 [0.56, 0.94]
Tariciotti 2018	214	19	50	12	500.0	0.81 [0.76, 0.86]	0.39 [0.22, 0.58]



In subgroup analysis, studies were separated into those using a threshold less than or equal to 500 pg/ml (seven studies, 809 participants, 697 with ADD) (Figure 12; Figure 13), and those using a threshold above 500 pg/ml (three studies, 194 participants, 133 with ADD) (Figure 14; Figure 15). The pooled sensitivity for studies using a threshold less than or equal to 500 pg/ml was 79% (95% CI 74% to 82%), and the pooled specificity was 68% (95% CI 51% to 82%). For studies using a threshold above 500 pg/ml, the pooled sensitivity was 86% (95% CI 74% to 93%), and the pooled specificity was 65% (95% CI 37% to 85%). We excluded one study (Spies 2010) that did not report a test threshold from the subgroup analyses.

Figure 12. Summary ROC Plot of CSF ABeta42 for differentiating ADD from VaD (threshold ≤ 500 pg/ml). Summary statistics: sensitivity: 79% (95% CI 74%-82%), specificity: 68% (95% CI 51%-82%).

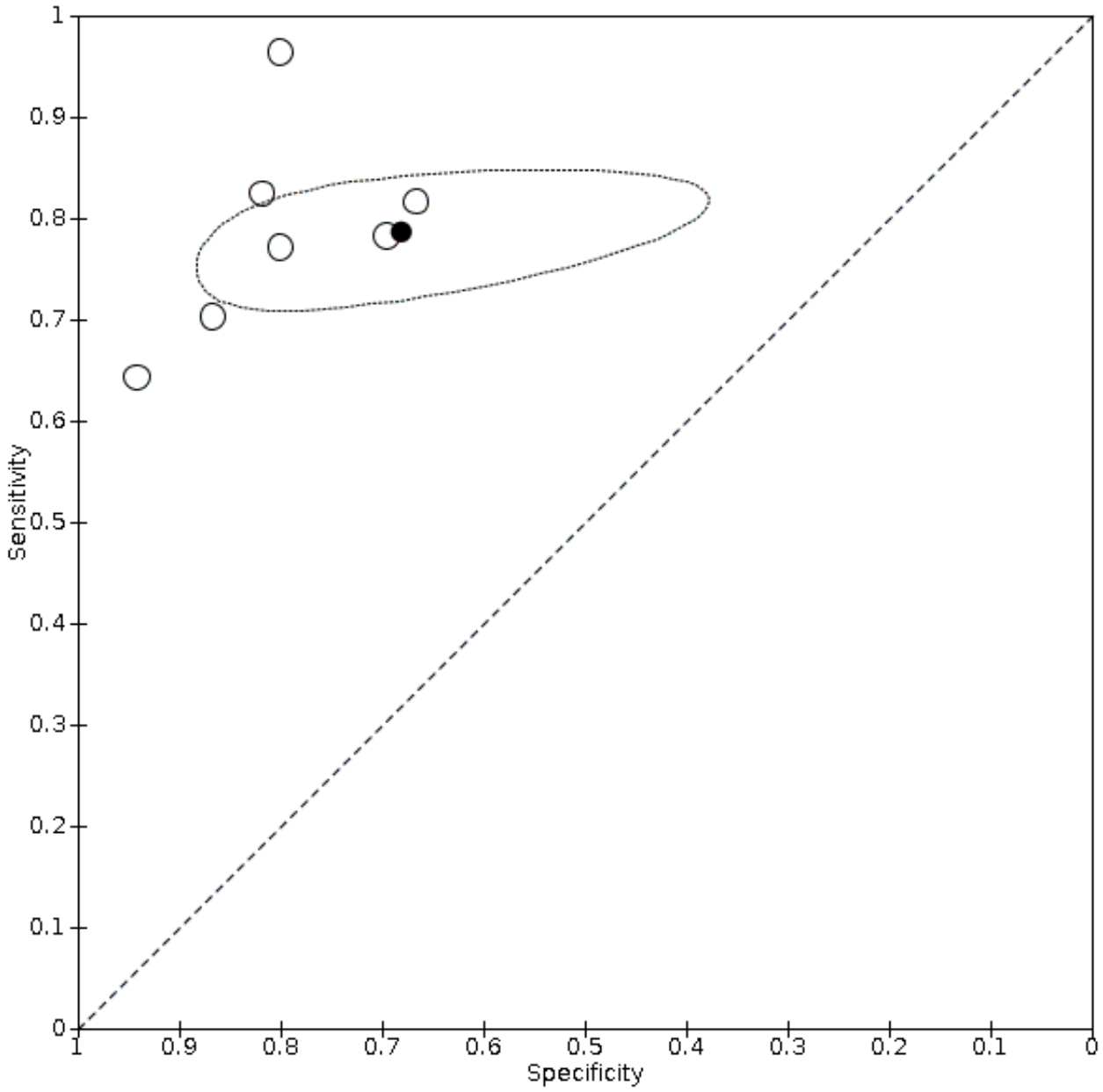


Figure 13. Forest plot of CSF ABeta42 for differentiating ADD from VaD (threshold ≤ 500 pg/ml).

Study	TP	FP	FN	TN	Cut-off point (pg/ml)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Herbert 2014	45	2	19	13	500.0	0.70 [0.58, 0.81]	0.87 [0.60, 0.98]		
Kapaki 2003	40	2	9	4	435.0	0.82 [0.68, 0.91]	0.67 [0.22, 0.96]		
Marchegiani 2019	45	1	25	16	431.0	0.64 [0.52, 0.75]	0.94 [0.71, 1.00]		
Paraskevas 2009	72	7	20	16	461.0	0.78 [0.68, 0.86]	0.70 [0.47, 0.87]		
Santangelo 2017	136	2	29	9	500.0	0.82 [0.76, 0.88]	0.82 [0.48, 0.98]		
Stefani 2005	27	4	8	16	493.0	0.77 [0.60, 0.90]	0.80 [0.56, 0.94]		
Tariciotti 2018	214	4	8	16	500.0	0.96 [0.93, 0.98]	0.80 [0.56, 0.94]		

Figure 14. Summary ROC Plot of CSF ABeta42 for differentiating ADD from VaD (threshold > 500 pg/ml). Summary statistics: sensitivity: 86% (95% CI 74%-93%), specificity: 65% (95% CI 37%-85%).

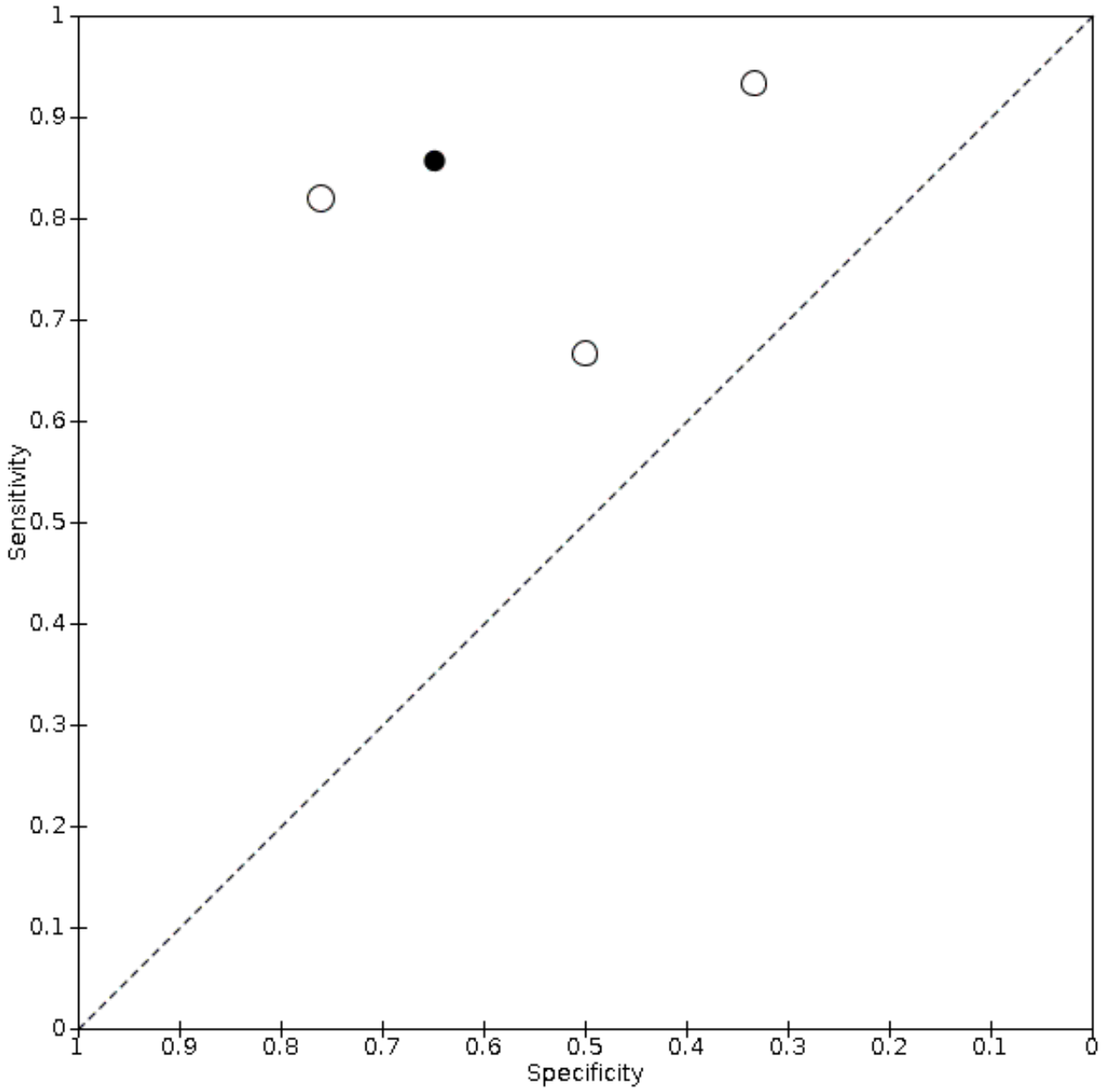


Figure 15. Forest plot of CSF ABeta42 for differentiating ADD from VaD (threshold > 500 pg/ml).

Study	TP	FP	FN	TN	Cut-off point (pg/ml)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
de Jong 2006	50	6	11	19	520.0	0.82 [0.70, 0.91]	0.76 [0.55, 0.91]		
Lins 2004	8	6	4	6	562.0	0.67 [0.35, 0.90]	0.50 [0.21, 0.79]		
Sjogren 2000	56	16	4	8	537.0	0.93 [0.84, 0.98]	0.33 [0.16, 0.55]		

CSF ABeta42 for differentiating ADD from FTD

The accuracy of ABeta42 to differentiate ADD from FTD subtypes was evaluated in a total of 17 studies (1948 participants, 1371 with ADD). The pooled sensitivity at all thresholds was 85% (95% CI 79% to 89%), and the pooled specificity was 72% (95% CI 55% to 84%) (Figure 16 Figure 17).

Figure 16. Summary ROC Plot of CSF ABeta42 for differentiating ADD from FTD (all studies). Summary statistics: sensitivity: 87% (95% CI 80%-92%), specificity: 51% (95% CI 21%-80%).

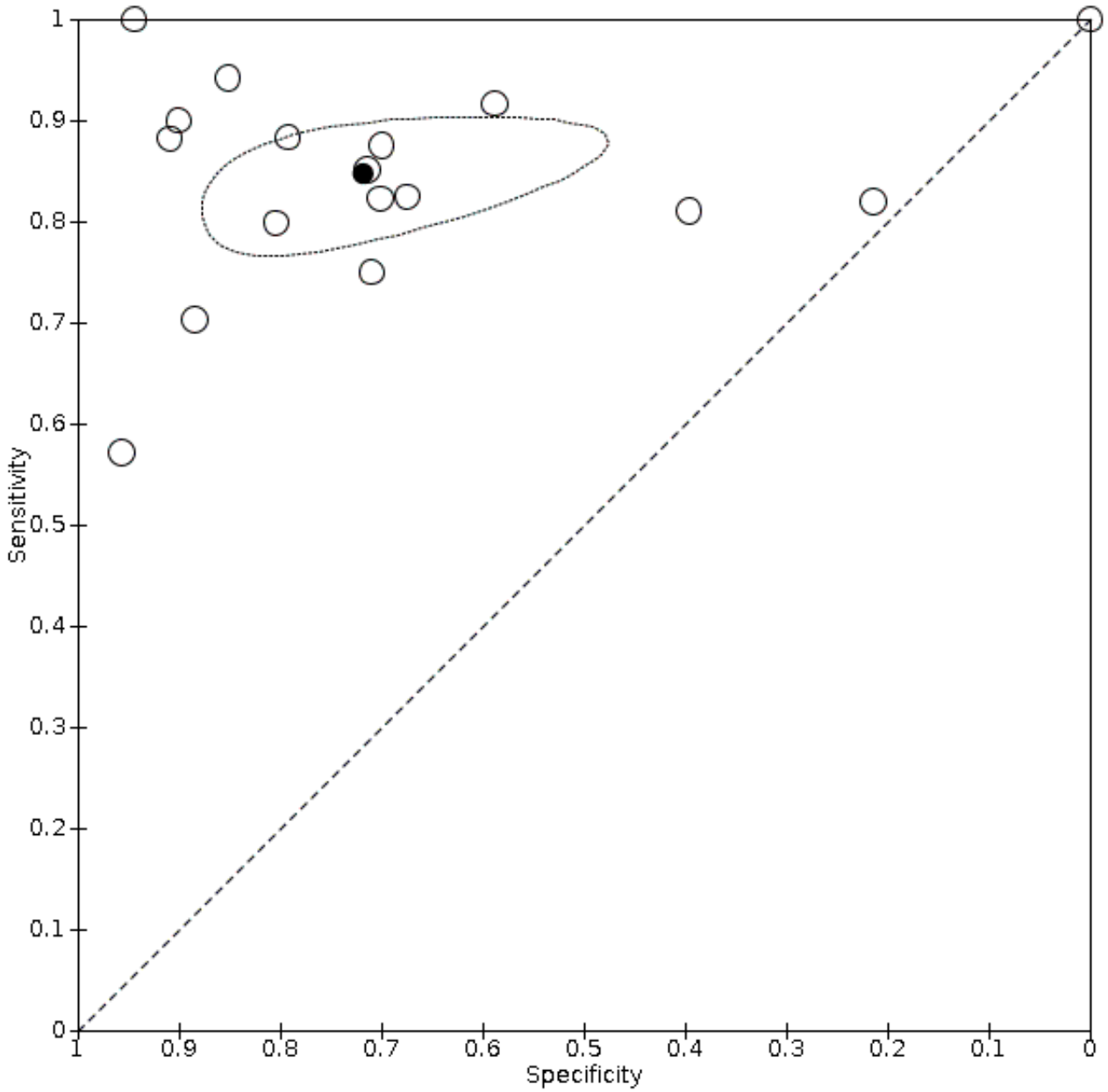
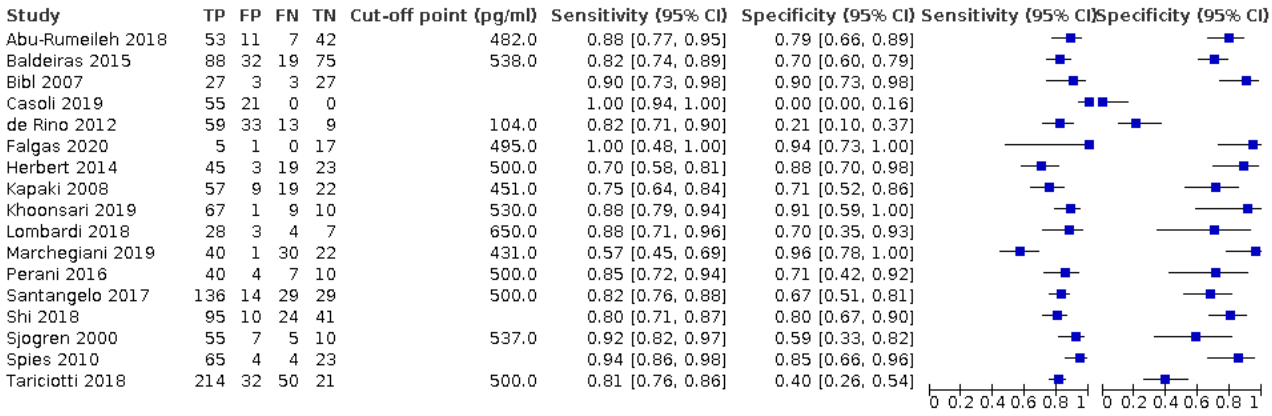


Figure 17. Forest plot of CSF ABeta42 for differentiating ADD from FTD (all studies).



In subgroup analysis, studies were separated into those using a threshold less than or equal to 500 pg/ml (eight studies, 1033 participants, 753 with ADD Figure 18; Figure 19), and those using a threshold above 500 pg/ml (five studies, 513 participants, 345 with ADD) (Figure 20; Figure 21). The pooled sensitivity for studies using a threshold less than or equal to 500 pg/ml was 87% (95% CI 80% to 92%), and the pooled specificity was 51% (95% CI 21% to 80%). For studies using a threshold above 500 pg/ml, the pooled sensitivity was 81% (95% CI 73% to 88%), and the pooled specificity was 84% (95% CI 72% to 91%). We excluded four studies (Bibl 2007; Casoli 2019; Shi 2018; Spies 2010) that did not report a test threshold from the subgroup analyses.

Figure 18. Summary ROC Plot of CSF ABeta42 for differentiating ADD from FTD (threshold ≤ 500 pg/ml). Summary statistics: sensitivity: 80% (95% CI 77%-84%), specificity: 69% (95% CI 49%-84%).

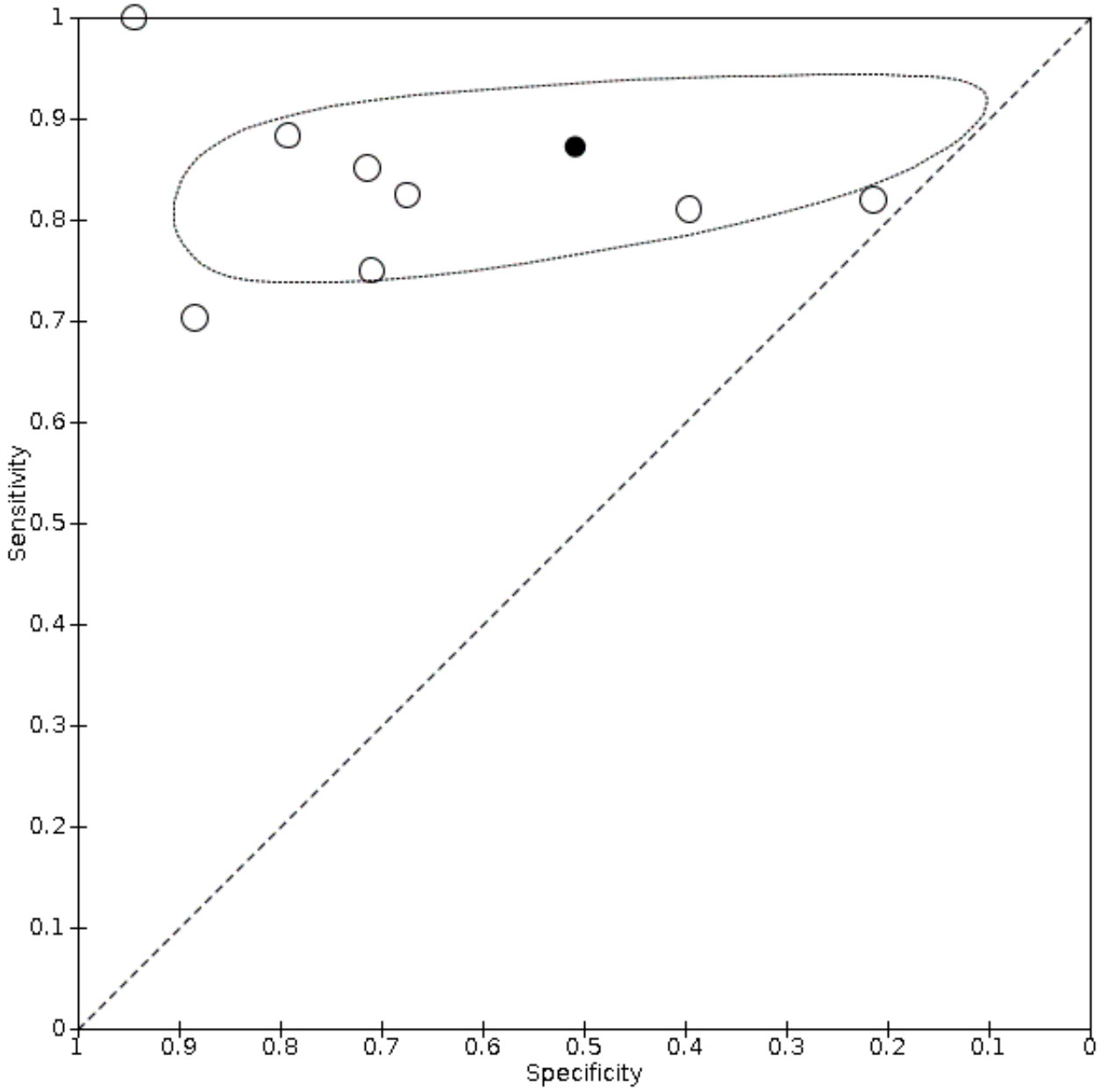


Figure 19. Forest plot of CSF ABeta42 for differentiating ADD from FTD (threshold ≤ 500 pg/ml).

Study	TP	FP	FN	TN	Cut-off point (pg/ml)	Sensitivity (95% CI)	Specificity (95% CI)
Abu-Rumeileh 2018	53	11	7	42	482.0	0.88 [0.77, 0.95]	0.79 [0.66, 0.89]
de Rino 2012	59	33	13	9	104.0	0.82 [0.71, 0.90]	0.21 [0.10, 0.37]
Falgas 2020	5	1	0	17	495.0	1.00 [0.48, 1.00]	0.94 [0.73, 1.00]
Herbert 2014	45	3	19	23	500.0	0.70 [0.58, 0.81]	0.88 [0.70, 0.98]
Kapaki 2008	57	9	19	22	451.0	0.75 [0.64, 0.84]	0.71 [0.52, 0.86]
Perani 2016	40	4	7	10	500.0	0.85 [0.72, 0.94]	0.71 [0.42, 0.92]
Santangelo 2017	136	14	29	29	500.0	0.82 [0.76, 0.88]	0.67 [0.51, 0.81]
Tariciotti 2018	214	32	50	21	500.0	0.81 [0.76, 0.86]	0.40 [0.26, 0.54]

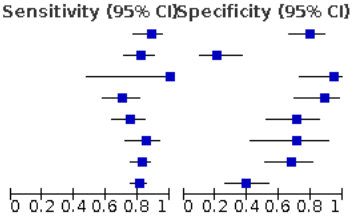


Figure 20. Summary ROC Plot of CSF ABeta42 for differentiating ADD from FTD (threshold > 500 pg/ml). Summary statistics: sensitivity: 83% (95% CI 71%-91%), specificity: 76% (95% CI 58%-87%).

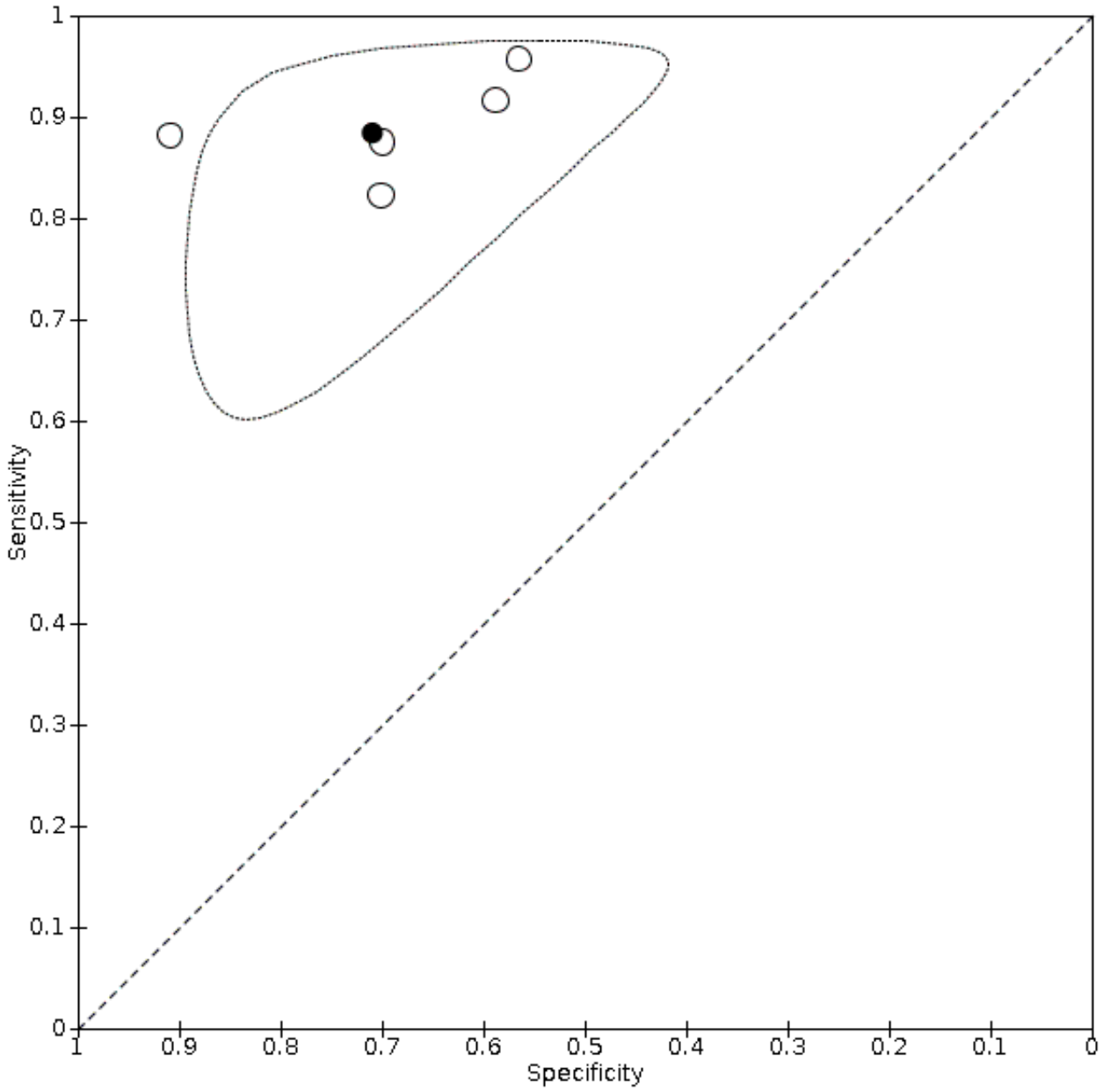


Figure 21. Forest plot of CSF ABeta42 for differentiating ADD from FTD (threshold > 500 pg/ml).

Study	TP	FP	FN	TN	Cut-off point (pg/ml)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Baldeiras 2015	88	32	19	75	538.0	0.82 [0.74, 0.89]	0.70 [0.60, 0.79]		
Khoonsari 2019	67	1	9	10	530.0	0.88 [0.79, 0.94]	0.91 [0.59, 1.00]		
Lombardi 2018	28	3	4	7	650.0	0.88 [0.71, 0.96]	0.70 [0.35, 0.93]		
Marchegiani 2019	67	10	3	13	431.0	0.96 [0.88, 0.99]	0.57 [0.34, 0.77]		
Sjogren 2000	55	7	5	10	537.0	0.92 [0.82, 0.97]	0.59 [0.33, 0.82]		

Test accuracy was investigated in two clinical subgroups of FTD (bvFTD and PPA). In the bvFTD subgroup (eight studies, 898 participants, 651 with ADD), the pooled sensitivity at all thresholds was 85% (95% CI 80% to 89%), and the pooled specificity was 68% (95% CI 51% to 81%). In the PPA subgroup (three studies, 192 participants, 171 with ADD) the pooled sensitivity at all thresholds

was 94% (95% CI 50% to 100%), and the pooled specificity was 23% (95% CI 0% to 98%).

CSF ABeta42 for differentiating ADD from DLB

The accuracy of ABeta42 to differentiate ADD from DLB subtypes was evaluated in a total of nine studies (1929 participants, 1521 with ADD). The pooled sensitivity at all thresholds was 77% (95% CI 70% to 83%), and the pooled specificity was 66% (95% CI 51% to 78%) (Figure 22, Figure 23).

Figure 22. Summary ROC Plot of CSF ABeta42 for differentiating ADD from DLB (all studies). Summary statistics: sensitivity: 77% (95% CI 70%-83%), specificity: 66% (95% CI 51%-78%).

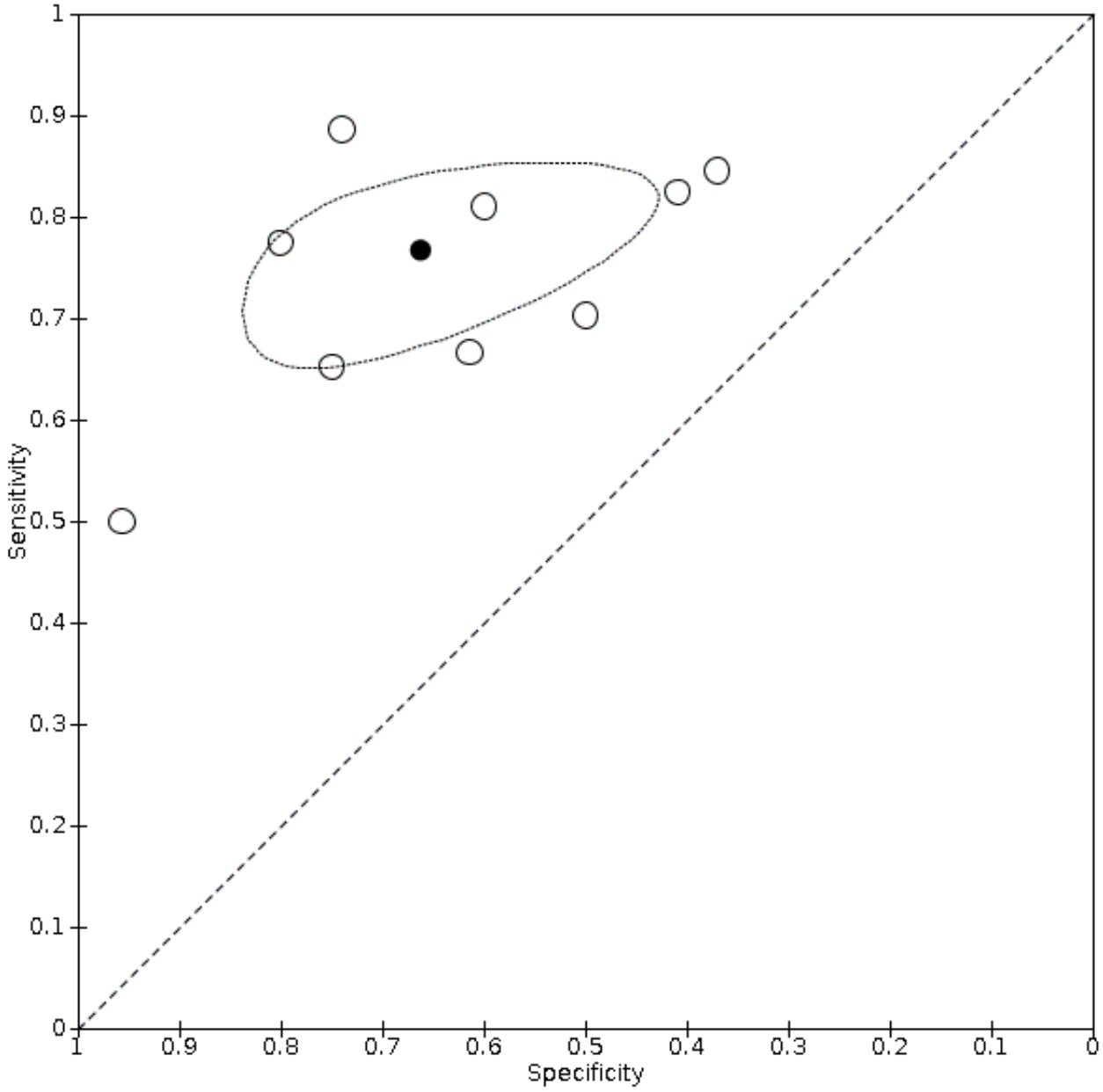
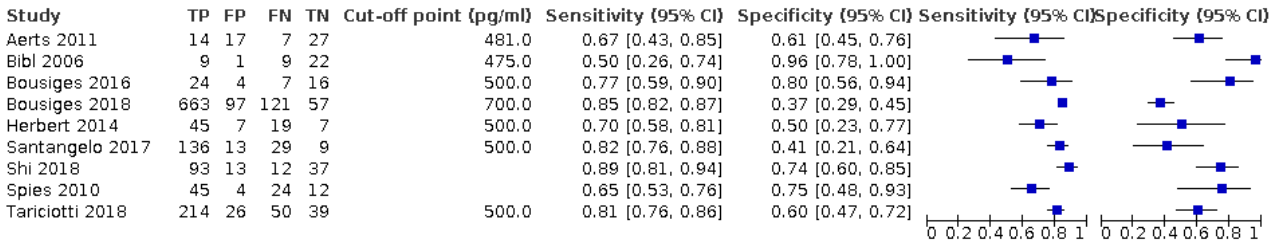


Figure 23. Forest plot of CSF ABeta42 for differentiating ADD from DLB (all studies).



In subgroup analysis, there were only sufficient studies investigating thresholds of less than or equal to 500 pg/ml to allow for meta-analysis (six studies, 751 participants, 563 with ADD) (Figure 24; Figure 25). The pooled sensitivity for studies using a threshold of less than or equal to 500 pg/ml was 79% (95% CI 69% to 86%), and the pooled specificity was 68% (95% CI 46% to 85%). Two studies did not specify the test threshold (Shi 2018; Spies 2010), and were excluded from the subgroup analysis. Only one study used a threshold above 500 pg/ml (700 pg/ml, Bousiges 2018); this study reported sensitivity 71% and specificity 53%.

Figure 24. Summary ROC Plot of CSF ABeta42 for differentiating ADD from DLB ≤ 500 (pg/ml). Summary statistics: sensitivity: 79% (95% CI 69%-86%), specificity: 68% (95% CI 45%-85%).

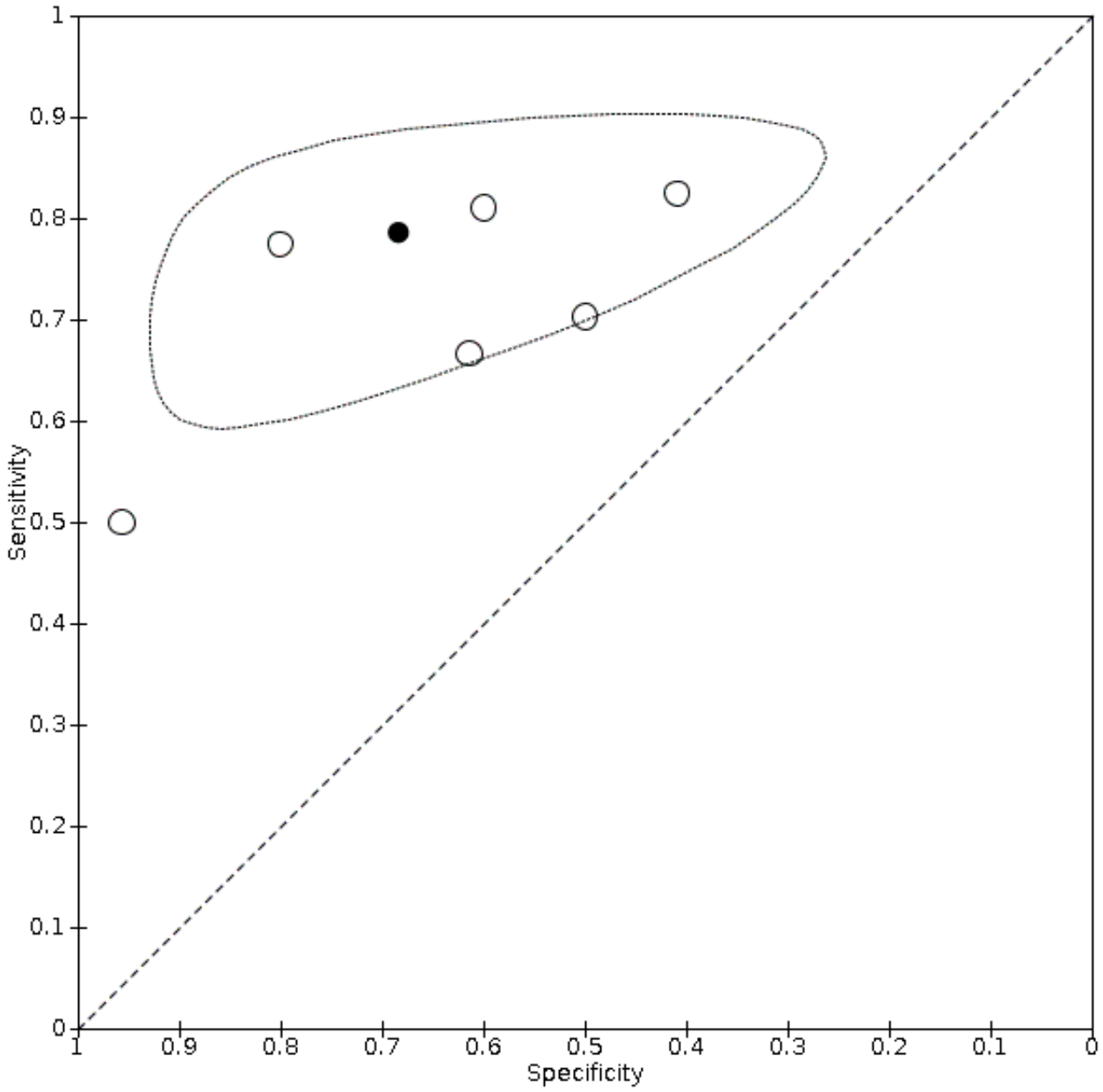
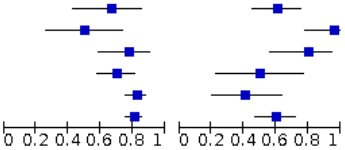


Figure 25. Forest plot of CSF ABeta42 for differentiating ADD from DLB ≤ 500 (pg/ml).

Study	TP	FP	FN	TN	Cut-off point (pg/ml)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Aerts 2011	14	17	7	27	481.0	0.67 [0.43, 0.85]	0.61 [0.45, 0.76]		
Bibl 2006	9	1	9	22	475.0	0.50 [0.26, 0.74]	0.96 [0.78, 1.00]		
Bousiges 2016	24	4	7	16	500.0	0.77 [0.59, 0.90]	0.80 [0.56, 0.94]		
Herbert 2014	45	7	19	7	500.0	0.70 [0.58, 0.81]	0.50 [0.23, 0.77]		
Santangelo 2017	136	13	29	9	500.0	0.82 [0.76, 0.88]	0.41 [0.21, 0.64]		
Tariciotti 2018	214	26	50	39	500.0	0.81 [0.76, 0.86]	0.60 [0.47, 0.72]		



CSF ABeta42 for differentiating ADD from NPH

The accuracy of ABeta42 to differentiate ADD from NPH related dementia subtypes was evaluated in a total of four studies (336 participants, 258 with ADD). The pooled sensitivity at all thresholds was 84% (95% CI 79% to 88%), and the pooled specificity was 42% (95% CI 26% to 60%) (Figure 26, Figure 27). There were insufficient studies for meta-analysis at different test thresholds.

Figure 26. Summary ROC Plot of CSF ABeta42 for differentiating ADD from vs NPH. Summary statistics: sensitivity: 84% (95% CI 79%-88%), specificity: 42% (95% CI 26%-60%).

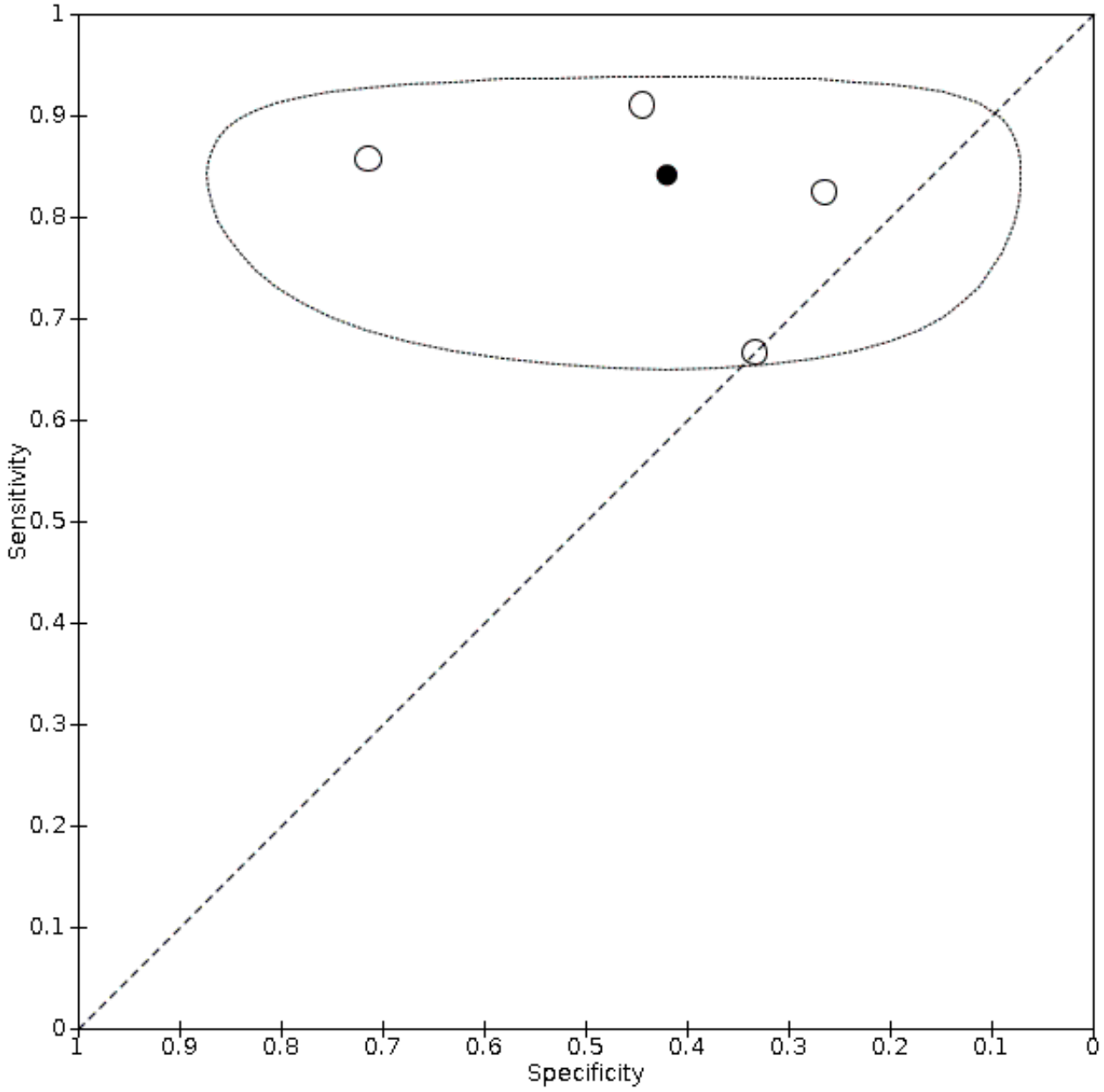
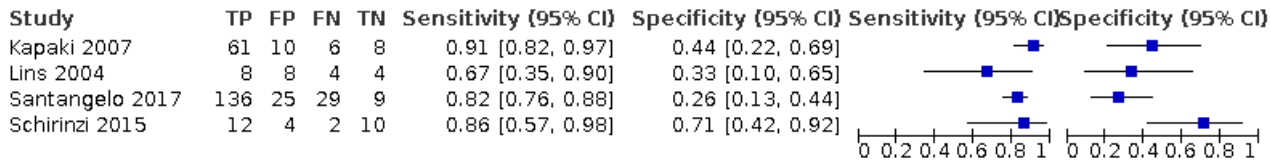


Figure 27. Forest plot of CSF ABeta42 for differentiating ADD from vs NPH.



CSF ABeta42 for differentiating ADD from CJD

The accuracy of ABeta42 to differentiate ADD from CJD subtypes was evaluated in a total of three studies (382 participants, 321 with ADD). The pooled sensitivity at all thresholds was 82% (94%CI:77% to 86%), and the pooled specificity was 46% (95% CI 34% to 58%) (Figure 28, Figure 29). There were insufficient studies for meta-analysis at different test thresholds.

Figure 28. Summary ROC Plot of CSF ABeta42 for differentiating ADD from CJD. Summary statistics: sensitivity: 82% (95% CI 77%-86%), specificity: 46% (95% CI 34%-58%).

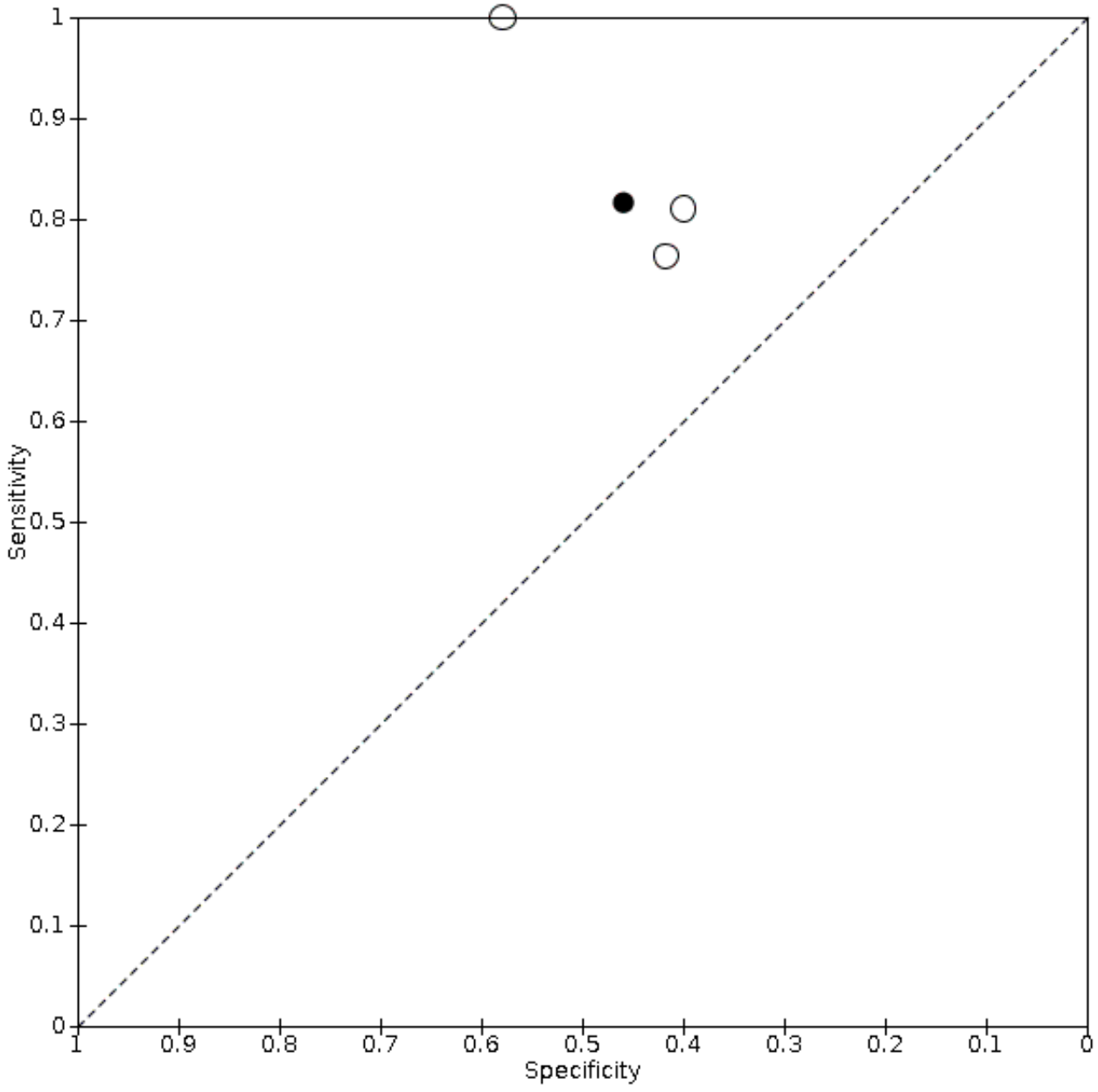
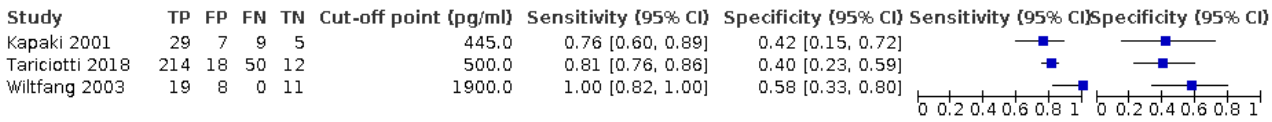


Figure 29. Forest plot of 1CSF ABeta42 for differentiating ADD from CJD.



CSF ABeta42 for differentiating ADD from ARCD

Only one study (53 participants, 33 with ADD) investigated the accuracy of ABeta42 to differentiate ADD from ARCD. Sensitivity was 80% and specificity was 85%.

Investigation of heterogeneity

We conducted sensitivity analyses for studies with a younger population of ADD participants, and studies with a drop-out rate of more than 30% of participants. Summary of findings 2 summarises the results of the subgroup analyses.

Effect of age

Three studies (Falgas 2020; Rosler 2001; Sjogren 2000) specifically enrolled participants with early-onset ADD (age equal to or under 65 years), corresponding to 100%, 40% and 62% of the ADD sample in each of the respective studies. Four studies had mean ages of under 66 years (Bibl 2007; Kapaki 2005; Knapskog 2018; Montine 2001), but did not specifically enrol participants with early-onset ADD. Kapaki 2005 was excluded from sensitivity analyses as data were only present for ADD versus ARCD (one study).

For ADD versus non-ADD, removal of three studies (Knapskog 2018; Montine 2001; Rosler 2001) did not substantially alter pooled estimates of sensitivity (79% versus 80%), or specificity (60% versus 62%).

Removal of one study (Sjogren 2000) in the ADD versus VaD analysis did not substantially alter pooled sensitivity (80% versus 80%), or specificity (69% versus 68%).

For ADD versus FTD, amongst three studies (Bibl 2007; Falgas 2020; Sjogren 2000) of younger participants, the pooled estimates of specificity (68% versus 86%), but not of sensitivity (85% versus 82%), were higher in younger than in older participants.

Effect of studies with high drop-out rates

Three studies (Herbert 2014; Santangelo 2017; Shi 2018) had drop-out rates, missing data, or excluded more than 30% of participant data.

For ADD versus VaD, removal of two studies (Herbert 2014; Santangelo 2017) did not substantially alter the pooled estimates of sensitivity (79% versus 79%), or specificity (69% versus 70%).

For ADD versus FTD, removal of three studies (Herbert 2014; Santangelo 2017; Shi 2018) did not substantially alter the pooled estimates of sensitivity (85% versus 81%) or specificity (72% versus 75%).

For ADD versus DLB, removal of three studies (Herbert 2014; Santangelo 2017; Shi 2018) did not substantially alter the pooled estimates of sensitivity (77% versus 75%), or specificity (66% versus 68%).

For ADD versus NPH, removal of one study (Santangelo 2017) also did not substantially alter the pooled estimates of sensitivity (84% versus 86%) or specificity (42% versus 49%).

Effect of studies without a pre-specified test threshold

For ADD versus non-ADD, removal of eight studies did not substantially alter pooled estimates of sensitivity (79% versus 79%) or specificity (60% versus 60%).

For ADD versus VaD, removal of eight studies did not substantially alter the pooled estimate of sensitivity (80% versus 80%), but the pooled estimate of specificity increased (73% versus 59%).

For ADD versus FTD, removal of 10 studies did not substantially alter the pooled estimates of sensitivity (81% versus 85%) or specificity (75% versus 72%).

For ADD versus DLB, removal of six studies did not substantially alter the pooled estimates of sensitivity (77% versus 70%) or specificity (66% versus 70%).

DISCUSSION

Summary of main results

We reviewed the diagnostic test accuracy of the ABeta42 biomarker for differential diagnosis in dementia. Specifically, we assessed accuracy of ABeta42 for differentiating ADD from other dementia subtypes. There were no suitable studies of plasma ABeta42 so our review evidence is limited to CSF-based studies.

In specialist settings, CSF ABeta42 may help differentiate ADD from other forms of dementia, but the test is imperfect. The pattern of higher sensitivity than specificity suggests that CSF ABeta42 is better at making a true ADD diagnosis than excluding other dementia types. The accuracy of ABeta42 for differentiating ADD was generally higher in those studies that compared a population of ADD and another specific dementia subtype; for example, vascular dementia. This situation does not mirror the real world, where patients present to memory clinics with undifferentiated memory problems and will include a variety of differing dementia subtypes. The studies that looked at differentiating ADD from mixed populations offer more generalisable data.

For those studies that assessed specific dementia pathologies, there was a suggestion that ABeta42 may work better at distinguishing certain dementia pathologies from ADD. This result has biological plausibility, as certain non-ADD types may involve abnormal amyloid production as part of the pathological cascade underlying the neurodegeneration.

We found that accuracy of CSF ABeta42 was dependent on the threshold used to define test positivity. The pattern of sensitivity and specificity will alter depending on the threshold employed. In this regard, it is disappointing that so few studies assessed CSF ABeta42 at a pre-specified threshold. Studies that explore various cut-off points until they find the threshold that works best are at risk of artificially inflating the test accuracy reported.

In general, we found that papers describing ABeta42 for differential diagnosis were at high risk of bias. This is a limitation that is common across much of the dementia biomarker literature.

Strengths and weaknesses of the review

We performed a systematic search of the literature, based on a sensitive search strategy. We followed best practice in all aspects of study selection, data extraction, quality assessment and meta-analysis.

Our interpretation is limited by issues with the included studies. None of the included studies were rated as low risk of bias across all the domains. Major issues were with patient selection and use of the index test (ABeta42). The ideal patient selection design would be random or consecutive enrolment. For many of the included studies, there was some degree of enrichment of the population, with researchers adding participants with the dementia subtypes of interest. For less common dementia subtypes, this approach may be necessary, unless very large populations can be included. However, this selection method risks bias, as the included patients may represent phenotypic extremes. The index test issue of greatest concern was around the choice of ABeta42 threshold used to define a positive test. There is no consensus on the optimal level of CSF ABeta42 to make an ADD diagnosis and limited agreement on levels to help determine one dementia subtype from another. To allow for a quantitative evidence synthesis, we accepted data from the threshold presented as the primary analysis in each parent study. Thus, there was no common threshold in our primary meta-analyses. Best practice in biomarker test accuracy studies is to pre-specify a threshold of interest. When we re-ran analyses at predefined thresholds of interest, we found that patterns of sensitivity and specificity were dependent on the threshold used. As biomarkers move from research tool to clinical practice, it is essential that consensus thresholds to define test positivity are agreed and used.

We pre-defined dementia subtypes of interest. However, there are potential further levels of granularity within these diagnostic groups. For example, FTD can be further subdivided into three main clinical categories, namely bvFTD, progressive non-fluent aphasia and semantic variant PPA. In addition to variable clinical presentation, these FTD subgroups are also genetically and pathologically heterogeneous. We were able to investigate the test accuracy of ABeta42 in two of the three FTD subgroups (bvFTD and PPA). Sensitivity to detect ADD was high in both subgroups, but specificity was considerably lower in the PPA compared to the bvFTD group. This suggests that certain clinical dementia classifications may be too broad, and biomarker-based diagnostics may be better suited to refined diagnosis. This aligns with the moves towards personalised medicine. We did not include a subgroup of 'mixed' dementia in our analyses, although this is probably one of the commonest dementia pathologies seen in older adults. Some argue that most dementia seen in older age is likely to have a degree of Alzheimer's disease and vascular pathology. If this is the case, then biomarkers specific to amyloid may be less helpful in this group.

Applicability of findings to the review question

We found no suitable studies assessing the test accuracy of plasma ABeta42. This is disappointing, as a biomarker that does not require invasive sampling of CSF would be preferable.

The analyses assessing ABeta42 for differentiating ADD from mixed dementias answer the question of greatest clinical relevance. The included studies were predominantly based in specialist secondary care settings. This is not a concern, as this is the setting where CSF biomarkers are at present most likely to be used. The case mix of participants in the studies did not always reflect the common diagnoses seen in general memory clinics, with a preponderance of more unusual dementia types. This is likely due to the highly specialist clinics participating in the studies.

Our condition of interest was the subtype of dementia, as assessed by clinical classification criteria. However, even the best validated clinical criteria are imperfect, and there are often differences between ante-mortem clinical diagnosis and post-mortem neuropathological diagnosis. Thus, it is possible that the accuracy data for ABeta42 are biased by erroneous clinical classification. In practice, clinical assessment, informed by informant review, neuroimaging and neuropsychological testing, remains the gold standard. In research, there is a move towards a biomarker-based diagnosis. There would be a circularity to comparing CSF ABeta42 to a pathological diagnosis based on amyloid beta testing, so clinicians will continue for now to use expert clinical assessment as the reference standard for now. We recognise, however, that dementia diagnostics is a rapidly evolving space, and best practice may change in the next years.

Our review answers the question: What is the accuracy of ABeta42 for distinguishing ADD from other dementias? However, this question assumes that the biomarker would be used in isolation. In practice, biomarkers will be used alongside clinical assessment, neuropsychological testing, and neuroimaging to inform a diagnostic formulation. A more pertinent question would be: What is the additive value of ABeta42 over usual practice for distinguishing ADD from other dementias?

AUTHORS' CONCLUSIONS

Implications for practice

Our results suggest that ABeta42 could be useful in improving differential diagnosis of the dementia syndrome, but the test is imperfect. As already discussed, it is unlikely that the ABeta42 biomarker would be used in isolation in clinical practice and ideally it should be used to support the diagnosis alongside full clinical, radiological, and neuropsychological assessment. Our review does not help answer questions around the added value of the test over routine diagnostics.

It is interesting that the test accuracy of cerebrospinal fluid (CSF) ABeta42 is similar to the accuracy seen in reviews of brief cognitive screening tools (Beishon 2019, Davis 2015, Quinn 2014). The studies are not comparable, but it does suggest that more expensive and more invasive tests are not necessarily better than the standard approach. Although a relatively safe procedure, CSF assessment via lumbar puncture has secondary complications and risks such

as headaches (Sadashivaiah 2009). There are also time and cost implications of this procedure and the subsequent assays. Before ABeta42 could be recommended for implementation at scale, there would need to be an assessment of feasibility, acceptability and economics.

The motivation for differentiating pathological dementia types is to allow personalised management of the dementia syndrome. At present, this is more of a theoretical issue than a practical concern. There are few approved drug treatments for dementia and no treatments specific to a certain dementia pathology. The main pharmacological intervention used in dementia care is symptomatic treatment with cholinesterase inhibitors or memantine. These agents seem to have a differential treatment response, dependent on dementia type. This supports the concept of tailoring drug therapy to the underlying pathology, although in international practice these agents are often prescribed for most dementia types anyway, perhaps due to the lack of any other therapeutic option.

As our understanding of the pathology underlying dementia improves, we find increasing evidence that the dementia of older age is often mixed with components from amyloid pathology, vascular disease, Lewy bodies etc. Our review did not include studies of 'mixed' dementias and the performance of the test in this group remains unknown.

Implications for research

Our review has implications for future dementia research and for future evidence synthesis of this research.

The test accuracy demonstrated does lend some support to the concept of using biomarkers to differentiate dementia type for tailored therapy. Clinical trials of anti-amyloid interventions could consider using quantification of ABeta42 for patient selection. As discussed above, the biomarker does not guarantee an exclusively ADD population, but it may help select those people most likely to benefit from the intervention. These two groups are not synonymous; a person with mixed dementia may not meet criteria for clinical ADD, but may still benefit from disruption of pathological amyloid pathways. The field of dementia biomarkers is rapidly evolving, other biomarkers and combinations of biomarkers are becoming available and it may be that a battery of biomarkers, rather than a single test, offers even greater precision in pathological diagnosis (Shaw 2009, Ritchie 2017). Such an approach is being used in projects such as EPAD (Ritchie 2016) and PREVENT Dementia (Ritchie 2012).

These arguments around utility of an ABeta biomarker to guide therapy only hold if the amyloid is the cause of the underlying neurodegeneration. This fundamental question remains unanswered. The relevance of amyloid pathology to clinical symptoms and dementia progression remains unclear and may be differential among different clinical syndromes; e.g. amyloidosis in vascular dementia may be a less potent driver of symptoms than in Alzheimer's disease dementia (ADD) (Iadecola 2014). Mechanistic research that explores the biological role of amyloid in neurodegeneration is still needed. Based on our results, such studies should not limit themselves to clinical ADD. Going forward, it will be important to understand interactions among pathologies and how they relate to risk factors and clinical phenotypes (Ritchie 2018).

As seen in other diagnostic test accuracy studies in dementia, we found issues with reporting of the science, which complicated our evidence synthesis. It would benefit the field to apply better and more consistent standards to the original research undertaken. Application of the Standards for Reporting of Diagnostic Accuracy Studies in Dementia (STARDdem) reporting checklist could help in this regard (Noel-Storr 2014). The clinical arguments around the need for greater consistency in the thresholds used to dichotomise ABeta42 are also true when considering research.

ACKNOWLEDGEMENTS

We would like to thank peer reviewers Suzanne Schindler and Charlotte Teunissen and consumer reviewer Alan Cassels for their comments and feedback.

REFERENCES

References to studies included in this review

Abu-Rumeileh 2018 *{published and unpublished data}*

Abu-Rumeileh S, Mometto N, Bartoletti-Stella A, Polischi B, Oppi F, Poda R, et al. Cerebrospinal fluid biomarkers in patients with frontotemporal dementia spectrum: a single-center study. *Journal of Alzheimer's Disease* 2018;**66**:551-563.

Aerts 2011 *{published data only}*

Aerts MB, Esselink RAJ, Claasen JAHR, Abdo WF, Bloem BR, Verbeek MM. CSF tau, Aβ42 and MHPG differentiate dementia with Lewy bodies from Alzheimer's disease. *Journal of Alzheimer's Disease* 2011;**27**:377-84.

Baldeiras 2015 *{published data only}*

Baldeiras I, Santana I, Joao Leitao M, Helena Ribeiro M, Pascoal R, Duro D, et al. Cerebrospinal fluid Aβ40 is similarly reduced in patients with frontotemporal lobar degeneration and Alzheimer's disease. *Journal of the Neurological Sciences* 2015;**358**(1-2):308-16.

Bibl 2006 *{published data only}*

* Bibl M, Mollenhauer B, Esselmann H, Lewczuk P, Trenkwalder C, Brechlin P, et al. CSF diagnosis of Alzheimer's disease and dementia with Lewy bodies. *Journal of Neural Transmission* 2006;**113**(11):1771-8.

Bibl 2007 *{published data only}*

Bibl M, Mollenhauer S, Wolf S, Esselmann H, Lewczuk P, Kornhuber J, et al. Reduced CSF carboxyterminally truncated Aβ peptides in frontotemporal lobe degenerations. *Journal of Neural Transmission* 2007;**114**:621-8.

Bousiges 2016 *{published data only}*

Bousiges O, Cretin B, Lavaux T, Philippi N, Jung B, Hezard S, et al. Diagnstic value of cerebrospinal fluid biomarkers (Phospho-Tau181, total-Tau, Aβ42, and Aβ40) in prodromal stage of Alzheimer's disease and dementia with Lewy bodies. *Journal of Alzheimer's Disease* 2016;**51**:1069-1083.

Bousiges 2018 *{published data only}*

Bousiges O, Bombois S, Schraen S, Wallon D, Quillard MM, Gabelle A, et al. Cerebrospinal fluid Alzheimer biomarkers can be useful for discriminating dementia with Lewy bodies from Alzheimer's disease at the prodromal stage. *Journal of Neurology, Neurosurgery & Psychiatry* 2018;**89**:467-475.

Brettschneider 2006 *{published data only}*

* Brettschneider J, Petzold A, Schöttle D, Claus A, Riepe M, Tumani H. The neurofilament heavy chain (NfH SMI35) in the cerebrospinal fluid diagnosis of Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders* 2006;**21**:291-95.

Casoli 2019 *{published and unpublished data}*

Casoli T, Paolini S, Fabbietti P, Fattoretti P, Paciaroni L, Fabi K, et al. Cerebrospinal fluid biomarkers and cognitive status in differential diagnosis of frontotemporal dementia and Alzheimer's disease. *Journal of International Medical Research* 2019;**47**:4968-4980.

de Jong 2006 *{published data only}*

* de Jong D, Jansen RWMM, Kremer BPH, Verbeek MM. Cerebrospinal fluid amyloid β_{42} / phosphorelated tau ratio discriminates between Alzheimer's disease and vascular dementia. *Journal of Gerontology* 2006;**61A**(7):755-8.

de Rino 2012 *{published data only}*

de Rino F, Martinelli-Boneschi F, Caso F, Zuffi M, Zabeo M, Passerini G. CSF metabolites in the differential diagnosis of Alzheimer's disease from frontal variant of frontotemporal dementia. *Neurological Sciences* 2012;**33**:973-7.

Falgas 2020 *{published data only}*

Falgas N, Ruiz-Peris M, Perez-Milan A, Sala-Llonch R, Antonell A, Balasa M. Contribution of CSF biomarkers to early-onset Alzheimer's disease and frontotemporal dementia neuroimaging signatures. *Human Brain Mapping* 2020;**41**:2004-2013.

Herbert 2014 *{published data only}*

Herbert MK, Aerts MB, Kuiperij BH, Claassen JA, Spies PE, Esselink RAJ, et al. Addition of MHPG to Alzheimer's disease biomarkers improves differentiation of dementia with Lewy bodies from Alzheimer's disease but not other dementias. *Alzheimer's and Dementia* 2014;**10**(4):448-455.

Kapaki 2001 *{published data only}*

* Kapaki E, Kilidireas K, Paraskevas GP, Michalopoulou M, Patsouris E. Highly increased CSF tau protein and decreased β-amyloid$_{1-42}$ in sporadic CJD: a discrimination from Alzheimer's disease? *Journal of Neurology, Neurosurgery and Psychiatry* 2001;**71**:401-3.

Kapaki 2003 *{published data only}*

* Kapaki E, Paraskevas GP, Zalonis I, Zournas C. CSF tau protein and beta-amyloid (1-42) in Alzheimer's disease diagnosis: discrimination from normal ageing and other dementias in the Greek population. *European Journal of Neurology* 2003;**10**(2):119-28.

Kapaki 2005 *{published data only}*

* Kapaki E, Lippas I, Paraskevas GP, Theotoka I, Rabavilas A. The diagnostic value of tau protein, β-amyloid$_{1-42}$ and their ratio for the discrimination of alcohol-related cognitive disorders from Alzheimer's disease in the early stages. *International Journal of Geriatric Psychiatry* 2005;**20**:722-9.

Kapaki 2007 *{published data only}*

* Kapaki EN, Paraskevas GP, Tzerakis NG, Sfagos C, Seretis A, Kararizou E, et al. Cerebrospinal fluid tau, phospho-tau181 and b-amyloid1)42 in idiopathic normal pressure hydrocephalus: a discrimination from Alzheimer's disease. *European Journal of Neurology* 2007 Feb;**14**(2):168-173.

Kapaki 2008 *{published data only}*

Kapaki E, Paraskevas GP, Papageorgiou SG, Bonakis A, Kalfakis N, Zalonis I, et al. Diagnostic value of CSF biomarker

profile in frontotemporal lobar degeneration. *Alzheimer Disease and Associated Disorders* 2008;**22**(1):47-53.

Khoonsari 2019 *{published and unpublished data}*

Khoonsari PE, Shevchenko G, Herman S, Remnestal J, Giedraitis V, Brundin RM, et al. Improved differential diagnosis of Alzheimer's disease by integrating ELISA and mass spectrometry-based cerebrospinal fluid biomarkers. *Journal of Alzheimer's Disease* 2019;**67**:639-651.

Knapskog 2018 *{published and unpublished data}*

Knapskog AB, Braekhus A, Engedal K. The effect of changing the amyloid β42 cut-off of cerebrospinal fluid biomarkers on Alzheimer disease diagnosis in a memory clinic population in Norway. *Alzheimer Disease and Associated Disorders* 2019;**33**:72-74.

Lewczuk 2004 *{published data only}*

* Lewczuk P, Esselmann H, Otto M, Maler JM, Henkel AW, Henkel MK, et al. Neurochemical diagnosis of Alzheimer's dementia by CSF A42, A42/A40 ratio and total tau. *Neurobiology of Aging* 2004;**25**:273–281.

Lins 2004 *{published data only}*

* Lins H, Wichart I, Bancher C, Walle C, Jellinger KA, Rosler N. Immunoreactivities of amyloid beta peptide((1-42)) and total tau protein in lumbar cerebrospinal fluid of patients with normal pressure hydrocephalus. *Journal of Neural Transmission* 2004;**111**(3):273-80.

Lombardi 2018 *{published and unpublished data}*

Lombardi G, Polito C, Berti V, Ferrari C, Lucidi G, Begnoli S, et al. Biomarkers study in atypical dementia: proof of a diagnostic work-up. *Neurological Sciences* 2018;**39**:1203-1210.

Maddalena 2003 *{published data only}*

* Maddalena A, Papassotiropoulos A, Muller-Tillmanns B, Jung HH, Hegi T, Nitsch RM, et al. Biochemical diagnosis of Alzheimer disease by measuring the cerebrospinal fluid ratio of phosphorylated tau protein to beta-amyloid peptide 42. *Archives of Neurology.* 2003;**60**(9):1202-1206.

Marchegiani 2019 *{published and unpublished data}*

Marchegiani F, Matacchione G, Ramini D, Marcheselli F, Recchioni R, Casoli T, et al. Diagnostic performance of new and classic CSF biomarkers in age-related dementias. *Aging (Albany NY)* 2019;**11**:2420-2429.

Montine 2001 *{published data only}*

* Montine TJ, Kaye JA, Montine KS, McFarland L, Morrow JF, Quinn JF. Cerebrospinal fluid abeta42, tau, and f2-isoprostane concentrations in patients with Alzheimer disease, other dementias, and in age-matched controls. *Archives of Pathology and Laboratory Medicine.* 2001;**125**(4):510-512.

Paraskevas 2009 *{published data only}*

* Paraskevas GP, Kapaki E, Papageorgiou SG, Kalfakis N, Andreadou E, Zalonis I, et al. CSF biomarker profile and diagnostic value in vascular dementia. *European Journal of Neurology.* 16;**2**:205-11.

Perani 2016 *{published data only}*

Perani D, Cerami C, Caminiti SP, Santangelo R, Coppi E, Ferrari L, et al. Cross-validation of biomarkers to early differential diagnosis and prognosis of dementia in a clinical setting. *European Journal of Nuclear Medicine and Molecular Imaging* 2015;**43**(3):499-508.

Rosler 2001 *{published data only}*

* Rosler N, Wichart I, Jellinger KA. Clinical significance of neurobiochemical profiles in the lumbar cerebrospinal fluid of alzheimer's disease patients. *Journal of Neural Transmission.* Jun 2001;**108**(2):231-246.

Santangelo 2017 *{published data only}*

Santangelo R, Cecchetti G, Bernasconi MP, Cardamone R, Barbieri A, Pinto P, et al. Cerebrospinal fluid amyloid-β 42, total tau and phosphorylated tau are low in patients with normal pressure hydrocephalus: analogies and differences with Alzheimer's disease. *Journal of Alzheimer's Disease* 2017;**60**:183-200.

Schirinzi 2015 *{published data only}*

Schirinzi T, Sancesario GM, Lalongo C, Imbriani P, Madeo G, Toniolo S, et al. A clinical and biochemical analysis in the differential diagnosis of idiopathic normalpressure hydrocephalus. *Frontiers in Neurology* 2015;**6**(86).

Shi 2018 *{published data only}*

Shi M, Tang L, Toledo J, Ginghina C, Wang H, Ar P, et al. CSF α-synuclein contributes to the differential diagnosis of Alzheimer disease. *Alzheimers Dementia* 2018;**14**:1052-1062.

Sjogren 2000 *{published data only}*

* Sjogren M, Minthon L, Davidsson P, Granerus AK, Clarberg A, Vanderstichele H, et al. CSF levels of tau, beta-amyloid 1-42 and GAP-43 in frontotemporal dementia, other types of dementia and normal aging. *Journal of Neural Transmission.* 2000;**107**(5):563-79.

Smach 2008 *{published data only}*

* Smacha MA, Charfeddinea B, Lammouchib T, Harrabic I, Othmana LB, Dridia H, et al. CSF -amyloid 1–42 and tau in Tunisian patients with Alzheimer's disease: The effect of APOE 4 allele. *Neuroscience Letters* 2008;**440**:145-149.

Spies 2010 *{published data only}*

* Spies PE, Slats D, Sjögren JM, Kremer BP, Verhey FR, Rikkert MG, et al. The cerebrospinal fluid amyloid beta42/40 ratio in the differentiation of Alzheimer's disease from non-Alzheimer's dementia. *Current Alzheimer Research* 2010;**7**(5):470-6.

Stefani 2005 *{published data only}*

* Stefani A, Bernardini S, Panella M, Pierantozzi M, Nuccetelli M, Kocha G, et al. AD with subcortical white matter lesions and vascular dementia: CSF markers for differential diagnosis. *Journal of the Neurological Sciences* 2005;**237**:83-88.

Tapiola 2000 *{published data only}*

* Tapiola T, Pirttila T, Mehta PD, Alafuzoff I, Lehtovirta M, Soininen H. Relationship between apoE genotype and CSF beta-

amyloid (1-42) and tau in patients with probable and definite Alzheimer's disease. *Neurobiology of Aging* 2000;**21**(5):735-40.

Taricciotti 2018 *{published data only}*

Taricciotti L, Casadei M, Honig LS, Teich AF, McKhann II GM, Tosto G, et al. Clinical experience with cerebrospinal fluid $A\beta_{42}$, total and phosphorylated tau in the evaluation of 1,016 individuals for suspected dementia. *Journal of Alzheimer's Disease* 2018;**65**:1417-1425.

Wiltfang 2003 *{published data only}*

* Wiltfang J, Esselmann H, Smirnov A, Bibl M, Cepek L, Steinacker P, et al. β-amyloid peptides in cerebrospinal fluid of patients with Creutzfeldt–Jakob disease. *Annals of Neurology* 2003;**54**:263-67.

References to studies excluded from this review

Alcolea 2014 *{published data only}*

Alcolea D, Carmona-Iragui M, Suarez-Calvet M, Sanchez-Saudinos MB, Sala I, Anton-Aguirre S, et al. Relationship between beta-secretase, inflammation and core cerebrospinal fluid biomarkers for Alzheimer's disease. *Journal of Alzheimer's Disease* 2014;**42**(1):157-67.

Alcolea 2017 *{published data only}*

Alcolea D, Vilaplana E, Suarez-Calvet, Ilan-Gala I, Blesa R, Clarimon J, et al. CSF sAPPbeta, YKL-40, and neurofilament light in frontotemporal lobar degeneration. *American Academy of Neurology* 2017;**89**(2):178-188.

Balasa 2014 *{published data only}*

Balasa M, Sanchez-Valle R, Antonell A, Bosch B, Olives J, Rami L, et al. Usefulness of biomarkers in the diagnosis and prognosis of early-onset cognitive impairment. *Journal of Alzheimer's Disease* 2014;**40**:919-927.

Berlyand 2016 *{published data only}*

Berlyand Y, Weintraub D, Xie SX, Mellis IA, Doshi J, Rick J, et al. An Alzheimer's disease-derived biomarker signature identifies Parkinson's disease patients with dementia. *PloS One* 2016;**11**.

Bertens 2017 *{published data only}*

Bertens D, Tijms BM, Scheltens P, Teunissen CE, Visser PJ. Unbiased estimates of cerebrospinal fluid beta-amyloid 1-42 cutoffs in a large memory clinic population. *Alzheimer's Research and Therapy* 2017;**9**.

Bibl 2007b *{published data only}*

Bibl M, Esselmann H, Mollenhauer B, Weniger G, Welge V, Liess M et al. Blood-based neurochemical diagnosis of vascular dementia: a pilot study. *Journal of Neurochemistry* 2007;**103**:467-74.

Bibl 2008a *{published data only}*

Bibl M, Lewczuk P, Esselmann H, Mollenhauer B, Klafki HW, Welge V et al. CSF amyloid-β 1-38 and 1-4 2in FTD and AD: biomarker performance critically depends on the detergent accessible fraction. *Proteomics - Clinical Applications* 2008;**2**:1548-56.

Brandt 2008 *{published data only}*

Brandt C, Bahl JC, Heegaard NH, Waldemar G, Johannsen P. Usability of cerebrospinal fluid biomarkers in a tertiary memory clinic. *Dementia and Geriatric Cognitive Disorders* 2008;**25**(6):553-8.

Carandini 2019 *{published data only}*

Carandini T, Arighi A, Sacchi L, Fumagalli GG, Pietroboni AM, Ghezzi L, et al. Testing the 2018 NIA-AA research framework in a retrospective large cohort of patients with cognitive impairment: from biological biomarkers to clinical syndromes. *Alzheimer's Research and Therapy* 2019;**11**.

Hall 2012 *{published data only}*

Hall S, Ohrfelt A, Constantinescu R, Andreasson U, Surova Y, Bostrom F, et al. Accuracy of a panel of 5 cerebrospinal fluid biomarkers in the differential diagnosis of patients with dementia and/or parkinsonian disorders. *Archives of Neurology* 2012;**69**(11):1445-1452.

Hampel 2018 *{published data only}*

Hampel H, Toschi N, Baldacci F, Zetterberg H, Blennow K, Kilimann I. Alzheimer's disease biomarker-guided diagnostic workflow using the added value of six combined cerebrospinal fluid candidates: abeta1-42, total-tau, phosphorylated-tau, NFL, neurogranin, and YKL-40. *Alzheimer's and Dementia* 2018;**14**:492-501.

Han 2012 *{published data only}*

Han Y, Jia J, Jia X-F, Qin W, Wang S. Combination of plasma biomarkers and clinical data for the detection of sporadic Alzheimer's disease. *Neuroscience Letters* 2012;**516**:232-6.

Illan-gala 2019 *{published data only}*

Illan-Gala I, Pegueroles J, Montal V, Alcolea D, Vilaplana E, Bejanin A, et al. APP-derived peptides reflect neurodegeneration in frontotemporal dementia. *Annals of Clinical and Translational Neurology* 2019;**6**:2518-2530.

Karadas 2017 *{published data only}*

Karadas O, Koc G, Ozon AO, Ozturk B, Konukoglu D. Biomarkers of Alzheimer's disease and vascular dementia simultaneously sampled from serum and cerebrospinal fluid. *Turkish Journal of Geriatrics* 2017;**20**(1):1-7.

Parnetti 2011 *{published data only}*

Parnetti L, Chiasserini D, Bellomo G, Giannandrea D, de Carlo C, Qureshi MM et al. Cerebrospinal fluid tau/a-Synuclein ratio in Parkinson's disease and degenerative dementias. *Movement Disorders* 2011;**26**(8):1429-35.

Prikrylova Vranova 2014 *{published data only}*

Prikrylova Vranova H, Henykova E, Kaiserova M, Mensikova K, Vastik M, Mares J, et al. Tau protein, beta-amyloid 1-42 and clusterin CSF levels in the differential diagnosis of Parkinsonian syndrome with dementia. *Journal of the Neurological Sciences* 2014;**343**:120-124.

Skillback 2015 *{published data only}*

Skillback T, Farahmand BY, Rosen C, Mattsson N, Nagga K, Kilander L, et al. Cerebrospinal fluid tau and amyloid -β1-42 in patients with dementia. *Brain* 2015;**138**:2716-2731.

Smach 2008a *{published data only}*

Smach MA, Charfeddine B, Lammouchi T, Dridi H, Othman LB, Bennamou S, et al. Interest of CSF –amyloid1-42 and t-tau protein level determinations for the diagnosis of Alzheimer's disease [Intérêt du dosage de la protéine amyloïde Aβ1-42 et de la protéine tau dans le LCR pour le diagnostic de la maladie d'Alzheimer: une étude tunisienne]. *Annales de Biologie Clinique* 2008;**66**(5):531-5.

Stoeck 2014 *{published data only}*

Stoeck K, Schmitz M, Ebert E, Schmidt C, Zerr I. Immune responses in rapidly progressive dementia: a comparative study of neuroinflammatory markers in Creutzfeldt-Jakob disease, Alzheimer's disease and multiple sclerosis. *Journal of Neuroinflammation* 2014;**11**(170).

Toledo 2012 *{published data only}*

Toledo JB, Brettschneider J, Grossman M, Arnold SE, Hu WT, Xie SX et al. CSF biomarkers cutoffs: the importance of coincident neuropathological disaease. *Acta Neuropathologica* 2012;**124**(1):23-35.

Uslu 2012 *{published data only}*

Uslu S, Akarkarasu ZE, Ozbabalik D, Ozkan S, Colak O, Demirkan ES et al. Levels of amyloid beta-42, interleukin-6 and tumor necrosis factor-alpha in Alzheimer's disease and vascular dementia. *Neurochemical Research* 2012;**37**:1554-9.

van Steenoven 2018 *{published data only}*

van Steenoven I, Majbour NK, Vaikath NN, Berendse HW, van der Flier WM, van de Berg WDJ, et al. Alpha-synuclein species as potential cerebrospinal fluid biomarkers for dementia with Lewy bodies. *Movement Disorders* 2018;**33**:1724-1733.

van Steenoven 2019 *{published data only}*

van Steenoven I, Noli B, Cocco C, Ferri GL, Oeckl P, Otto M, et al. VGF peptides in cerebrospinal fluid of patients with dementia with Lewy bodies. *International Journal of Molecular Sciences* 2019;**20**:4674.

Vergallo 2017 *{published data only}*

Vergallo A, Carlessi C, Pagni C, Sean Giorgi F, Baldacci F, Petrozzi L, et al. A single centre study: Aβ42/p-Tau181 CSF ratio to discriminate AD from FTD in clinical setting. *Neurological Sciences* 2017;**38**:1791-1787.

Wennstrom 2015 *{published data only}*

Wennstrom M, Hall S, Nagga K, Londos E, Minthon L, Hansson O. Cerebrospinal fluid levels of IL-6 are decreased and correlate with cognitive status in DLB patients. *Alzheimer's Research and Therapy* 2015;**7**(63).

Zwan 2014 *{published data only}*

Zwan M, van Harten A, Ossenkoppele R, Bouwman F, Teunissen C, Adriaanse S, et al. Concordance between cerebrospinal fluid biomarkers and [11C]PIB PET in a memory clinic cohort. *Journal of Alzheimer's Disease* 2014;**41**:801-807.

Additional references

Albert 2011

Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendation from the National Institute on Ageing and Alzheimer's Association workgroup. *Alzheimer's & Dementia* 2011;**7**(3):270-9.

APA 1987

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 3rd revised edition. American Psychiatric Association 1987.

APA 1994

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th edition. American Psychiatric Association 1994.

APA 2000

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th edition (revised). American Psychiatric Association, Washington, DC 2000.

Beach 2012

Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer's disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. *Journal of Neuropathology and Experimental Neurology* 2012;**71**.

Beishon 2019

Beishon LC, Batterham AP, Quinn TJ, Nelson CP, Panerai RB, Robinson T, Haunton VJ. Addenbrooke's Cognitive Examination III (ACE-III) and mini-ACE for the detection of dementia and mild cognitive impairment. *Cochrane Database of Systematic Reviews* 2019, Issue 12. Art. No: CD013282. [DOI: 10.1002/14651858.CD013282.pub2]

Beyer 2009

Beyer K, Domingo-Sabat M, Ariza A. Molecular pathology of Lewy body diseases. *International Journal of Molecular Science* 2009;**10**:724-745.

Beynon 2013

Beynon R, Leeflang MM, McDonald S, Eisinga A, Mitchell RL, Whiting P, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *Cochrane Database of Systematic Reviews* 2013;(9):Art. No.: MR000022 DOI: 10.1002/14651858.MR000022.pub3.

Blurton-Jones 2006

Blurton-Jones M, Laferla FM. Pathways by which Abeta facilitates tau pathology. *Current Alzheimer Research* 2016;**3**(5):437-48.

Boxer 2005

Boxer AL, Miller BL. Clinical features of frontotemporal dementia. *Alzheimer Disease and Associated Disorders* 2005;**19**:S3-6.

Burns 2005

Burns A, O'Brien J, Ames D. Dementia. Oxford University Press 2005.

Chui 1992

Chui HC, Victoroff JI, Margolin D, Jagust W, Shankle R, Katzman R. Criteria for the diagnosis of ischemic vascular dementia proposed by the State of California Alzheimer's Disease Diagnostic and Treatment Centers. *Neurology* 1992;**42**(3 pt 1):473-80.

Cochrane 2020 [Computer program]

The Cochrane Collaboration Review Manager (RevMan). The Cochrane Collaboration, Version Version 5.4. The Cochrane Collaboration, 2020.

Cummings 2019

Cummings J. The role of biomarkers in Alzheimer's disease drug development. *Advances in Experimental Medicine and Biology* 2019;**1118**:29-61.

Davis 2015

Davis D, Creavin S, Yip J, Noel-Storr A, Brayne C, Cullum S. Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. *Cochrane Database of Systematic Reviews* 2015, Issue 10. Art. No: CD010775. [DOI: 10.1002/14651858.CD010775.pub2]

De Strooper 2010

De Strooper B, Vassar R, Golde T. The secretases: Enzymes with therapeutic potential in Alzheimer's disease. *Nature Reviews: Neurology* 2010;**6**:99-107.

DeTure 2019

DeTure MA, Dickson DW. The neuropathological diagnosis of Alzheimer's disease. *Molecular Neurodegeneration* 2019;**14**.

Dubois 2007

Dubois B, Feldman HH, Jacova C, DeKosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. *Lancet Neurology* 2007;**6**:734-746.

Dubois 2010

Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, et al. Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurolology* 2010;**9**(11):1119-27.

Dubois 2014

Dubois B, Feldman HH, Jacova C, Hampal H, Molinuevo JL, DeKosky ST, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurology* 2014;**6**:614-29.

Fantoni 2018

Fantoni ER, Chalkidou A, O' Brien JT, Farrar G, Hammers A. A systematic review and aggregated analysis on the impact of amyloid PET brain imaging on the diagnosis, diagnostic confidence, and management of patients being evaluated for Alzheimer's disease. *Journal of Alzheimer's Disease* 2018;**63**:783-796.

Freeman 2019

Freeman SC, Kerby CR, Patel A, Cooper NJ, Quinn T, Sutton AJ. Development of an interactive web-based tool to conduct and interrogate meta-analysis of diagnostic test accuracy studies. *BMC Medical Research Methodology* 2019;**19**:81.

Hakim 1965

Hakim S, Adams RD. The special clinical problem of symptomatic hydrocephalus with normal cerebrospinal fluid pressure. Observations on cerebrospinal fluid hydrodynamics. *Journal of the Neurological Sciences* 1965;**2**(4):307-327.

Hansson 2019

Hansson O, Lehmann S, Otto M, Zetterberg H, Lewczuk P. Advantages and disadvantages of the use of the CSF Amyloid β (Aβ) 42/40 ratio in the diagnosis of Alzheimer's Disease. *Alzheimer's Research & Therapy* 2019;**11**.

Iadecola 2014

Iadecola C. The pathobiology of vascular dementia. *Neuron* 2013;**4**.

Iadecola 2019

Iadecola C, Duering M, Hachinski V, Joutel A, Pendelbury S, Schneider J, et al. Vascular cognitive impairment and dementia. *Journal of the American College of Cardiology* 2019;**73**:3326-3334.

Jansen WJ 2015

Jansen WJ, Ossenkoppele R, Knol D, Tijms BM, Scheltens P, Verhey FR, et al. Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis. *JAMA* 2015;**313**(19):1924-38.

Karantzoulis 2011

Karantzoulis S, Galvin JE. Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Review of Neurotherapeutics* 2011;**11**:1579-1591.

Khoury 2019

Khoury R, Ghossoub E. Diagnostic biomarkers of Alzheimer's disease: A state-of-the-art review. *Biomarkers in Neuropsychiatry* 2019;**1**:100005.

Klohs 2019

Klohs J. An integrated view on vascular dysfunction in Alzheimer's disease. *Neurodegenerative Diseases* 2019;**19**:109-127.

Knapp 2007

Knapp, M, Prince, M. Dementia UK. Alzheimer's Society 2014. Available from: https://www.alzheimers.org.uk/sites/default/files/2018-10/Dementia_UK_Full_Report_2007.pdf?fileID=2.

Kril 1999

Kril JJ, Halliday GM. Brain shrinkage in alcoholics: a decade on and what have we learned? *Progress in Neurobiology* 1999;**58**(4):381-7.

Lopes 2010

Lopes MA, Furtado EF, Ferrioli E, Litvoc J, Bottino CM. Prevalence of alcohol-related problems in an elderly population and their association with cognitive impairment and dementia. *Alcoholism: Clinical and Experimental Research* 2010;**34**(4):726-33.

Lopez 1999

Lopez OL, Litvan I, Catt KE, Stowe R, Klunk W, Kaufer DI, et al. Accuracy of four clinical diagnostic criteria for the diagnosis of neurodegenerative dementias. *Neurology* 1999;**53**(6).

Lund Manchester Groups 1994

The Lund Manchester Groups. Clinical and neuropathological criteria for frontotemporal dementia. *Journal of Neurology, Neurosurgery, and Psychiatry* 1994;**57**(4):416-8.

McKeith 1996

McKeith IG, Galasko D, Kosaka K, Perry EK, Dickson DW, Hansen LA, et al. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop. *Neurology* 1996;**47**(5):1113-24.

McKeith 2002

McKeith IG. Dementia with Lewy bodies. *British Journal of Psychiatry* 2002;**180**:144-7.

McKeith 2005

McKeith IG, Dickson DW, Lowe J, Emre M, O'Brien JT, Feldman H, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology* 2005;**65**(12):1863-72.

McKhann 1984

McKhann GM, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;**34**(7):939-44.

McKhann 2011

McKhann GM, Knopman DS, Chertkow H, Hyman BT, Clifford RJ, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendation from the National Institute on Ageing and Alzheimer's Association workgroup. *Alzheimer's & Dementia* 2011;**7**(3):263-9.

Murphy 2010

Murphy MP, LeVine H. Alzheimer's disease and the β-amyloid peptide. *Journal of Alzheimer's Disease* 2010;**19**:311.

Nakamura 2018

Nakamura A, Kaneko N, Villemagne VL, Kato T, Doecke J, Dore V, et al. High performance plasma amyloid-beta biomarkers for Alzheimer's disease. *Nature* 2018;**554**:249-254.

Neary 1998

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**:1546-54.

NICE 2018

National Institute of Clinical Excellence (NICE). Dementia: assessment, management and support for people living with dementia and their carers. NICE 2018.

Niemantsverdriet 2017

Niemantsverdriet E, Valckx S, Bjerke M, Engelborghs S. Alzheimer's disease CSF biomarkers: clinical indications and rational use. *Acta Neurologica Belgica* 2017;**117**:591-602.

Noel-Storr 2014

Noel-Storr A, McCleery JM, Richard E, Ritchie CW, Flicker L, Cullum SJ, et al. Reporting standards for studies of diagnostic test accuracy in dementia: The STARDdem Initiative. *Neurology* 2014;**83**(4):364-73.

O'Brien 2017

O'Brien JT, Holmes C, Jones M, Jones R, Livingston G, McKeith I, et al. Clinical practice with anti-dementia drugs: A revised (third) consensus statement from the British Association for Psychopharmacology. *Journal of Psychopharmacology* 2017;**31**:147-168.

Ossenkoppele 2015

Ossenkoppele R, Jansen WJ, Rabinovici GD, Knol DL, van der Flier WM, van Berckel BNM, et al. Prevalence of amyloid PET positivity in dementia syndromes: a meta-analysis. *JAMA* 2015;**313**:1939-1949.

Otto 2000

Otto M, Esselmann H, Schulz-Shaeffer W, Neumann M, Schröter A, Ratzka P, et al. Decreased ß-amyloid1-42 in cerebrospinal fluid of patients with Creutzfeldt-Jakob disease. *Neurology* 2000;**54**(5):1099-102.

Outeiro 2019

Outeiro TF, Koss DJ, Erskine D, Walker L, Kurzawa-Akanbi M, Burn D, et al. Dementia with Lewy bodies: an update and outlook. *Molecular Neurodegeneration* 2019;**14**:5.

Parkkinen 2008

Parkkinen L, Pirttilä T, Alafuzoff I. Applicability of current staging/categorization of á-synuclein pathology and their clinical relevance. *Acta Neuropathologica* 2008;**115**(4):399-407.

Patel 2020

Patel A, Cooper NJ, Freeman SC, Sutton AJ. Graphical enhancements to summary receiver operating characteristic plots to facilitate the analysis and reporting of meta-analysis of diagnostic test accuracy data. *Research Synthesis Methods* 2020:1-11.

Prince 2015

Prince M, Wimo A, Guerchet M, Ali GC, Wu YT, Prina M. World Alzheimer Report 2015 The Global Impact of Dementia An

analysis of prevalence, incidence, cost and trends. Alzheimer's Disease International 2015:1-84.

Quinn 2012

Quinn TJ, McShane R, Fearon P, Young C, Noel-Storr A, Stott DJ. IQCODE for the diagnosis of Alzheimer's disease dementia and other dementias within a community setting. *Cochrane Database of Systematic Reviews* 2012, Issue 9. Art. No: CD010079. [DOI: 10.1002/14651858.CD010079]

Quinn 2014

Quinn T, Fearon P, Noel-Storr A, Young C, McShane R, Stott D. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations. *Cochrane Database of Systematic Reviews* 2014, Issue 4. Art. No: CD010079. [DOI: 10.1002/14651858.CD010079.pub2]

Rabinovici 2019

Rabinovici GD, Gatsonis C, Apgar C, Chaudhary K, Gareen I, Hanna L, et al. Association of Amyloid Positron Emission Tomography with Subsequent Change in Clinical Management Among Medicare Beneficiaries With Mild Cognitive Impairment or Dementia. *JAMA* 2019;**321**(13).

Ritchie 2012

Ritchie CW, Ritchie K. The PREVENT study: a prospective cohort study to identify mid-life biomarkers of late-onsetAlzheimer's disease. *BMJ Open* 2012;**2**(6).

Ritchie 2014

Ritchie C, Smailagic N, Noel-Storr AH, Takwoingi Y, Flicker L, Mason SE, McShane R. Plasma and cerebrospinal fluid amyloid beta for the diagnosis of Alzheimer's disease dementiaand other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev* 2014;**6**.

Ritchie 2016

Ritchie CW, Molinuevo JL, Truyen L, Satlin A, Van der Geyten S, Lovestone S. Development of interventions for the secondary prevention of Alzheimer's dementia: the EuropeanPrevention of Alzheimer's Dementia (EPAD) project. *Lancet Psychiatry* 2016;**3**(2):179-86.

Ritchie 2017

Ritchie C, Smailagic N, Noel-Storr A, Ukoumunne O, Ladds EC, Martin S. CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementia in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews* 2017, Issue 3. Art. No: CD010803. [DOI: 10.1002/14651858.CD010803.pub2]

Ritchie 2018

Ritchie CW, Muniz-Terrera G. Models for dementia risk prediction: so much activity brings a need for coordination and clarity. *Journal of Neurology, Neurosurgery and Psychiatry* 2018;**90**:372.

Rizzo 2018

Rizzo G, Arcuti S, Copetti M, Alessandria M, Savica R, Fontana A, Liguori R, Logroscino G. Accuracy of clinical diagnosis of dementia with Lewy bodies: a systematic review and meta-analysis. *Journal of Neurology, Neurosurgery, and Psychiatry* 2018;**89**(4).

Robinson 2015

Robinson L, Taylor JP. Dementia: timely diagnosis and early intervention. *BMJ* 2015;**350**:h3029.

Roman 1993

Roman GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, et al. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. Neurology 1993;**43**(2):250-60.

Ryan 2018

Ryan J, Fransquet P, Wrigglesworth J, Lacaze P. Phenotypic heterogeneity in dementia: a challenge for epidemiology and biomarker studies. *Frontiers of Public Health* 2018;**6**:181.

Sadashivaiah 2009

Sadashivaiah J, McLure. Tuohy needle can reduce the incidence of severe post dural puncture headache. *Anaesthesia* 2009;**64**(12):1379-80.

Shaw 2009

Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals of Neurology* 2009;**65**(4).

Takami 2009

Takami M, et al. Gamma-secretase: Successive tripeptide and tetrapeptide release from the transmembrane domain of beta-carboxyl terminal fragment. *J Neurosci* 2009;**29**:13042-13052.

Thomas 2001

Thomas VS, Rockwood KJ. Alcohol abuse, cognitive impairment, and mortality among older people. Journal of the American Geriatrics Society 2001;**49**(4):415-20.

Van Everbroeck 1999

Van Everbroeck B, Green AJ, Pals P, Martin JJ, Cras P. Decreased levels of amyloid-beta 1-42 in cerebrospinal fluid of Creutzfeldt-Jakob Disease patients. Journal of Alzheimer's Disease 1999;**1**(6):419-24.

Wetterling 1996

Wetterling T, Kanitz RD, Borgis KJ. Comparison of different diagnostic criteria for vascular dementia (ADDTC, DSM-IV, ICD-10, NINDS-AIREN). *Stroke* 1996;**27**(1):30-6.

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529-36.

WHO 1993

World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders. World Health Organization 1993.

WHO 1998

World Health Organization. Global surveillance, diagnosis, and therapy of human transmissible spongiform encephalopathies: report of WHO consultation. World Health Organization 1998.

Wilkosz 2010

Wilkosz PA, Seltman HJ, Devlin B, Weamer EA, Lopez OL, DeKosky ST, et al. Trajectories of cognitive decline in Alzheimer's disease. *International Psychogeriatrics* 2010;**22**:281-290.

Young 2018

Young JJ, Lavakumar M, Tampi D, Balachandran S, Tampi RR. Frontotemporal dementia: latest evidence and clinical implications. *Ther Adv Psychopharmacol* 2018;**8**:33-48.

Zerr 2009

Zerr I, Kallenberg K, Summers DM, et al. Updated clinical diagnostic criteria for sporadic Creutzfeldt-Jakob disease. *Brain* 2009;**132**:2659-2668.

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Abu-Rumeileh 2018

Study characteristics	
Patient Sampling	Retrospective analysis of CSF samples at the Institute of Neurological Sciences of Bologna obtained between 2005 and 2016. Samples were taken from patients with a clinical, genetic, or pathologically confirmed diagnosis of FTD or ADD, and cognitively healthy controls. A sub-sample of 141 FTD patients were selected who did not have co-existing DLB, ADD, prion diease, or vascular dementia. Sampling procedure: not reported. Separate data were available for the performance of biomarkers in distinguishing between ADD from FTD. We did not include data on performance of the index test to discriminate ADD participants from controls. Exclusion criteria: patients with CBS were excluded, as were those with significant cerebrovascular pathology on brain imaging. DLB was excluded clinically. No other exclusion criteria were detailed.
Patient characteristics and setting	The sample considered in the review comprised of 201 participants, 60 ADD and 141 FTD. All participants underwent clinical history, neurological examination, neuropsychological testing, and neuroimaging. In addition, some participants had post-mortem diagnoses and results from molecular genetic testing. Education, gender, and age at the time of lumbar puncture were similar in ADD and FTD. MMSE score was lower in ADD ($p < 0.05$). Sex: 33 males, 27 females for ADD; 75 males and 66 females for FTD Age mean (SD) (y): 67.1±8.7 for ADD; 64.9 ±9.8 for FTD MMSE: 20.7±4.8 for ADD; 25.0±3.7 for FTD Disease duration (y): not reported Education (y): 10.8±4.8 for ADD; 8.9±4.0 for FTD Sources of recruitment: CSF samples submitted for analysis at the Institute of Neurological Sciences of Bologna
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: >482 ng/L; not prespecified; determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [No]

Abu-Rumeileh 2018 *(Continued)*

Target condition and reference standard(s)	Target condition: Alzheimer's disease (differential diagnosis of ADD from FTD) Reference standards: International Working Group 2 (IWG-2) criteria for ADD and CSF biomarker profile. FTD were classified using criteria for the following subtypes: behavioural variant, non-fluent variant of primary progressive aphasia, sematic variant of primary progressive aphasia, amyotrophic lateral sclerosis, corticobasal syndrome, progressive supranuclear palsy and FTD with parkinsonism. FTD was neuropathologically confirmed in four cases, and 22 cases had additional molecular genetic findings which supported the diagnosis. Ten participants with FTD were excluded where the CSF biomarker profile was in-keeping with a diagnosis of ADD. The final clinical diagnosis was confirmed after at least two years of follow-up. The reference standard results were reported using knowledge of the results of index test.
Flow and timing	The final clinical diagnosis was established after 24 months of follow-up. AD vs FTD (n=201) AD=60; bvFTD=53; Sensitivity=89%; Specificity=80% (Table 2, p381) TP=53; FP=11; FN=7; TN=42 (calculated in RevMan5) Missing data: Data were requested from the author on the bvFTD subtype and ADD. The interval between established clinical diagnosis and CSF sample collection was not reported.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Abu-Rumeileh 2018 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Aerts 2011

Study characteristics	
Patient Sampling	A study with retrospective design (retrospective analysis) of data from patients with DLB and ADD. Consecutive patients with clinical diagnosis of DLB, who were referred to either the movement disorder clinic of the Department of Neurology or the memory clinic of the Department of Geriatric Medicine at the Radboud University Nijmegen Medical Centre, and who underwent a lumbar puncture between December 2003 and June 2008, were included. Out of 93 eligible ADD patients from the memory clinic database, an age and gender matched group of 45 ADD patients was randomly drawn. Exclusion criteria: not reported.
Patient characteristics and setting	The sample considered in the review comprised of 68 participants, 45 ADD and 23 DLB. Disease duration, gender, and age at the time of lumbar puncture were similar in AD and DLB. MMSE score was lower in AD ($p < 0.05$). Sex: 34 males and 11 females for ADD; 18 males and 5 females for DLB Age mean (SD) (y): 71.6±9.4 for ADD; 71.6 ±9.4 for DLB Disease duration (months): 33.0 for ADD; 38.8 for DLB Sources of recruitment: memory clinic and movement disorder clinic, the Radboud University Nijmegen Medical Centre, The Netherlands
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed (within 4 weeks).

Aerts 2011 *(Continued)*

	Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: >482 ng/L; not prespecified; determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of ADD from DLB) Reference standards: NINCDS-ADRDA criteria for ADD. Clinical diagnosis of DLB was based on McKeith criteria. Initial clinical diagnosis was established by a multidisciplinary team consisting of a geriatrician, a neurologist a neuropsychologist prior CSF sample. The final clinical diagnosis was reassessed by a single rater after a follow-up period of 12 months or longer. Not reported whether the reference standard results were reported without knowledge of the results of index test.
Flow and timing	The final clinical diagnosis was established (reassessed) 12 months or longer after CSF sampling. AD vs DLB (n=65) AD=44; DLB=21; Sensitivity=62%; Specificity=65% (Table 2, p381) TP=13; FP=15; FN=8; TN=29 (calculated in RevMan5) Missing data: CSF Abeta42 sample was unavailable from 2 DLB and 1 AD participants (Total: 23 DLB and 44 ADD, p379)
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			

Aerts 2011 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Baldeiras 2015

Study characteristics	
Patient Sampling	Participants were recruited at the Dementia clinic, Neurology Department of Coimbra University Hospital. All patients were followed for two years after which the clinical diagnosis was revised.
Patient characteristics and setting	The sample considered in the review comprised of 214 participants, 107 ADD and 107 FTD. Age of onset, gender, and age at the time of lumbar puncture were similar in AD and FTD. MMSE score was lower in AD ($p < 0.005$). Sex: 37 males and 70 females for ADD; 47 males and 60 females for FTD Age mean (SD) (y): 64.4 ±9.5 for ADD; 66.3 ±9.0 for FTD Age of onset (years): 62.0 ± 9.6for ADD; 62.6± 9.0for FTD Sources of recruitment: Dementia Clinic, Neurology Department of Coimbra University Hospital
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 538pg/ml, not prespecified; determined by ROC analysis.

Baldeiras 2015 *(Continued)*

	Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of ADD from FTD) Reference standards: NINCDS-ADRDA criteria and McKhann et al for ADD. Clinical diagnosis of FTD was based on the Lund and Manchester clinical criteria. The reference standard results were reported without knowledge of the results of index test.
Flow and timing	
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			

Baldeiras 2015 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Unclear risk

Bibl 2006

Study characteristics	
Patient Sampling	Prospective investigation of participants with probable AD, probable DLB and non-demented disease controls from initially consecutively referred sample to a laboratory for neurochemical evaluation. Separate data were available for the performance of biomarkers in distinguishing between AD from DLB. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria: not reported. Exclusion criteria were only reported for the control group.
Patient characteristics and setting	The sample considered in the review comprised of 43 participants, 18 AD and 25 DLB. CSF was collected from hospitalised DLB patients from a clinic specialising in the diagnosis and treatment of Parkinson's disease. CSF of AD patients came from a memory clinic. The mean age and the mean MMSE score did not significantly differ between AD and DLB participants. Sex: 5 males and 13 females for AD; 21 males and 4 females for DLB Age mean (SD) (y): 69.7 ± 10.6 for AD; 72.0 ± 7.5 for DLB Disease duration (y): not reported Sources of recruitment: AD patients from the memory clinic, University of Goettingen; DLB patients: inpatients from a Paracelsus-Elena Klinic, Kassel; Germany
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed (within 2 days). Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 475pg/ml, not prespecified; determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [Yes]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from DLB) Reference standards: NINCDS-ADRDA and DSM-IV criteria for AD. Clinical diagnosis of DLB was based on McKeith and DSM-IV criteria. Diagnosis was established by a psychiatrist and a neurologist (blinded to biomarker results) thorough anamnesis, clinical examination, results of neuropsychological assessment, clinical records of the patients and the best clinical judgement.
Flow and timing	The interval between established clinical diagnosis and blood sample collection was not reported.However, it appears that CSF samples were collected short after establishing the clinical diagnosis of AD and DLB. At baseline: 18 AD; 25 DLB

Bibl 2006 *(Continued)*

	Sample included in the analysis: 18 AD; 23 DLB AD vs DLB (n=41) Disease^{+}: 18; Disease^{-}: 23 Sensitivity=50%; Specificity=96% (Calculated in Revman5) TP=9; FP=1; FN=9; TN=22 (calculated in RevMan5) Missing data: CSF Abeta42 sample was unavailable from 2 DLB participants (p1772)
Comparative	
Notes	Author contacted: there is some discrepancy between our findings and findings data reported in the Table 2, p1775. No reply.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard			Low concern

Bibl 2006 *(Continued)*

does not match the question?		
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		Unclear risk

Bibl 2007

Study characteristics	
Patient Sampling	A total of 90 patients (30 ADD; 30 FTLD; 30 non-demented disease controls) were selected on wards and the dementia outpatient clinic of the Universitiy of Goettingen and the dementia outpatient clinic of the Universitiy of Erlangen between 2000 and 2004. Sampling procedure: not reported. Separate data were available for the performance of biomarkers in distinguishing between ADD from FTLD. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria: not reported
Patient characteristics and setting	The sample considered in the review comprised of 60 participants, 30 ADD and 30 FTLD. 30 non-demented disease controls were not included. Diagnosis was established by a psychiatrist and a neurologist (blinded to biomarker results), all highly experienced in clinical differential diagnosis of dementias, on the basis of thorough anamnesis, clinical examination, results of neuropsychological assessment, clinical records of the patients and the best clinical judgement Sex: 13 males and 17 females for ADD; 21 males and 9 females for FTLD Age mean (SD) (y): 65.4 ± 7.3 for ADD; 61.6 ± 11.5 for FTLD. The mean age did not significantly differ between those two groups. MMSE: 19.3 ± 5.4 for ADD; 20.7 ± 8.9 for FTLD (for 26 participants). The mean age did not significantly differ between those two groups. Disease duration (y): not reported Sources of recruitment: mixed setting: the wards and the dementia outpatient clinic of the Univerity of Goettingen; 5 AD patients were recruited from the dementia outpatient clinic of the Universitiy of Erlangen; Germany
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed (within 2 days). Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: not reported; determined by ROC analysis.

Bibl 2007 *(Continued)*

	Were the index test results reported without knowledge of the reference standard? [Yes]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from FTLD) Reference standards: NINCDS-ADRDA and DSM-IV criteria for ADD. Diagnosis for FTLD was established on the McKhann 2001 and Neary 1988 criteria. Clinicians were blinded to biomarker results.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Bibl 2007 *(Continued)*

DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Bousiges 2016

Study characteristics	
Patient Sampling	A total of 151 patients were selected between January 2013 and January 2015. Sampling procedure: Not reported Separate data were available for the performance of biomarkers in distinguishing probable AD and probable DLB as well as mixed diagnosis of ADD and DLB with the other diagnostic groups. In accordance with inclusion criteria in the current review we only included data to differentiate between ADD and DLB with dementia diagnoses.
Patient characteristics and setting	The sample considered in the review comprised of 51 participants, 31 ADD and 20 DLB. Diagnosis was established double-blinded to biomarker results by clinicians and the biologist. Sex: 12 males and 19 females for ADD; 14 males and 6 females for DLB Age mean (SD) (y): 67.2±9.3 for ADD; 68.8±9.7 for DLB. MMSE: 20.2±4.7 for ADD; 21±4.7 for DLB . Disease duration (y): not reported Sources of recruitment: The tertiary memory clinic of Strasbourg University Hospital
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 500ng/L, pre-specified Were the index test results reported without knowledge of the reference standard? Not reported
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from DLB) Reference standards: McKhann's criteria and Duboi's criteria for ADD. Diagnosis for DLB was established on the McKeith's and DSM-V criteria. Clinicians were blinded to biomarker results.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported.

Bousiges 2016 *(Continued)*

Comparative
Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Bousiges 2018

Study characteristics

Bousiges 2018 *(Continued)*

Patient Sampling	Retrospective multicentre study from six French memory research centres undertaking clinical and biological diagnoses of dementia. All centres used the same diagnostic procedures. Patients were selected from a database between January 2010 and December 2015. 1221 patients were included in the study: 95 control subjects, 57 prodromal-DLB, 154 DLB with dementia, 132 prodromal-ADD, and 783 ADD with dementia. Sampling procedure: not reported. Separate data were available for the performance of biomarkers in distinguishing between ADD from DLB. We did not include data on performance of the index test to discriminate AD participants from controls or prodromal syndromes. Exclusion criteria: patients with mixed diagnoses (e.g. ADD and DLB). No other exclusion criteria were detailed.
Patient characteristics and setting	The sample considered in the review comprised of 937 participants, 783 ADD and 154 DLB. 95 non-demented disease controls were not included. All participants underwent physical, neurological, and neuropsychological assessments, laboratory tests, and brain imaging. ADD was diagnosed according to Albert's and Dubois criteria. DLB was diagnosed according to McKeith's and DSM-V criteria. Sex: 333 males and 450 females for ADD; 93 males and 61 females for DLB Age mean (SD) (y): 67.5 ± 9 for ADD; 70.5 ± 10.5 for DLB. Participants with DLB were significantly older than those with ADD. MMSE: 19.0 ± 5.8 for ADD; 19.2 ± 5.5 for DLB. MMSE score did not differ significantly between ADD and DLB. Disease duration (y): not reported Sources of recruitment: six French memory centres undertaking clinical and biological diagnoses of dementia.
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed (within 4 hours). Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: pre-specified threshold <700ng/L, optimal cut-offs also determined by ROC curve analysis (</= 606ng/L). Were the index test results reported without knowledge of the reference standard? [Unlcear]
Target condition and reference standard(s)	Target condition: Alzheimer's disease (differential diagnosis of ADD from DLB) Reference standards: Albert's and Dubois criteria for ADD. Diagnosis for DLB was established on the McKeith and DSM-V criteria. CSF criteria were not used in the diagnosis of ADD but does not state if clinicians were blinded to the biomarker results.
Flow and timing	AD vs FTD (n=937) AD=783; DLB=154; Sensitivity=71%; Specificity=53% (Table 2, p381) TP=556; FP=72; FN=227; TN=81 (calculated in RevMan5) Missing data: None. The interval between established clinical diagnosis and CSF sample collection was not reported.
Comparative	

Bousiges 2018 *(Continued)*

Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Were all patients included in the analysis?	No		

Bousiges 2018 *(Continued)*

Could the patient flow have introduced bias?	High risk

Brettschneider 2006

Study characteristics	
Patient Sampling	248 patients (109 AD, 41 VD,15 FTD, 25 MCI and 58 controls) were recruited from the Memory Clinic of the Department of Neurology, University Hospital of Ulm over 3 years. Sample procedure not reported. Separate data were available for the performance of biomarkers in distinguishing between AD and other types of dementia. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria: not reported.
Patient characteristics and setting	248 participants were included in the study: 109 AD, 41 VD,15 FTD, 25 MCI and 58 controls. Medical history, neurological, neuropsychiatric, neuroradiological and neuropsychological examinations were obtained. Control group: 34 patients presented with tension-type headache and showed no evidence of a structural, hemorrhagic or inflammatory lesion; 24 patients fulfilled the criteria of a major depressive disorder. CSF samples were collected over 3 years. Separate data were extractable for the accuracy of biomarkers in distinguishing AD dementia from i) FTD & VD and ii) non-AD dementia. The sample considered in the review comprised of 165 participants (109 AD, 41 VD,15 FTD). Sex: 39 males and 70 females for AD; 24 males and 17 females for VD; 8 males and 7 females for FTD Age: 71 (43-88) for AD; 75 (47-88) for VD; 68 (43-77) for FTD Disease duration (y): 2 (0.5-10) for AD; 1.75 (0.5-9) for VD; 2 (0.5-4) for FTD Sources of referral: secondary care. Not reported Sources of recruitment: Memory Clinic of the Department of Neurology, University Hospital of Ulm, Germany
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 612ng/L, not pre-specified, cut-offs were derived from ROC analysis. Were the index test results reported without knowledge of the reference standard? Not reported
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (1. differential diagnosis of AD from VD & FTD; 2. differential diagnosis of AD from non-AD dementia) Reference standards: NINCDS-ADRDA criteria Alzheimer's disease dementia Clinical diagnosis of VD was based on NINDS-AIREN criteria, of FTD on Neary 1998 criteria, of MCI on Pettersen 1999, prior the results of the index test.
Flow and timing	The interval between established clinical diagnosis and blood sample collection was not reported.However, it appears that CSF samples were collected short after establishing the clinical diagnosis of the participants included in the study.

Brettschneider 2006 *(Continued)*

Sample included in the analysis: 109 AD; 56 non-AD (41 VD; 15 FTD)

AD vs non-AD (n=165)

Sensitivity=82%; Specificity=46% (Table 3, p294)

TP=89; FP=30; FN=20; TN=26 (calculated in RevMan5)

All recruited participants with diagnosed dementia were included in the analysis.

Comparative
Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Brettschneider 2006 *(Continued)*

DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk

Casoli 2019

Study characteristics	
Patient Sampling	Participants were recruited at the INRCA hospital Neurology Unit, Ancona, Italy. Participants were included where brain atrophy was present as defined by the Pasquier scale (</=2). 95 participants were included: 55 ADD, 21 FTD, and 20 non-demented controls. Sampling procedure: not reported. Separate data were available for the performance of biomarkers in distinguishing between ADD from FTD. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria were: age <60 years, family history of disease, cerebrovascular accidents, anamnesis of delirium, cognitive decline induced by head injury, recently diagnosed or untreated thyroid disease, vitamin B12 or folic acid deficiency, intoxication with drugs or medications, severe depression (pseudodementia), chromosome 21 trisomy (Down syndrome), neurosyphilis, and human immunodeficiency virus dementia.
Patient characteristics and setting	Participants underwent clinical history, neuropsychological and functional assessments, neuroimaging, and laboratory tests. Sex: 23 males and 32 females for ADD; 9 males and 12 females for FTD. Age mean (SD) (y): 77.3 ± 7.1 for ADD; 72.0 ± 5.8 for FTD. Participants with ADD were significantly older than those with FTD. MMSE: 14.5 ± 6.1 for ADD; 19.0 ± 6.2 for FTD. MMSE score was significantly lower in ADD compared to FTD. Disease duration (y): not reported Sources of recruitment: Participants were recruited at the INRCA hospital Neurology Unit, Ancona, Italy.
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed (within 3 hours). Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Fujirebio Inc., Tokyo, Japan. Threshold: not pre-specified, optimal cut-offs were calculated. Were the index test results reported without knowledge of the reference standard? [Yes]

Casoli 2019 *(Continued)*

Target condition and reference standard(s)	Target condition: Alzheimer's disease (differential diagnosis of ADD from FTD) Reference standards: NIA/AA and IWG-2 criteria for ADD. FTD was diagnosed according to the EFNS-ENS Guidelines. Participants with FTD were subclassifed according to criteria for behavioural variant and primary progressive aphasia subtypes. Diagnosis was confirmed after at least 24 months of follow-up. It was not clear if clinicians were blinded to the results of the index test.
Flow and timing	Data were provided by the author upon request. AD vs FTD (n=76) AD=55; FTD=21; Sensitivity=100%; Specificity=0% (Table 2, p381) TP=55; FP=21; FN=0; TN=0 (calculated in RevMan5) Missing data: None. The interval between established clinical diagnosis and CSF sample collection was not reported.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

Casoli 2019 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

de Jong 2006

Study characteristics	
Patient Sampling	Patients with mild to moderate AD (n=61) or VD (n=25) were selected from a large database containing 260 patients with cognitive impairment or dementia of various origins (e.g., degenerative, vascular, hereditary, inflammatory, metabolic) who visited an outpatient clinic between 1992 and 2004. Thirty controls, aged >50 years, with no neurological disorder, were also included. We only considered data on performance of the index test to discriminate between patients with AD and VD. Excluded criteria: not reported
Patient characteristics and setting	The sample considered in the review comprised of 86 participants, 61 AD and 25 VD. Separate data were reported for the performance of biomarkers to distinguish between AD and VD. The control group was not included. The mean age did not significantly differ between AD and VD participants. Sex: 25 males, 36 females for AD; 14 males, 11 females for VD Age (SD) (y): 68 (8.8) for AD; 72 (8.4) for VD Sources of recruitment: database of patients from an outpatient clinic, the Radboud University Nijmegen Medical Centre, The Netherlands
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 520pg/mL, not pre-specified, determined by ROC analysis. Cutoff values with the most optimal combination of sensitivity and specificity to discriminate between these AD and VD groups were calculated.

de Jong 2006 *(Continued)*

	Were the index test results reported without knowledge of the reference standard? Not reported
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from VD) Reference standards: NINCDS-ADRDA criteria for AD. Clinical diagnosis of VD was based on NINDS-AIREN criteria (Roman 1993). Clinical diagnosis was established prior to study entry.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported.However, it appears that CSF samples were collected shortly after establishing the clinical differential diagnosis of AD and VD. Lumbar punctures were performed after written informed consent was obtained from the patient and the patient's legal representatives. Sample included in the analysis: 61 AD; 25 VD AD vs VD (n=86) Sensitivity=82%; Specificity=76% (Table 2, p756) TP=50; FP=6; FN=11; TN=19 (calculated in RevMan5)
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

de Jong 2006 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

de Rino 2012

Study characteristics	
Patient Sampling	The enrolment of patients in this prospective study started in January 2006 and ended in December 2009. All consecutive patients admitted to two tertiary memory clinics with an ambiguous diagnosis of AD or fvFTD according to current research criteria (Neary 1998; McKhann 1984) underwent lumbar puncture as a diagnostic tool. 75 ADD patients and 42 fvFTD patients were enrolled. Exclusion criteria: not reported.
Patient characteristics and setting	The sample considered in the review comprised of 114 participants, 72 ADD and 42 fvFTD. MMSE adjusted score was significantly higher (p = 0.04) in fvFTD than in ADD. Sex: 32 males and 40 females for ADD; 26 males and 16 females for fvFTD Age mean (SD) (y): 67±6.8 for ADD; 69±7.1 for fvFTD Disease duration (months): 24.1±12.6 for ADD; 26.9±15.0 for fvFTD MMSE: 18.3±4.2 for ADD; 25.5±4.8 for fvFTD Sources of referral: not reported Sources of recruitment: two tertiary memory clinics, Department of Neurology, IRCCS MUltimedica and Vita-Salute S. Raffaele University, Milan, Italy
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed (within 15 days). Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 104pg/mL, not pre-specified, determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? Yes

de Rino 2012 *(Continued)*

Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of ADD from bvFTD) Reference standards: NINCDS-ADRDA criteria for ADD. Clinical diagnosis of FTD was based on Neary 1998 criteria. Initial clinical diagnosis, independent of CSF metabolite levels, were established, which during the study were known only to researchers not further involved in the follow-up. Afterwards, patients were evaluated at 6-months intervals by three expert neurologists blind to the CSF results, who had to confirm or discard the initial clinical diagnosis. After at least 2 years of follow up, the last clinical diagnosis was considered as the gold standard to be compared with CSF biomarkers.
Flow and timing	The final clinical diagnosis was established (reassessed) at least 2 years of follow up after CSF sampling. ADD vs bvFTD (n=114) ADD=72; bvFTD=42 Sensitivity=82%; Specificity=21% (calculated in RevMan5) TP=59; FP=33; FN13; TN=9 (calculated in RevMan5)
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowl-	Yes		

de Rino 2012 *(Continued)*

edge of the results of the index tests?			
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Falgas 2020

Study characteristics	
Patient Sampling	A cross-sectional study of participants under the age of 65 undergoing assessment at the Alzheimer's Disease and Other Cognitive Disorders Unit at the Hospital Clinic de Barcelona. 138 participants were recruited between 2009 and 2016 with the following diagnoses: 64 ADD, 26 FTD, and 48 healthy controls. Sampling procedure: not reported. Separate data were available for the performance of biomarkers in distinguishing between ADD from FTD. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria: not detailed.
Patient characteristics and setting	Participants underwent neurological and neuropsychological assessments, and neuroimaging. Sex: 28 males and 36 females for ADD; 14 males and 12 females for FTD. Age mean (SD) (y): 56.6 (54.5-60.5) for ADD; 60.6 (55.9-64.7) for FTD. Participants with FTD were significantly older than those with ADD. MMSE: 23 (19-26.5) for ADD; 26.0 (24.0-27.0) for FTD. MMSE score was not significantly different in ADD compared to FTD. Disease duration (y): 2.9 (1.61-3.79) for ADD; 2.88 (1.9-3.78) for FTD. Sources of recruitment: Participants were recruited at the Alzheimer's Disease and Other Cognitive Disorders Unit at the Hospital Clinic de Barcelona.
Index tests	Patients gave CSF samples.

Falgas 2020 *(Continued)*

	Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: pre-specified at <550 pg/ml and 750 pg/ml, but optimal thresholds were used for analysis. Were the index test results reported without knowledge of the reference standard? [Unclear].
Target condition and reference standard(s)	Target condition: Alzheimer's disease (differential diagnosis of ADD from FTD). Reference standards: NIA/AA for ADD: NIA/AA criteria. All participants with ADD had a typical CSF biomarker pattern. FTD was diagnosed by criteria in two subtpyes: behavioural variant and semantic variant of primary progressive aphasia. It was not clear if clinicians were blinded to the results of the index test.
Flow and timing	Data were provided by the author upon request. AD vs FTD (n=23) AD=18; FTD=5; Sensitivity=100%; Specificity=94% (Table 2, p381) TP=5; FP=1; FN=0; TN=17 (calculated in RevMan5) Missing data: None. The interval between established clinical diagnosis and CSF sample collection was not reported.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 3: Reference Standard			

Falgas 2020 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Herbert 2014

Study characteristics	
Patient Sampling	Patients who were referred to either the movement disorders clinic of the department of Neurology or the memory clinic of the Department of Geriatric Medicine at the Radboud University Medical Centre during the period May 1996 to December 2009. Patients who had received a lumbar puncture and had relevant CSF parameters and not included in a previous study were included.
Patient characteristics and setting	The sample considered in the review comprised of, 64 ADD, 14 DLB, 15 VaD and 26 FTD subjects. MMSE findings and disease duration were not available for all patients. Sex: 13 males and 51 females for ADD; 10 males and 4 females for DLB, 10 males and 5 females for VaD and 17 males and 9 females for FTD. Age mean (SD) (y): 73.1 ± 8.3 for ADD; 72.4 ± 8.0 for DLB, 76.5 ± 4.8 for VaD and 61.6 ± 8.4 for FTD. Disease duration (months): 15 ± 15.6 or ADD (n=61); 24 ± 24.0 for DLB (n=6), 17 ± 15 for VaD (n= 12) and 7.3 ± 14 for FTD (n= 12). MMSE: 20 ± 4 for ADD (n= 61); 22 ± 5 for DLB (n=4), 18 ± 3.7 for VaD (n=12) and 18 ± 7.3 for FTD (n= 10).
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed.

Herbert 2014 *(Continued)*

	Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 500pg/mL, not pre-specified, determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? Yes
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD dementia from DLB, VaD and FTD) Reference standard: NINCDS-ADRDA criteria for AD The clinical diagnosis of DLB was based on McKeith criteria, for VaD on NINDS-AIREN criteria and for FTD on the Lund and Manchester Groups criteria. It is not stated whether the reference standard was performed before applying the index test.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported. However, it appears that CSF samples were collected short after the diagnosis of dementia was confirmed. Sample included in the analysis: 64 AD; 14 DLB; 26 FTD; 15 VaD
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

Herbert 2014 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		Unclear risk

Kapaki 2001

Study characteristics	
Patient Sampling	A total of 99 subjects were included in the study: 38 patients with AD, 14 patients with CJD and 47 controls. Sampling procedure not reported. We only considered data on performance of the index test to discriminate between patients with AD and CJD. Exclusion criteria not reported.
Patient characteristics and setting	The sample considered in the review comprised of 52 participants: 38 patients with ADD, 14 patients with CJD. Sex: 15 M, 23 F AD; 7 M, 7F CJD Age (y): 68±10 years AD; 59±4 CJD Disease duration (y): 3.6±2.4 AD; 0.4±0.2 CJD Sources of recruitment: Department of Neurology, Athens National University, Greece. Not reported whether inpatients or outpatients
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 445pg/ml; not prespecified; Receiver operating characteristics (ROCs) curve analysis was used to define the cut off concentrations of tau protein and $A\beta_{42}$ with the corresponding optimal sensitivity and specificity (Fig 1B, p402). Were the index test results reported without knowledge of the reference standard? [Not reported]

Kapaki 2001 *(Continued)*

Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD dementia from CJD) Reference standard: NINCDS-ADRDA criteria for AD The clinical diagnosis of CJD was based on progressive dementia of less than 2 years, periodic sharp wave complexes in the EEG recording, and two of the following: (1) myoclonus, (2) visual or cerebeller symptoms, (3) pyramidal or extrapyramidal tract signs, and (4) akinetic mutism. All patients had a positive test for 14-3-3 protein, a sensitive marker of the disease. The reference standard was performed before applying the index test. Method of confirming diagnosis was not specified for two patients.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported. However, it appears that CSF samples were collected short after the diagnosis of dementia was confirmed. Sample included in the analysis: 38 AD; 12 CJD AD vs CJD (50) TP=29; FP=7; FN=9; TN=5 (Fig 1B, p402) Sensitivity=76%; Specificity=42% (Calculated in RevMan5) Missing data: from 2 CJD participants. it was not reported whether those two patients with clinical diagnosis of CJD, which were not confirmed either postmortem or by biopsy, were excluded from the analysis.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			

Kapaki 2001 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		Unclear risk	

Kapaki 2003

Study characteristics	
Patient Sampling	Participants from an outpatient clinic diagnosed with AD and non-AD dementia were followed-up for at least three years in an effort to ensure the correct diagnosis, and doubtful cases were rejected. 70 patients with dementia (49 AD; 15 non-AD; 6 VD) were recruited. 49 controls were also included. Sample procedure not reported. Separate data were available for the performance of the biomarkers in distinguishing AD from non-AD dementia, and AD from VD. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria: patients with dementia due to metabolic causes and patients with a history of alcohol abuse, MRI infarctions (except VD patients), or B12 deficiency were excluded.
Patient characteristics and setting	The sample considered in the review comprised of 70 participants: 49 AD, 15 non-AD (6 DLB; 4 FTD; 1 with Parkinson's disease; 2 with progressive supranuclear pulsy; 2 with corticobasal-ganglionic degeneration) and 6 with VD. All participants had detailed evaluation (medical history, physical and neurological examination, blood tests to exclude metabolic causes of dementia) and MRI. Sex: 31 males and 18 females for AD; 11 males and 4 females for non-AD dementia; 4 males and 2 females for VD Age (SD) (y): 67.6 ± 9.3 for AD; 61.3 ± 5.1 for non-AD dementia; 69 ± 4 for VD Sources of recruitment: an outpatient clinic, Athens National University, Greece.

Kapaki 2003 *(Continued)*

Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -70°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 435 pg/ml; not prespecified; Cut-offs were determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (1. differential diagnosis of AD from non-AD dementia; 2. differential diagnosis of AD from VD) Reference standards: NINCDS-ADRDA for AD. Clinical diagnosis of VD was based on NINDS-AIREN criteria, of DLB and Parkinson's dementia on McKeith criteria, of FTD on Neary 1999 criteria, of progressive supranuclear palsy according on NINDS-SPSP criteria. Criteria of corticobasal-ganglionic degeneration were not not specified. Clinical diagnosis was established prior the results of the index test.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported. However, it appears that CSF samples were collected shortly after the clinical diagnosis was established. Sample included in the analysis: 49 AD; 6 VD; 15 non-AD (6 DLB; 4 FTD; 1 with PD dementia; 2 with progressive supranuclear pulsy; 2 with corticobasal-ganglionic degeneration) 1. AD vs non-AD dementia (n=64) Sensitivity=71%; Specificity=80% (Abstract) TP=35; FP=3; FN=14; TN=12 (calculated in Revman5) 2. AD vs VD (n=55) Sensitivity=82%; Specificity=67% (Abstract) TP=40; FP=2; FN=9; TN=4 (calculated in Revman5)
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		

Kapaki 2003 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Kapaki 2005

Study characteristics	
Patient Sampling	A total of 103 subjects were included in the study: 33 patients with AD, 20 patients with ARCD and 50 controls (healthy elderly). ARCD patients were recruited during a two-year period from a larger pool of 82 detoxified alcoholic subjects. No further details about sampling procedure. Separate data were available for the performance of biomarkers in distinguishing between AD from ACRD. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria not reported.

Kapaki 2005 *(Continued)*

Patient characteristics and setting	The sample considered in the review comprised of 53 participants: were included in the review: 33 with AD and 20 with ARCD, which completed a detoxification program. AD patients were subjected to a detailed evaluation (medical history, physical and neurological examination, computed tomography and/or magnetic resonance imaging and blood tests to exclude metabolic causes of dementia). There was no history of alcohol use or abuse and all had a sufficient follow-up (for at least two years) to ensure diagnosis. No one of the patients was under any medication for dementia at the time of lumbar puncture. Evaluation of alcohol abuse was made by the Pattern of Abuse tool (Hughes 1980), the section on alcoholism of Composite International Diagnostic Interview (WHO 1990) and the Diagnostic Interview Schedule (Wells 1994). The mean duration of alcohol consumption was 29 years (range 6–40 years). Only 23 of the 83 subjects met the DSM-IV criteria of alcohol-induced persisting dementia. Three out of the 23 patients were under the age of 40 years (out of the range of AD patients), and were not included in the study. Sex: 14 M, 19 F AD; 18 M, 2 F ACRD Age: 63±11 years AD; 60±12 ACRD MMSE: 23 (15-27) AD; 25 (15-28) ACRD Resources of recruitment: i) in-patients: Drug and Alcohol Addiction Clinic, Department of Psychiatry, Athens National University, Greece; ii) not reported for AD participants
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -70°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 562 pg/ml; not prespecified; Cut-offs were determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD dementia from ARCD) Reference standard: NINCDS-ADRDA criteria Clinical diagnostic criteria for ARCD: the Pattern of Abuse tool (Hughes 980), the section on alcoholism of Composite International Diagnostic Interview (WHO 1990) and the Diagnostic Interview Schedule (Wells 1994). The reference standard was performed before applying the index test.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported.However, it appears that CSF samples were collected short after neuropsychological examination that was performed two months after detoxification for alcohol-induced dementia. Sample included in the analysis: 33 AD; 20 ARCD AD vs ACRD (53) TP=28; FP=4; FN=5; TN=16 (Fig 1B, p402) Sensitivity=85%; Specificity=80% (Abstract)
Comparative	
Notes	
Methodological quality	

Kapaki 2005 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Kapaki 2007

Study characteristics	
Patient Sampling	A total of 85 patients and 72 elderly controls were recruited. Sample procedure not described. Separate data were available on biomarkers for differentiating AD and idiopathic normal presure hydrocephalus (iNPH) patients. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria: patients with secondary NPH (e.g. following meningitis, hemorrhage, brain tumor or trauma) were excluded
Patient characteristics and setting	The sample considered in the review comprised of 85 participants: 67 with AD and 18 with iNPH. All the patients underwent extensive neuropsychological evaluation in an effort to further reduce the possibility of AD comorbidity. At least a 2-year follow-up was required to ensure correct diagnosis. No AD patients were under cholinesterase inhibitor therapy at the time of lumbar puncture Sex: 26 males and 41 females for AD; 11 males and 7 females for AD for iNPH Age (SD) (y): 66 ± 10 for AD; 69 ± 14 for iNPH MMSE:18 (14–22) for AD; 21 (16–26) for iNPH Disease duration (y): 3.2 ± 2.3 for AD; 0.7 ± 0.4 for iNPH Sources of recruitment: specialist care setting, Athens National University, Greece. Not reported whether inpatients or outpatients
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 268 pg/ml; not prespecified; Cut-offs were determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD dementia from iNPH) Reference standard: NINCDS-ADRDA criteria for AD Clinical diagnostic criteria for iNPH: the standard classic triad of gait impairment, urinary incontinence and impaired mental function, supported by ventricular dilation in neuroimaging without significant cerebral atrophy, with Evan's index >0.3 on CT or MRI scan. The reference standard was performed before applying the index test.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported. However, it appears that CSF samples were collected short after establishing the clinical diagnosis of AD and iNPH. Sample included in the analysis: 67 AD; 18 iNPH AD vs iNPH (n=85) Sensitivity=91%; Specificity=44% (Table 2, p171)

Kapaki 2007 *(Continued)*

	TP=61 FP=10; FN=6; TN=8 (calculated in RevMan5)		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Were all patients included in the analysis?	Yes		

Kapaki 2007 *(Continued)*

Could the patient flow have introduced bias?	Low risk

Kapaki 2008

Study characteristics	
Patient Sampling	A total of 203 participants (76 AD; 34 FTLD; 93 healthy controls) were prospectively enrolled in the study. No further information on sampling procedure. Separate data were available for the performance of biomarkers in distinguishing between AD from FTD and AD from FTLD. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria: secondary causes of dementia.
Patient characteristics and setting	110 participants were considered in the review: 76 AD and 34 FTLD (24 FTD; 5 PPA; 5 FTD with motor neuron signs). All patients underwent detailed clinical, neuropsychologic, biochemical, and neuroimaging examination (magnetic resonance imaging in all patients and, additionally, single photon emission computed tomography in all FTLD patients), to exclude secondary causes of dementia and establish the diagnosis. In addition, at least 2-years follow-up was available to ensure the correct diagnosis. None of the patients were under cholinesterase inhibitors at the time of lumbar puncture. Sex: 28 males and 48 females for AD; 20 males and 14 females for FTLD Age mean (SD) (y): 66.0 ± 10.0 for AD; 3.1 ± 2.7 for FTLD Disease duration (y): 3.4 ± 2.8 for AD; 61.0 ± 9.0 for FTLD Sources of referral: not reported Sources of recruitment: specialist care setting, Athens National University, Greece. Not reported whether inpatients or outpatients
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 451 pg/ml; not prespecified; Cut-offs were determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD dementia from FTD, and AD from FTLD) Reference standard: NINCDS-ADRDA criteria for AD The clinical diagnosis of FTLD was established on Neary 1998 criteria. At least 2-years follow-up was available to ensure the correct diagnosis, prior the results of the index test. Disease duration was defined as the time between the onset of the symptom(s) and CSF sampling.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported. However, it appears that CSF samples were collected shortly after establishing the clinical diagnoses. Sample included in the analysis: 76 AD and 34 FTD (FTLD: 24 FTD; 5 PPA; 5 FTD with motor neuron signs) AD vs FTD (FTLD) (N=107)

Kapaki 2008 *(Continued)*

	TP=57; FP=9; FN=19; TN=22 (Fig 1b, p49) Sensitivity=75%; Specificity=71% (Calculated in RevMan) Missing data: 3 FTLD were not included in the analysis
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Kapaki 2008 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Yes	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		Unclear risk

Khoonsari 2019

Study characteristics	
Patient Sampling	Analysis of CSF samples from 76 ADD, 74 MCI, 11 FTD, and 45 non-dementia controls. Participants with MCI were followed-up for 4-8 years and 21 converted to AD, 53 remained stable. Recruitment procedure: not specified. Sampling procedure: not specified. Separate data were available for the performance of biomarkers in distinguishing between ADD from FTD. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria: not detailed.
Patient characteristics and setting	Participants underwent clinical history, cognitive assessment, and neuroimaging. Sex: 29 males and 47 females for ADD; 7 males and 4 females for FTD. Age median (range) (y): 72 (54-88) for ADD; 66 (50-75) for FTD. Participants with ADD were significantly older than those with FTD. MMSE: 23.6 ± 4.3 for ADD; 25.20 ± 4.2 for FTD. MMSE score was significantly lower in ADD compared to FTD. Disease duration (y): not specified. Sources of recruitment: not specified.
Index tests	Patients gave CSF samples. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: pre-specified at <530 ng/L. Were the index test results reported without knowledge of the reference standard? [Unclear].
Target condition and reference standard(s)	Target condition: Alzheimer's disease (differential diagnosis ADD from FTD) Reference standards: NINCDS-ADRDA and DSM-IV criteria for ADD. FTD diagnostic criteria not stated. It was not clear if clinicians were blinded to the results of the index test.
Flow and timing	Data were provided by the author upon request.

Khoonsari 2019 *(Continued)*

	AD vs FTD (n=87) AD=76; FTD=11; Sensitivity=88%; Specificity=91% (Table 2, p381) TP=67; FP=1; FN=9; TN=10 (calculated in RevMan5) Missing data: None. The interval between established clinical diagnosis and CSF sample collection was not reported.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Were all patients included in the analysis?	No		

Khoonsari 2019 *(Continued)*

Could the patient flow have introduced bias?	High risk

Knapskog 2018

Study characteristics	
Patient Sampling	A cross-sectional study at the memory clinic at Oslo University Hospital, Ullevaal, Norway. 205 patients were referred for diagnostic work-up between January 2009 and July 2014. 138 participants had a diagnosis of ADD, and 17 were "other dementia". Separate data were available for the performance of biomarkers in distinguishing between ADD from FTD. We did not include data on performance of the index test to discriminate AD participants from MCI or subjective cognitive impairment. Sampling procedure: not reported. Inclusion criteria: CSF biomarkers available. Exclusion criteria: none.
Patient characteristics and setting	Participants underwent clinical history, neuropsychological examination, laboratory tests, neuroimaging. Consensus diagnosis was made by two experienced physicians. Sex: 46.3% of the total sample were female. Age mean (SD): 84.8 ±8.8 for the total sample. MMSE: 23.5 ± 4.1 for ADD; 24.3 ± 3.6 for other dementia. MMSE score was significantly lower in ADD compared to FTD. Disease duration (y): not specified. Sources of recruitment: outpatient memory clinic at the Oslo University Hospital, Ullevall, Norway.
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -20°C and analysed (within 1 day). Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: pre-specified at >550 ng/L and >700 ng/L. Were the index test results reported without knowledge of the reference standard? [Unclear]
Target condition and reference standard(s)	Target condition: Alzheimer's disease (differential diagnosis ADD from other dementia) Reference standards: no diagnostic criteria specified; by consensus between two experiences physicians. Physicians were blinded to the results of the index test.
Flow and timing	Data were provided by the author upon request. AD vs FTD (n=71) AD=59; FTD=12; Sensitivity=43%; Specificity=35% (Table 2, p381) TP=25; FP=8; FN=34; TN=4 (calculated in RevMan5)

Knapskog 2018 *(Continued)*

	Missing data: Yes. The interval between established clinical diagnosis and CSF sample collection was not reported.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Were all patients included in the analysis?	Yes		

Knapskog 2018 *(Continued)*

Could the patient flow have introduced bias?	Unclear risk

Lewczuk 2004

Study characteristics	
Patient Sampling	In total 68 participants were recruited (22 AD; 11 non-AD; 35 controls). Sampling procedure not reported. Separate data were available on the performance of biomarkers to distinguish between ADD and other types of dementia. We did not include data on performance of the index test to discriminate AD participants from controls. No details of recruitment, or exclusion criteria were reported.
Patient characteristics and setting	The sample considered in the review comprised of 33 participants: 22 AD and 11 non-AD dementia (5 VD; 1 mixed dementia; 1 subcortical arterial sclerotic encephalopathy; 1 senile dementia of vascular origin; 1 FTD accompanied by Still-Richardson-Olszewski syndrome; 1 dementia due to alcohol abuse; 1 dementia of unclear etiology). All subjects underwent clinical examination, routine blood, urine and CSF tests, magnetic resonance imaging or computed tomography and neuropsychological tests when applicable. Sex: 6 males and 16 females for AD; 6 males and 5 females for non-AD dementia Age (SD) (y): 68 (62–77) for AD; 75 (65–80) for non-AD dementia MMSE: 14 (12–19) for AD; 22 (21–25) for non-AD dementia Sources of recruitment: specialist care setting, University of Goetting, Germany. Not reported whether inpatients or outpatients.
Index tests	Patients gave CSF samples. The samples were stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 550 pg/ml; not prespecified; Cut-offs were determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [Yes]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from non-AD dementia) Reference standards: NINCDS-ADRDA for AD. Clinical diagnosis was established prior the results of the index test.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported.However, it appears that CSF samples were collected short after establishing the clinical diagnosis. Sample included in the analysis: 21 AD; 11 non-AD (5 VD; 1 mixed dementia; 1 subcortical arterial sclerotic encephalopathy; 1 SD; 1 FTD; 1 dementia due to alcohol abuse; 1 unspecified) AD vs non-AD (n=33) Sensitivity=86%; Specificity=82% (Table 2, p275)

Lewczuk 2004 *(Continued)*

	TP=19; FP=2; FN=3; TN=9 (calculated in RevMan5)		
Comparative			
Notes			

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Were all patients included in the analysis?	Yes		

Lewczuk 2004 *(Continued)*

Could the patient flow have introduced bias?	Low risk

Lins 2004

Study characteristics	
Patient Sampling	CSF samples archived for research purposes from patients with probable AD, VD, iNHP dementia, Parkinson disease without dementia and controls were selected. Separate data on the performance of biomarkers to distinguish between AD from VD and iNPH dementia have been reported. Sample procedure not reported. Exclusion criteria not reported.
Patient characteristics and setting	CSF samples from 36 participants: 12 ADD, 12 VD and 12 iNPH. Sex: 5 males and 7 females for AD; 4 males and 8 females for VD; 9 males and 3 females for iNPH Age (SD) (y): 71.8 ±1.7 AD; 76.4 ±1.9 for VD; 75.0 ±1.9 for iNPH Sources of recruitment: not reported. Not reported whether the study was conducted in Germany or Austria.
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 562 pg/ml; not prespecified; Cut-offs were determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (1. differential diagnosis of AD from VD; 2. differential diagnosis of AD from iNPH) Reference standards: NINCDS-ADRDA criteria for ADD. Clinical diagnosis of VD was based on NINDS-AIREN and ICD-10 criteria. Clinical diagnosis of iNPH was based on clinical symptoms (Keifer index), the results of neuroimaging and improvement after CSF withdrawal. Clinical diagnosis was established prior the results of the index test.
Flow and timing	Retrospective analysis. The interval between established clinical diagnosis and CSF sample collection was not reported. Sample included in the analysis: 12 AD, 12 VD; 12 iNPH. AD vs VD (n=24) TP=8; FP=6; FN=4; TN=6 (Fig 1, p277) Sensitivity=67%; Specificity=50% (Calculated in RevMan5) AD vs iNPH (n=24)

Lins 2004 *(Continued)*

	TP=8; FP=8; FN=4; TN=4 (Fig 1, p277) Sensitivity=67%; Specificity=33% (Calculated in RevMan5)
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Lombardi 2018

Study characteristics	
Patient Sampling	A single-centre retrospective observational study. 45 consecutive patients with an atypical presentation were recruited between 2014 and 2015. Patients were included where the diagnosis was uncertain after clinical evaluation, and who had CSF biomarkers available. Final diagnoses were: 32 ADD, 10 FTD, and 3 unclassified cognitive decline (UCD). Sampling procedure: not reported. Exclusion criteria: high vascular burden, prevailing extrapyramidal signs, or pathogenic mutations.
Patient characteristics and setting	Cases were selected by an expert neurologist who administered the diagnosis after at least one year of follow-up. Two further neurologists who were blinded to the final diagnosis, determined the diagnosis in three different scenarios: clinical information only (neuropsychological assessment and neuroimaging), pathological information (amyloid-PET imaging and/or CSF biomarkers), and FDG-PET (brain metabolism). All participants underwent neuropsychological testing and brain imaging. Sex: 19 male, 13 female for ADD; 5 male, 5 female for FTD; 0 male, 3 female for UCD. Age mean (SD): 66.5 ± 9.9 for ADD; 67.4 ± 8.5 for FTD; 59.3 ± 11.9 for UCD. MMSE: 21.7 ± 4.3 for ADD; 22.6 ± 2.4 for FTD; 23 ± 3.5 for UCD. MMSE score was not significantly different in ADD compared to FTD. Disease duration (y): not reported. Sources of recruitment: retrospective, observational study.
Index tests	Patients gave CSF samples. The samples were collected at 8am, immediately centrifuged, and stored at -80°C and analysed (within 1 day). Abeta42 was measured using enzyme-linked immunosorbent assays, (kit not specified). Threshold: pre-specified at >650 pg/ml. Were the index test results reported without knowledge of the reference standard? [Unclear]
Target condition and reference standard(s)	Target condition: Alzheimer's disease (differential diagnosis of ADD from FTD or ADD from UCD). Reference standard: NIA-AA criteria for ADD. FTD was diagnosed according to Gorno-Tempini Rascovsky criteria. The final diagnosis was not blinded to the results of the index test.
Flow and timing	Data were provided by the author upon request. AD vs FTD (n=42) AD=32; FTD=10; Sensitivity=87%; Specificity=70% (Table 2, p381) TP=28; FP=3; FN=4; TN=7 (calculated in RevMan5) AD vs UCD (n=35) AD=32; UCD=3; Sensitivity=87%; Specificity=64% (Table 2, p381) TP=28; FP=1; FN=4; TN=2 (calculated in RevMan5)

Lombardi 2018 *(Continued)*

	Missing data: No. The interval between established clinical diagnosis and CSF sample collection was not reported.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Lombardi 2018 *(Continued)*

Were all patients included in the analysis?	Unclear
Could the patient flow have introduced bias?	Unclear risk

Maddalena 2003

Study characteristics	
Patient Sampling	Prospective study recruiting 100 consecutive dementia patients through a memory disorders clinic. 31 controls were also recruited among cognitively intact patients and added to the sample. Separate data were available on the performance of biomarkers to distinguish between AD and non-AD dementia. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria not reported. Referral through health services such as GP, community health etc. 31 controls were included. No exclusion criteria were specified.
Patient characteristics and setting	The sample considered in the review comprised of 81 participants, 51 AD and 30 non-AD dementia (8 VD; 2 cerebral amyloid angiopathy; 2 DLB; 3 FTLD; 4 Parkinson's dementia; 1 progressive supranuclear palsy; 2 corticobasal degeneration; 3 CJD; 2 Huntington disease; 2 cerebral autosomal dominant arteriopathy with subcortical infarctions and leukoencephalopathy; 1 neuroacanthocytosis). Ninteen participants with other neurological disorders and thirty one controls were not considered in this review. Patients underwent thorough clinical examination, including providing medical and family history; neurological, internal, and psychiatric examinations; routine laboratory testing; and CT or MRI of brain. Sex: 54 males and 46 females (total cohort) Age (SD) (y): 70.1±8.7 (range=51-87) for AD; 66.3±11.2 (range=40-90) for non-AD dementia MMSE: 21.3±5.3 for AD; 21.1±5.7 for non-AD dementia Sources of referral: GP, community health services, specialists in neurology, psychiatry or geriatrics. Sources of recruitment: memory disorders unit, outpatients, University of Zurich, Switzerland
Index tests	Patients gave CSF samples. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 490 pg/ml; not prespecified; Cut-offs were determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from non-AD dementia) Reference standards: NINCDS-ADRDA for AD. Clinical diagnosis of DLB was based on McKeith criteria, of VD on NINDS-AIREN criteria, of FTD on The Lund and Manchester Group criteria. Clinical diagnosis was established prior the results of the index test.

Maddalena 2003 *(Continued)*

Flow and timing	Lumbar puncture was performed and CSF samples were obtained within one week of neuropsychological testing. Sample included in the analysis: 51 AD and 30 non-AD dementia (8 VD; 3 FTD; 2 DLB; 2 PDD; 2 CJD; 2 cerebral amyloid angiopathy; 11 other) AD vs non-AD (n=81) Sensitivity=78%; Specificity=70% (Table, p1205) TP=40; FP=9; FN=11; TN=21 (Calculated in RevMan5)
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard			Low concern

Maddalena 2003 *(Continued)*

does not match the question?		
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk

Marchegiani 2019

Study characteristics	
Patient Sampling	Consecutive patients who were admitted to the Neurology Unit of the Geriatric Hospital of Ancona, Italy between July 2010 and July 2017. Participants with CSF sample available were included in the study. 153 participants were included: 70 ADD, 23 tauopathy (19 FTD, 3 progressive supranuclear palsy, 3 corticobasal syndrome), 17 vascular dementia, and 43 cognitively healthy participants. Separate data were available for the performance of biomarkers in distinguishing between ADD from FTD or vascular dementia. We did not include data on performance of the index test to discriminate AD participants from cognitively healthy participants. Sampling procedure: consecutive patients with CSF samples. Exclusion criteria: patients with unidentified neurodegenerative disease or patients with different various diagnoses (e.g. psychiatric disorders, traumatic brain injury, alcoholism, metabolic encephalopathy).
Patient characteristics and setting	All the participants underwent physical, neurological and neuropsychological assessments, including laboratory tests, brain imaging and the MMSE evaluation. Sex: 26 male, 44 female for ADD; 12 male, 11 female for FTD; 8 male, 9 female for vascular dementia. Age mean (SD): 77 ± 7.7 for ADD; 68.6 ± 8.3 for tauopathy; 79.4 ± 6.2 for vascular dementia. MMSE: 14.9 ± 6.3 for ADD; 18.2 ± 7.7 for tauopathy; 20.3 ± 7.8 for vascular dementia. MMSE score was not significantly different in ADD compared to tauopathy or vascular dementia. Disease duration (y): not reported. Sources of recruitment: Neurology Unit at the Geriatric Hospital of Ancona, Italy.
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Fujirebio Inc., Japan. Threshold: pre-specified at <500 pg/ml. Were the index test results reported without knowledge of the reference standard? [Yes]

Marchegiani 2019 *(Continued)*

Target condition and reference standard(s)	Target condition: Alzheimer's disease (differential diagnosis of ADD from tauopathy or ADD from vascular dementia). Reference standard: NINCDS-ADRDA or NIA/AA criteria for ADD. FTD was diagnosed according to the Neary or Rascovky criteria, and vascular dementia according to the NINDS-AIREN criteria. It was unclear if the reference standard was blinded to the results of the index test.
Flow and timing	Data were provided by the author upon request. AD vs tauopathy (n=93) AD=70; tauopathy=23; Sensitivity=96%; Specificity=57% (Table 2, p381) TP=67; FP=10; FN=3; TN=13 (calculated in RevMan5) AD vs vascular dementia (n=87) AD=70; vascular dementia=17; Sensitivity=65%; Specificity=94% (Table 2, p381) TP=45; FP=1; FN=25; TN=16 (calculated in RevMan5) Missing data: No. The interval between established clinical diagnosis and CSF sample collection was not reported.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Marchegiani 2019 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Montine 2001

Study characteristics	
Patient Sampling	Participants with probable AD and dementias other than AD, who were under care at Oregon Health Science University or Vandebilt University Medical Center, were recruited. Age-matched non-demented controls were also recruited. Separate data were available for the performance of biomarkers in distinguishing between AD and non-AD dementia. We did not include data on performance of the index test to discriminate AD participants from controls. Sampling process and exclusion criteria not reported.
Patient characteristics and setting	The sample considered in the review comprises of 27 participants, 19 AD and 8 non-AD dementia (1 DLB; 3 NPH; 3 primary progressive aphasia; 1 hippocampal sclerosis). Ten controls were also recruited in the primary study. Most patients were evaluated by neuroimaging biomarkers. There was no significant difference in age or education level among the study groups. Duration of dementia was not significantly different between patients with probable Alzheimer disease or other dementias Sex: Not reported Age (SD) (y): 65.3±8.7 for AD; 66.6±4.4 for non-AD MMSE: 24 (19 to 27) for AD; 28 (25 to 29) for non-AD Duration of disease (y): 4.2±0.7 for AD; 4.2±0.7 for non-AD

Montine 2001 *(Continued)*

	Sources of recruitment: patients under care of the Oregon Health Science University or Vandebilt University Medical Center, Nashville, USA. Not reported whether inpatients or outpatients.
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using Athena Diagnostics (Worcester, Mass). Threshold: 1125 pg/ml; prespecified using the published cut-off (Fig 1, p512) Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from non-AD dementia) Reference standards: NINCDS-ADRDA criteria for AD. Clinical diagnosis of non-AD dementia was established according to 'best clinical judgement'. No further details reported. Clinical diagnosis was established prior the results of the index test
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported. However, it appears that CSF samples were collected short after establishing the clinical diagnosis and following informed consent Sample included in the analysis: 19 AD; 8 non-AD (1 DLB; 3 NPH; 3 primary progressive aphasia; 1 hippocampal sclerosis) AD vs non-AD (n=27) TP=19; FP=6; FN=0; TN=2 (Fig 1A and Fig 2, p512) Sensitivity=100%; Specificity=25% (Calculated in RevMan5)
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern

Montine 2001 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Paraskevas 2009

Study characteristics	
Patient Sampling	132 participants with dementia and 68 controls were recruited. Sampling procedure not reported. Separate data were available for the performance of the biomarkers in distinguishing AD from non-AD dementia, and AD from VD. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria: patients with one or more cardiovascular risk factors and patients with 1-2 white matter lacunes were excluded from AD group; patients with causes of secondary dementia (including thyroid dysfunction, B12 deficiency and possible neurosyphilis) and those using anticoagulant medication (contra-indication for lumbar puncture) were also excluded from the study.
Patient characteristics and setting	The sample considered in the review comprises of 115 participants: 92 AD, 23 VD. Seventeen participants with mixed dementia were not included in the analysis. 68 controls were also recruited, but not included in the analysis. All patients underwent clinical assessment. Both the VD and mixed groups had significant vascular disease on MRI or CT, either in the form of multiple infarctions, or multiple and/or confluent lacunar infarctions or 'leukoaraiosis of Binswanger type, together with multiple risk factors including hypertension, diabetes, obesity and/or carotid artery stenosis on ultrasound. None of the patients was under treatment for dementia at the time of lumbar puncture, but drugs for cardiovascular disease were allowed in patients with VD and mixed dementia.

Paraskevas 2009 *(Continued)*

	Sex: 36 males and 56 females for AD; 13 males and 10 females for VD; 9 males and 8 females for mixed dementia Age (SD) (y): 66 ± 10 for AD; 69 ± 10 for VD; 74 ± 7 for mixed dementia Disease duration (y): 3.4 ± 2.7 for AD; 2.9 ± 2.8 for VD; 3.1 ± 2.0 for mixed dementia Sources of recruitment: specialist care setting, Athens National University, Greece. Not reported whether inpatients or outpatients
Index tests	Patients gave CSF samples. The samples were stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 461 pg/ml; not prespecified; the cut-off levels (for individual markers, or their ratios) were calculated, with the resulting percentages of correct classification. Were the index test results reported without knowledge of the reference standard? [Yes]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (1. differential diagnosis of AD from VD; 2. differential diagnosis of AD from mixed dementia) Reference standards: NINCDS-ADRDA criteria Alzheimer's disease dementia Clinical diagnosis of VD and mixed dementia was based on NINDS-AIREN criteria. Clinical diagnosis was established prior the results of the index test
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported. However, it appears that CSF samples were collected short after establishing the clinical diagnosis and following informed consent. Sample included in the analysis: 92 ADD; 23 VD AD vs VD (n=115) TP=72; FP=7; FN=20; TN=16 (Fig 1, p207) Sensitivity=78%; Specificity=70% (Calculated in RevMan5)
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		

Paraskevas 2009 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Perani 2016

Study characteristics	
Patient Sampling	86 early dementia patients were recruited. Patients were referred to the memory clinics of the San Raffaele Hospital (Milan, Italy). They underwent clinical evaluation. Separate data were available for the performance of the biomarkers in distinguishing AD from MCI. We did not include data on performance of the index test to discriminate AD participants from MCI.

Perani 2016 *(Continued)*

	Exclusion criteria: reported
Patient characteristics and setting	The sample considered in the review comprises of 75 patients with dementia: 47 AD, 14 FTLD and 14 DLB. All patients underwent clinical assessment. Sex: 26 males and 21 females for AD; 8 males and 6 females for FTLD; 11 males and 3 females for DLB Age (SD) (y): 66±6.8 for AD; 65± 7.3 for FTLD; 72± 6 for DLB Disease duration (y): 39 ± 24 for AD; 32±19 for FTLD; 42±22 for mixed dementia
Index tests	Patients gave CSF samples. The samples were stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 500 pg/mL; pre-specified Were the index test results reported without knowledge of the reference standard? [Yes]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (1. differential diagnosis of AD from FTLD and DLB; 2. differential diagnosis of AD from FTLD only) Reference standards: NINCDS-ADRDA criteria for Alzheimer's disease dementia McKeith criteria for DLB and Rascovsky et al., 2013 for FTLD.
Flow and timing	All biomarker data were collected within 3 months from the baseline clinical visit.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		

Perani 2016 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		Unclear risk	

Rosler 2001

Study characteristics	
Patient Sampling	170 patients were recruited: 27 patients probable AD, 24 with non-AD dementias, 70 with various infectious, immunological, neurodegenerative, neoplastic and vascular central nervous system (CNS) diseases without cognitive impairment (OND) and 49 without CNS disease (CO). Sample procedure not reported. Separate data were available for the performance of biomarkers in distinguishing between AD and non-AD dementia. We did not included data on performance of the index test to discriminate AD participants from controls. Exclusion criteria: not reported.
Patient characteristics and setting	Sample included in the review comprised of 51 participants: 27 patients with probable AD according to the NINCDS-ADRDA criteria (McKhann 1984); 11 patients had early onset and 16 patients late onset of the disease; 24 patients with non-AD dementias: 4 Parkinson's disease with dementia, 5 vascular dementia 2 diffuse Lewy body disease, 1 progressive supranuclear palsy, 2 multisystem degeneration, 1 Pick's disease, 1 Huntington's disease and 8 normal pressure hydrocephalus. Age: <65 years early onset AD; >65 years late onset AD; not reported for the non-AD group Sex: 9 males and 18 females for AD, 13 males and 11 females for non-AD dementias Sources of recruitment: not reported. Residual lumbar CSF samples archived for research purposes were enrolled. The study was conducted at the Ludwig Boltzman Institute of Clinical Neurobiology, Vienna, Austria.
Index tests	Patients gave CSF samples. The samples were stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 375 pg/ml; not pre-specified, Cut-offs were determined by ROC analysis.

Rosler 2001 *(Continued)*

	Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from non-AD dementia) Reference standards: NINCDS-ADRDA criteria for AD. It was not reported whether the results of the reference standard results were interpreted without knowledge of the results of the index test.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported. Sample included in the analysis: 27 AD; 24 non-AD participants (5 VD; 4 PDD; 2 DLB, 8 NPH; 1 progressive supranuclear palsy, 2 multisystem degeneration, 1 Pick's disease, 1 Huntington's disease) AD vs non-AD (N=51) Sensitivity=78%; Specificity=58% (p234) TP=21; FP=10; FN=6; TN=14 (Calculated in RevMan; Fig 1b, p236)
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

Rosler 2001 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Santangelo 2017

Study characteristics	
Patient Sampling	326 patients were included: 165 patients with AD, 34 with NPH, 43 with FTD, 22 with LBD, 19 with PSP/CBS, 11 with VaD. Sample procedure not reported. We did not include data on performance of the index test to discriminate AD participants from controls or AD participants from patients with PSP/CBS. Exclusion criteria: reported.
Patient characteristics and setting	Age at diagnosis and disease duration and education:Reported Sex: 64 males and 101 females for AD; 6 males and 5 females for VD, 26 males and 17 females for FTD, 14 males and 8 females for DLB, 23 males and 11 females for NPH Sources of recruitment: A sample who were admitted to the Memory Centre of IRCCS-San Raffaele Hospital, Milan, Italy between December 2008 and July 2015.
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 500 pg/ml; pre-specified, Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from non-AD dementia)

Santangelo 2017 *(Continued)*

	Reference standards: NINCDS-ADRDA criteria for AD.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported. Patients underwent lumbar puncture at the baseline visit.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		Unclear risk	

Schirinzi 2015

Study characteristics	
Patient Sampling	Patients received lumbar puncture for diagnostic purposes at the Neurology unit of Policlinico Tor Vergata, Rome-Italy between 2012 and 2014.
Patient characteristics and setting	CSF samples from 28 participants: 14 ADD and 14 iNPH. Sex: 6 males and 8 females for AD and 8 males and 6 females for iNPH Age (SD) (y): 69.85 ± 7.42AD; 73.21 ± 4.63 for iNPH
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, stored on ice and sent to local laboratory and analysed (within 1 hour). Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 371pg/mL, not pre-specified, determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? Not reported
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from idiopathic NPH) Reference standards: NINCDS-ADRDA criteria for AD. Subjects received a diagnosis according to iNPH guideline criteria for possible iNPH. It was not reported whether the results of the reference standard results were interpreted without knowledge of the results of the index test.
Flow and timing	Clinical diagnosis and CSF sample collection was done on the same day.
Comparative	
Notes	

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear

Schirinzi 2015 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Shi 2018

Study characteristics	
Patient Sampling	Patients were recruited from six centers: the AD Core Centre, the Penn Memory Center, the Frontotemporal Degeneration Center, the Amyotrophic Lateral Sclerosis Center, the Parkinson disease and Movement Disorder Clinic, and then Penn Udall Center for Parkinson's Research at the University of Pensylvania. Patients were divided into two cohorts (clinical and neuropathologically confirmed diagnoses). The Clinical cohort (n=540) excluded participants with CSF haemoglobin >500 ng/mL and included: 165 AD, 105 MCI, 70 FTD, 10 CBD, 79 Lewy-body disorders, 11 PSP, amd 69 healthy controls. Separate data were available for the performance of biomarkers in distinguishing between ADD from FTD or DLB. We did not include data on performance of the index test to discriminate AD participants from cognitively healthy participants. Sampling procedure: not reported. Exclusion criteria: not reported.
Patient characteristics and setting	Sex: 66 male, 99 female for ADD; 37 male, 23 female for FTD; 8 male, 8 female for DLB. Age mean (range): 72 (53-78) for ADD; 64 (56-67) for FTD; 67.5 (64.5-74.5) for DLB. MMSE: Not reported. Disease duration (y): 2 (1-4) for ADD; 2 (1-4) for FTD; 2 (1-3) for DLB. Sources of recruitment: six centers specialising in AD, FTD, ALS, and PD research at Pensylvania University.
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed.

Shi 2018 *(Continued)*

	Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics, Ghent, Belgium. Threshold: not pre-specified, optimal cut-offs calculated. Were the index test results reported without knowledge of the reference standard? [Unlcear].
Target condition and reference standard(s)	Target condition: Alzheimer's disease (differential diagnosis of ADD from FTD and DLB). Reference standard: NIA/AA criteria for ADD. FTD was diagnosed according to the Rascovsky criteria, DLB according to McKeith criteria. It was unclear if the reference standard was blinded to the results of the index test.
Flow and timing	AD vs DLB (n=156) AD=114; DLB= 42; Sensitivity=89%; Specificity=74% (Table 2, p381) TP=93; FP=13; FN=12; TN=37 (calculated in RevMan5) AD vs FTD (n=170) AD=114; FTD=56; Sensitivity=80%; Specificity=80% (Table 2, p381) TP=95; FP=10; FN=24; TN=41 (calculated in RevMan5) Missing data: 31.3% of samples were excluded if haemoglobin >500 ng/mL. The interval between established clinical diagnosis and CSF sample collection was not reported.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 3: Reference Standard			

Shi 2018 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Sjogren 2000

Study characteristics	
Patient Sampling	Patients were consecutively recruited from either a prospective longitudinal study of patients with dementia (the Mölndal prospective dementia study; demented patients and controls), or similar studies at the Clinic of Neuropsychiatry, University Hospital, Malmö (all the dysthymia and 5 FTD patients) or similar studies at the Department of Geriatrics, Linköping (all the PD patients). Control group (32) without history, symptoms, or signs of psychiatric or neurological disease, malignant disease or systemic disorders and with MMSE score or at least 28 was also recruited. We did not include data on performance of the index test to discriminate AD participants from controls. Separate data were available for the performance of the biomarkers in distinguishing ADD from VD, and ADD from FTD. Exclsion criteria: participants with un-specified dementia, mixed dementia, history of severe psychiatric disease, chronic alcoholism, non-degenerative neurological disease, severe head injury, severe CNS infections, systemic diseases (e.g. maliganant tumour, liver disease), or secondary causes for dementia according to DSM-III-R were excluded
Patient characteristics and setting	The sample considered in the review comprises of 102 participants: 37 early AD defined as onset at or before 65 years; 23 late AD defined as onset after 65 years; 17 FTD; 25 VD (subcortical white-matter dementia, SWD, 'a putative subtype of VD'). We did not consider 23 Parkinson's disease (PD), 19 dysthymia and 32 controls in the analyses. All patients underwent a thorough clinical investigation including medical history, physical, neurologic and psychiatric examinations, laboratory blood tests, routine CSF analysis, ECG, chest X-ray, EEG, CT or MRI of the brain and investigation of regional cerebral blood flow using SPECT or ^{133}xenon inhalation

Sjogren 2000 *(Continued)*

	technique. At all the localities, clinical evaluation and diagnosis were made according to a Swedish consensus (Wallin 1994) that complies with international standards. Sex: 27 males and 33 females for AD total sample; 62.4 ±10.2 for FTD; 18 males and 7 females for SWD; 17 males and 6 females for PD; 10 males and 9 females for dysthymia Age (SD) (y): 66.0 ±7.8 for AD total sample; 6 males and 11 females for FTD; 62.4 ±10.2 for SWD; 47.2±15.0 PD; 47.2±15.0 for dysthymia Disease duration (y): 3.5±2.3 for AD total sample; 4.9±3.1 for FTD; 2.8±1.9 for SWD Sources of recruitment: specialist care setting; multicentre; Institute of Clinical Neuroscience, Gobteborg University and Neuropsychiatric Clinic, Malmo University Hospital, Sweden. Not reported whether inpatients or outpatients
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, stored on ice and sent to local laboratory and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 537pg/mL, not pre-specified, Cut-off value, sensitivity and specificity were determined according to suggestions by Altman 1997. A specificity level of approximately 85% for controls (the proportion of true negative cases) was chosen when determining the cut-off values. From the cut-off levels, sensitivity values for each diagnostic group and CSF-marker were obtained. This specificity level has been recommended in a consensus report on biochemical markers for AD (The Ronald and Nancy Reagan Research Institute, 1998). Were the index test results reported without knowledge of the reference standard? Not reported
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (1. differential diagnosis of AD from VD; 2. differential diagnosis of AD from FTD) Reference standards: NINCDS-ADRDA for AD. Clinical diagnosis of VD was based on NINDS-AIREN criteria, of FTD on The Lund/Manchester criteria. Clinical diagnosis was established prior the results of the index test.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported.However, it appears that CSF samples were collected shortly after establishing the clinical diagnosis. Sample included in the analysis: 132 participants: 60 AD (37 early onset AD; 23 late onset AD); 17 FTD; 25 VD (SWD) AD vs VD (n=84) TP=56; FP=16; FN=4; TN=8 Sensitivity=93%; Specificity=33% (Calculated in RevMan5) AD vs FTD (n=77) TP=55; FP=7; FN=5; TN=10 Sensitivity=92%; Specificity=59% (Calculate in RevMan5) Missing data: CSF Abeta42 sample was unavailable from 1 VD participants
Comparative	
Notes	
Methodological quality	

Sjogren 2000 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval be-	Yes		

Sjogren 2000 *(Continued)*

tween index test and reference standard?		
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk

Smach 2008

Study characteristics	
Patient Sampling	181 participants were randomly selected from the population register and consecutively evaluated at Sahloul University Hospital. The study also included 53 age-matched controls with absence of memory complaints and cognitive symptoms, preservation of general cognitive function and no no other active neurological or psychological disease. Separate data were available on the performance of biomarkers to distinguish AD from non-AD dementia. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria: not reported.
Patient characteristics and setting	The sample considered in the review comprises of 108 participants: 73 AD and 35 non-AD dementia (18 VD; 7 mixed dementia; 5 FTD; 3DLB; 2 unclassified dementia). CSF was not obtained from 20 AD patients. Controls were not included in the review. Participants underwent a clinical examination inc. medical history, neurological and neuropsychological examination, MMSE, laboratory screening tests and MRI. Sex: 49 males and 44 females for AD; 17 males and 18 females for non-AD dementia Age (range) (y): 73 (48–85) for AD; 69 (58–85) for non-AD dementia MMSE: 14 (0–26) for AD; 18 (10–27) for non-AD dementia Disease duration (y): 2 (1–9) for AD; 2 (1–6) for non-AD dementia Sources of recruitment: specialist care setting; population register of the inhabitants in Tunis, Tunisian Republic, Africa. Not reported whether inpatients or outpatients.
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 505 pg/mL, not pre-specified, determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? Not reported
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from non-AD dementia) Reference standards: NINCDS-ADRDA for AD. Clinical diagnosis of non-AD dementia was based on DSM-IV. Clinical diagnosis was established prior the results of the index test.

Smach 2008 *(Continued)*

Flow and timing	The interval between established clinical diagnosis and blood sample collection was not reported.However, it appears that CSF samples were collected short after establishing the clinical diagnosis. Sample included in the analysis: 73 AD and 35 non-AD dementia (18 VD; 7 mixed dementia; 5 FTD; 3DLB; 2 unclassified) AD vs non-AD (n=108) TP=60; FP=10; FN=13; TN=25 (p147) Sensitivity=82%; Specificity=71% (Calculate in RevMan5) Missing data: adequate CSF sample was not obtained for 20 patients with AD.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Smach 2008 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		High risk

Spies 2010

Study characteristics	
Patient Sampling	Retrospective study using clinical and CSF information from a database at a university medical centre Alzheimer's centre. The database contains clinical data as well as biobanked CSF and serum of consecutive patients. All 138 patients with a clear cut diagnosis of dementia, whose CSF was available for Abeta42 and Abeta40 analysis, were included. In addition, 47 non-demented controls without neurological problems were included. Separate data were available for the performance of biomarkers in distinguishing AD from various other types of dementia. We did not include data on performance of the index test to discriminate AD participants from controls. Inclusion criteria: participants with clear diagnosis of dementia.
Patient characteristics and setting	The sample considered in the review comprises of 138 participants: 69 AD, 26 VD, 27 FTD and 16 DLB. Demographic details are not presented for all patients. Sex: 34 males and 35 females for AD; 17 males and 9 females for VD; 19 males and 8 females for FTD; 12 males and 4 females for DLB Age (SD) (y): 69±8 for AD; 35±29 for VD (n=20); 34 ±21 for FTD (n=26); 76±8 for DLB Disease duration (mo): 29±23 for AD (n=60); 72±9 for VD; 65±7 for FTD; 34±27 for DLB (n=8) Sources of recruitment: specialist care setting; CSF database of the Radboud University Nijmegen Medical Centre, The Netherlands. Not reported whether inpatients or outpatients
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: Not reported; Cut-offs were determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from VD, FTD and DLB)) Reference standards: NINCDS-ADRDA for AD

Spies 2010 *(Continued)*

	Clinical diagnosis of VD was based on NINDS-AIREN, of FTD on Neary criteria, of DLB on McKeith criteria. Clinical diagnosis was established prior the results of the index test.
Flow and timing	Dates not provided for CSF sample collection. Sample included in the analysis: 69 AD; 69 non-AD (26 VD, 27 FTD and 16 DLB) AD vs VD (n=95) Sensitivity=83%; Specificity=69% (Table 2, p475) TP=57; FP=8; FN=12; TN=18 (Calculated in RevMan5) AD vs FTD (n=96) Sensitivity=94%; Specificity=85% (Table 2, p475) TP=65; FP=4; FN=4; TN=23 (Calculated in RevMan5) AD vs DLB (n=85) Sensitivity=65%; Specificity=75% (Table 2, p475) TP=45; FP=4; FN=24; TN=12 (Calculated in RevMan5) AD vs non-AD (n=138) Sensitivity=83%; Specificity=74% (Table 2, p475) TP=57; FP=18; FN=12; TN=51 (Calculated in RevMan5)
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			

Spies 2010 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Stefani 2005

Study characteristics	
Patient Sampling	Patients (n=140) were consecutively evaluated at a university hospital Alzheimer's centre, 86 patients were subsequently enrolled. A control group of 24 non-demented participants were also recruited. We did not include data on performance of the index test to discriminate ADD participants from controls. Exclusion criteria: isolated deficits or mostly subjective memory loss and/or stable MMSE (+/>25/30) on revisit; neuropsychological profile and behavioural symptoms suggest a diagnosis of FTD; suspected diagnosis of DLB; clinically manifest stroke in the last six months
Patient characteristics and setting	110 participants were enrolled in the study: 35 ADD, 31 ADD with WMC, 20 VD and 24 controls. The sample considered in the review comprises of 55 participants: 35 ADD and 20 VD. All patients provided medical history and underwent neurological examination, MMSE, complete blood screening (including thyroid function and B12), neuropsychological examination and neuroimaging. Neuropsychological follow-up included more comprehensive neuropsychological testing, including a standardised neuropsychological battery (Mental Deterioration Battery) and a complete psychiatric evaluation Sex: 16 males and 19 females for AD; 16 males and 16 females for AD & WMC; 11 males and 9 females for VD Age (years at LP): 72.2±8.1 for AD; 71.2±7.7 for AD & WMC; 73.6±6.8 for VD MMSE: 18.2±1.7 for AD; 19.1±1.5 for AD & WMC; 20.1±2.0 for VD

Stefani 2005 *(Continued)*

	Disease duration (mo at time of LP): 44.2±9.5 for AD; 143.5±8.9 for AD & WMC; 60.5±15.5 for VD Sources of recruitment: Alzheimer Center of the Department of Neuroscience,Tor Vergata University Hospital, Rome, Italy. Outpatients
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 493 pg/ml; not prespecified; 750 pg/ml for AD & AD with WMC vs VD; Cut-offs were determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of 1. AD and 2. AD & AD with WMC from VD) Reference standards: NINCDS-ADRDA and DSM IV criteria for AD; NINCDS-ADRDA criteria and MRI showing brain imaging findings suggesting subcortical vascular lesions for AD with WMC. Clinical diagnosis of VD was based on NINDS-AIREN criteria. Clinical diagnosis was established prior the results of the index test.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported. However, it appears that CSF samples were collected short after establishing the clinical diagnosis and following informed consent. Sample included in the analysis: 35 ADD; 20 VD AD vs VD (n=55) (cut-off 493 pg/ml) Sensitivity=77%; Specificity=80% (p86) TP=27; FP=4; FN=8; TN=16 (Calculated in RevMan5) All ADD and VD patients enrolled in the primary study were included in analysis. We did not considered ADD participants with WMC in the analysis.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		

Stefani 2005 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Tapiola 2000

Study characteristics	
Patient Sampling	The study included 187 participants. The definite AD group was recruited from a follow-up study of hospitalised patients in the geriatric department of Harjula hospital in Kuopio. The probable AD patients, patients with other dementias and neurological controls were recruited from diagnostic investigations in the Department of Neurology, Kuopio University hospital. Sampling procedure not reported. Separate data were available for the performance of biomarkers in distinguishing between AD and other dementias. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria: not reported.

Tapiola 2000 *(Continued)*

Patient characteristics and setting	The sample included in the study comprised of 187 participants: 41 definite AD cases, 80 patients with probable AD, 27 with other dementias (8 VD; 4 FTD; 5 LBD; 3 Parkinson's disease dementia; 7 unclassified dementia) and 39 neurological controls. This review included 107 participants: 80 with probable AD and 27 with non-AD dementia (8 VD; 4 FTD; 5 LBD; 3 Parkinson's disease dementia; 7 unclassified dementia) Sex: 34 males and 46 females for probable AD; 13 males and 14 females for other dementias Age (mean/SD) (y): 71±8 for probable AD; 71±10 for other dementias Disease duration (y): 2.6±1.9 for probable AD; 1.9±1.4 for other dementias Sources of recruitment: research centre, Department of Neurology, Kuopio University Hospital, Finland.
Index tests	Patients gave CSF samples. The samples were aliquoted, and stored at -70°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from Innogenetics NV, Gent, Belgium. Threshold: 340 pg/ml; not prespecified; Cut-offs were determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [Yes]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from other dementias) Reference standards: NINCDS-ADRDA for AD. Clinical diagnosis of other dementias was based on DSM-IV criteria. Clinical diagnoses were established prior the results of the index test.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported.However, it appears that CSF samples were collected short after establishing the clinical differential diagnosis of AD and other dementia. Sample included in the analysis: 107: 80 probable AD and 27 non-AD dementia (8 VD; 4 FTD; 5 LBD; 3 Parkinson's disease dementia; 7 unclassified dementia) Probable AD vs non-AD dementia (n=107) Sensitivity=69%; Specificity=59% (p739) TP=55; FP=11; FN=25; TN=16 (Calculated in Revman5)
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		

Tapiola 2000 *(Continued)*

Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Taricotti 2018

Study characteristics	
Patient Sampling	Retrospective study of CSF samples from 1137 out- and inpatients at the New York Presbyterian Hospital between 2005 and 2017. The study included 264 participants with ADD, 53 MCI, 65 DLB, 53 FTD, 31 vascular dementia, 21 progressive supranuclear palsy, 14 corticobasal degeneration, 218 NPH, 30 CJD, 37 non-specific psychaitric disorders, and 230 with subjective memory complaints. Participants with NPH were only included where they underwent ventriculoperitoneal shunt placement.

Taricotti 2018 *(Continued)*

	Separate data were available for the performance of biomarkers in distinguishing between ADD from FTD or DLB. We did not include data on performance of the index test to discriminate AD participants from cognitively healthy participants. Sampling procuedure: participants were ascertained from medical records. Exclusion criteria: dementia of uncertain aetiology, or partially documented dementia diagnosis.
Patient characteristics and setting	Diagnoses were made by several different neurologists using standard criteria (see reference standar below). Sex: 106 male, 158 female for ADD; 33 male; 20 female for FTD; 33 male; 32 female for DLB, 18 male; 13 female for vascular dementia; 124 male, 94 female for NPH; 20 male, 10 female for CJD. Age mean (SD): 67.7 ± 10.4 for ADD; 63.6 ± 8.8 for FTD; 73.1 ± 7.9 for DLB; 70.2 ± 8.9 for vascular dementia; 76.8 ± 8.0 for NPH; 67.0 ± 9.9 for CJD. There was a significant difference in age between ADD and other dementia sub-types. MMSE: Not reported. Disease duration (y): not reported. Sources of recruitment: medical records of in- and outpatients at the New York Presbyterian Hospital.
Index tests	Patients gave CSF samples. The samples were collected in polypropylene tubes, centrifuged, aliquoted, and stored at -80°C and analysed. Abeta42 was measured using enzyme-linked immunosorbent assays, obtained from ADmark® ELISA kit. Threshold: pre-specified at <500 pg/ml. Were the index test results reported without knowledge of the reference standard? [Unlcear].
Target condition and reference standard(s)	Target condition: Alzheimer's disease (differential diagnosis of ADD from "other dementia", vascular dementia, DLB, FTD, CJD, and NPH with AD pathology). Reference standard: NINCDS-ADRDA criteria for ADD. FTD was diagnosed according to the Neary criteria, DLB according to McKeith criteria, referred criteria for CJH, NINDS-society for Progressive Supranuclear Palsy for PSP, Boeve criteria for CBD, and vascular dementia according to the NINDS-AIREN criteria. It was unclear if the reference standard was blinded to the results of the index test.
Flow and timing	AD vs other dementia (n=749) AD=264; other dementia=485; Sensitivity=81%; Specificity=54% (Table 2, p381) TP=197; FP=233; FN=46; TN=273 (calculated in RevMan5) AD vs DLB (n=329) AD=264; DLB= 65; Sensitivity=81%; Specificity=60% (Table 2, p381) TP=214; FP=26; FN=50; TN=39 (calculated in RevMan5) AD vs FTD (n=317) AD=264; FTD=53; Sensitivity=81%; Specificity=40% (Table 2, p381) TP=214; FP=32; FN=50; TN=21 (calculated in RevMan5) AD vs CJD (n=294) AD=264; CJD= 30; Sensitivity=81%; Specificity=40% (Table 2, p381)

Taricotti 2018 *(Continued)*

TP=214; FP=18; FN=50; TN=12 (calculated in RevMan5)

AD vs vascular dementia (n=295)

AD=264; vascular dementia=31; Sensitivity=81%; Specificity=39% (Table 2, p381)

TP=214; FP=19; FN=50; TN=12 (calculated in RevMan5)

Missing data: 121 (10.7%) excluded due to incomplete or uncertain diagnosis.

The interval between established clinical diagnosis and CSF sample collection was not reported.

Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Tariciotti 2018 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		High risk

Wiltfang 2003

Study characteristics	
Patient Sampling	The study included 19 patients with CJD, 19 patients with AD and 26 non-demented controls. Sampling procedure not reported. Separate data were available for the performance of biomarkers in distinguishing between AD and CJD participants. We did not include data on performance of the index test to discriminate AD participants from controls. Exclusion criteria: not reported.
Patient characteristics and setting	The sample considered in the review in the review comprised of 19 AD and 19 CJD participants. Sex: 5 males and 14 females for AD; 9 males and 10 females for CJD Age (median) (y): 76 (range, 54–80) for AD; 66 (range, 37–88) for CJD Sources of recruitment: specialist care setting. Not reported whether inpatients or outpatients. The study was conducted in Germany.
Index tests	Patients gave CSF samples. CSF sampling methods not described. Abeta42 was measured using SDS-PAGE immunoblot. Threshold: 1900 pg/ml; not prespecified; Cut-offs were determined by ROC analysis. Were the index test results reported without knowledge of the reference standard? [Not reported]
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia (differential diagnosis of AD from CJD) Reference standards: NINCDS-ADRDA and DSM-IV for AD. Clinical diagnosis of CJD was based on the clinical criteria (Otto 2002). 11/19 patients were later neuropathologically verified as definite CJD cases.

Wiltfang 2003 *(Continued)*

	Clinical diagnoses were established prior the results of the index test.
Flow and timing	The interval between established clinical diagnosis and CSF sample collection was not reported.However, it appears that CSF samples were collected short after establishing the clinical differential diagnosis of AD and CJD. Sample included in the analysis: 19 AD; 19 CJD AD vs CJD (n=38) Sensitivity: 100%; Specificity: 58% (p264) TP=19; FP=8; FN=0; TN=11 (Calculated in Revman5)
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Wiltfang 2003 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Alcolea 2014	Assessed temporal changes in the levels of CSF ABeta; therefore, data not available for creating 2 x 2 table
Alcolea 2017	Index text: threshold not used; data not available for creating 2 x 2 table
Balasa 2014	Data not available for creating 2 x 2 table
Berlyand 2016	Data not available for creating 2 x 2 table
Bertens 2017	Data not available for creating 2 x 2 table
Bibl 2007b	Data not available for creating 2 x 2 table
Bibl 2008a	Aim was not differential diagnosis of ADD from other dementia subtypes
Brandt 2008	Data presented not sufficient for constructing 2 x 2 table. Author contacted for the additional information. No reply.
Carandini 2019	Data not available for creating 2 x 2 table
Hall 2012	Data not available for creating 2 x 2 table
Hampel 2018	Data not available for creating 2 x 2 table
Han 2012	Data not available for creating 2 x 2 table
Illan-gala 2019	Data not available for creating 2 x 2 table (MCI combined with ADD)
Karadas 2017	Data not available for creating 2 x 2 table
Parnetti 2011	Index test: tau/a-Synuclein ratio. Data for 2 x 2 table for CSF Aβ1-42 biomarker not reported.
Prikrylova Vranova 2014	Data not available for creating 2 x 2 table
Skillback 2015	Data not available for creating 2 x 2 table
Smach 2008a	Index test: combined CSF ABeta42 and CSF t-tau. Author contacted for the relevant information regarding the accuracy of CSF ABeta only. No reply.
Stoeck 2014	Data not available for creating 2 x 2 table

Study	Reason for exclusion
Toledo 2012	Index test:combined CSF t-tau and CSF p-tau. The accuracy of CSF ABeta42 not assessed (email on 01/11/14 from Dr Toledo).
Uslu 2012	Data not available for creating 2 x 2 table
van Steenoven 2018	Data not available for creating 2 x 2 table
van Steenoven 2019	Data not available for creating 2 x 2 table
Vergallo 2017	Data not available for creating 2 x 2 table
Wennstrom 2015	Data not available for creating 2 x 2 table
Zwan 2014	Data not available for creating 2 x 2 table

ADDITIONAL TABLES

Table 1. Included studies and the index test accuracy at study level

Included studies and the accuracy of CSF Aβ42 for discriminating ADD from other dementia subtypes

Differential diagnosis	Study	Participants N (included in analysis)	Threshold assays	Threshold pre-specified	Test accuracy at study level	
					Sensitivity (%)	Specificity (%)
ADD versus non-ADD	Brettschneider 2006	N = 165: 109 ADD; 56 non-ADD (41 VaD; 15 FTD)	612 pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	82%	46%
	Kapaki 2003	N = 64: 49 ADD; 15 non-ADD (6 DLB; 4 FTD; 1 PDD; 2 PSP; 2 CBGD)	435 pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	71%	80%
	Knapskog 2018	N = 155: 138 ADD; 17 non-ADD (subtypes not specified)	550 pg/ml and 700 pg/ml ELISA, Innogenetics, Ghent, Belgium	Yes	43% and 35%	79% and 47%
	Lewczuk 2004	N = 33: 21 ADD; 11 non-ADD (5 VaD; 1 mixed; 1 SCASE; 1 SD; 1 FTD; 1 ARCD; 1 unspecified)	500 pg/ml ELISA, Innogenetics, Ghent, Belgium	No	86%	82%
	Lombardi 2018	N = 45: 32 ADD; 10 FTD; 3 unclassified cognitive decline	650 pg/ml 600 pg/ml ELISA (unspecified)	Yes No	73% and 87%	64%

Table 1. Included studies and the index test accuracy at study level *(Continued)*

	Maddalena 2003	N = 81: 51 ADD; 30 non-ADD (8 VaD; 3 FTD; 2 DLB; 2 PDD; 2 CJD; 2 CAA; 11 other)	490 pg/ml ELISA, Innogenetics, Belgium	No	78%	70%
	Montine 2001	N = 27: 19 ADD; 8 non-ADD (1 DLB; 3 NPH; 3 PPA; 1 hippocampal sclerosis)	1125 pg/ml Athena Diagnostics, Worcester, MA, USA.	Yes	100%	25%
	Rosler 2001	N = 51: 27 ADD (11 EO; 16LO); 24 non-AD (5 VaD; 4 PDD; 2 LBD; 8 NPH; 5 other)	375 pg/ml ELISA, Innogenetics, Ghent, Belgium	No	78%	58%
	Smach 2008	N = 108: 73 ADD; 35 non-ADD (18 VaD; 5 FTD; 3 DLB; 7 mixed; 2 unclassified)	505 pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	82%	71%
	Spies 2010	N = 138: 69 ADD; 69 non-ADD (26 VaD; 27 FD; 16 DLB)	Threshold not reported ELISA, Innogenetics NV, Ghent, Belgium	No	83%	74%
	Tapiola 2000	N = 107: 80 probable ADD; 27 non-ADD (8 VaD; 4 FTD; 5 LBD; 3 PDD; 7 unclassified) *Note: 41 definite ADD not included in analysis*	340 pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	69%	59%
	Taricciotti 2018	N = 749: 264 ADD; 485 non-ADD (65 DLB, 53 FTD, 31 VaD, 21 PSP, 14 CBD, 218 NPH, 30 CJD) *Note: 121 uncertain diagnosis not included in analysis*	500 pg/ml ADmark ELISA kit	Yes	81%	54%
	Perani 2016	N = 75: 47 ADD; 28 non-ADD (14 FTLD; 14 DLB)	500 ng/L ELISA, Innogenetics, Ghent, Belgium	Yes	85%	46%
ADD versus VaD	De Jong 2006	N = 86: 61 ADD; 25 VaD	520 pg/ml Innogenetics NV, Ghent, Belgium	No	82%	76%
	Kapaki 2003	N = 55: 49 ADD; 6 VaD	526 pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	82%	67%

Table 1. Included studies and the index test accuracy at study level *(Continued)*

	Lins 2004	N = 24: 12 ADD; 12 VaD	562 pg/ml ELISA, Innogenetics, Ghent, Belgium	No	67%	50%
	Marchegiani 2019	N = 87: 70 ADD, 17 VaD	431 pg/ml ELISA, Fujirebio Inc., Tokyo, Japan	No	65%	95%
	Paraskevas 2009	N = 115: 92 ADD; 23 VaD	461 pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	78%	70%
	Sjogren 2000	N = 85: 60 ADD (37 EO; 23 LO); 24 VaD (SWM dementia)	537 pg/ml ELISA, Innotest, Innogenetics, Belgium	Yes	93%	33%
	Spies 2010	N = 95: 69 ADD; 26 VaD	Threshold not reported ELISA, Innogenetics NV, Ghent, Belgium	No	83%	69%
	Stefani 2005	N = 55: 35 ADD; 20 VaD	493 pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	77%	80%
	Herbert 2014	N = 79: 64 ADD; 15 VaD	≤ 500pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	70%	87%
	Taricciotti 2018	N = 295: 264 ADD; 31 VaD (*Note: 121 uncertain diagnosis not included in analysis)*	500 pg/ml ADmark ELISA kit	Yes	81%	39%
	Santangelo 2017	N = 176: 165 ADD; 11 VaD	≤ 500pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	Yes	82%	82%
ADD versus FTD	de Rino 2012	N = 114: 72 ADD; 42 bvFTD	104 pg/ml ELISA, Innogenetics, Ghent, Belgium	No	82%	21%
	Abu-Rumeileh 2018	N = 113: 60 ADD; 53 bvFTD (*Note: 10 FTD not included in analysis)*	482 pg/ml Innotest, Innogenetics, Ghent, Belgium	No	89%	80%
	Bibl 2007	N = 60: 30 ADD; 30 FTD (FTLD: 24 FTD; 5 PPA; 1 SD)	Threshold not reported ELISA	No	90%	90%

Table 1. Included studies and the index test accuracy at study level *(Continued)*

Casoli 2019	N = 76: 55 ADD; 21 FTD (12 bvFTD and 9 PPA)	Various (minimum threshold 112 maximum 1006 and 837 pg/ml) ELISA, Fujirebio Inc., Tokyo, Japan	No	100%	0%
Falgas 2020	N = 90: 64 AD; 26 FTD *Note: only 23 (18 FTD and 5 ADD) included in the analysis as MCI excluded*	494.95 pg/ml Innotest, Innogenetics, Ghent, Belgium	No	100%	94%
Kapaki 2008	N = 107: 76 ADD; 31 FTD (FTLD: 24 FTD; 7 PPA & FTD) *Note: 3 FTLD not included in analysis*	≤ 451 pg/ml ELISA, Innogenetics, Ghent, Belgium	No	75%	71%
Khoonsari 2019	N = 87: 76 ADD; 11 FTD (subtype unspecified)	530 pg/ml Innotest, Innogenetics, Ghent, Belgium	Yes	88%	91%
Lombardi 2018	N = 45: 32 ADD; 10 FTD (subtype not specified); 3 non-ADD	650 pg/ml 600 pg/mlELISA (unspecified)	Yes No	73 and 87%	70%
Marchegiani 2019	N = 93: 70 ADD; 23 FTD (19 FTD, 3 PSP, 3 CBD)	613 pg/ml ELISA, Fujirebio Inc., Tokyo, Japan	No	96%	57%
Shi 2018	N = 170: 114 ADD; 56 FTD (48 bvFTD, 8 CBS) *Note: samples excluded where haemoglobin was >500 ng/ml*	Threshold not reported ELISA, Innogenetics NV, Ghent, Belgium	No	80%	80%
Sjogren 2000	N = 77: 60 ADD (37 EO; 23 LO); 17 FTD	537 pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	Yes	92%	59%
Spies 2010	N = 96: 69 ADD; 27 FTD	Threshold not reported ELISA, Innogenetics NV, Ghent, Belgium	No	94%	85%
Baldeiras 2015	N = 214: 107 ADD; 107 FTD	≤ 538pg/ml ELISA, Innotest, Innogenetics, Belgium	No	70%	82%
Herbert 2014	N = 90: 64 ADD; 26 FTD	≤ 500pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	70%	88%

Table 1. Included studies and the index test accuracy at study level *(Continued)*

	Santangelo 2017	N = 208: 165 ADD; 43 FTD	≤ 500pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	Yes	82%	67%
	Tariciotti 2018	N = 317: 264 ADD; 53 FTD (*Note: 121 uncertain diagnosis not included in analysis)*	500 pg/ml ADmark ELISA kit	Yes	81%%	40%
	Perani 2016	N = 61: 47 AD; 14 FTD	500pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	Yes	85%	71%
ADD versus DLB	Aerts 2011	N = 65: 44 ADD; 21 DLB	> 482 pg/ml ELISA, Innogenetics NV, Ghent, Belgium	No	62%	65%
	Bibl 2006	N = 41: 18 ADD; 23 DLB	475 pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	50%	96%
	Bousiges 2018	N = 937: 783 ADD; 154 DLB	700 np/ml 606 pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	Yes No	71% and 85%	53% and 37%
	Spies 2010	N = 85: 69 ADD; 16 DLB	Threshold not reported. ELISA, Innogenetics NV, Ghent, Belgium	No	65%	75%
	Herbert 2014	N = 78: 64 ADD; 14 DLB	≤ 500pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	70.3%	50%
	Santangelo 2017	N = 187: 165 ADD; 22 DLB	≤ 500pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	Yes	82%	41%
	Shi 2018	N = 156: 114 ADD; 42 DLB (*Note: samples excluded where haemoglobin was > 500 ng/ml)*	Threshold not reported ELISA, Innogenetics NV, Ghent, Belgium	No	89%	74%
	Tariciotti 2018	N = 329: 264 ADD; 65 DLB (10 LBD, 32 PDD)	500 pg/ml ADmark ELISA kit	Yes	81%	60%

Table 1. Included studies and the index test accuracy at study level *(Continued)*

	Bousiges 2016	N = 51: 31 ADD; 20 DLB	500pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	Yes	77%	80%
ADD versus CJD dementia	Kapaki 2001	N = 50: 38 ADD; 12 CJD	445 pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	76%	42%
	Taricciotti 2018	N = 294: 264 ADD; 30 CJD	500 pg/ml ADmark ELISA kit	Yes	81%	40%
	Wiltfang 2003	N = 38: 19 ADD; 19 CJD	1900 pg/ml Aβ-SDS-PAGE immunobolt	No	100%	58%
ADD versus NPH dementia	Kapaki 2007	N = 85: 67 ADD; 18 NPH	> 268 pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	91%	44%
	Lins 2004	N = 24: 12 ADD; 12 NPH	562 pg/ml ELISA, Innogenetics, Ghent, Belgium	No	67%	33%
	Schirinzi 2015	N = 28: 14 ADD; 14 NPH	371 pg/ml ELISA (unspecified)	No	73.3%	81.3%
	Santangelo 2017	N = 199: 165 ADD; 34 NPH	≤ 500pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	Yes	82%	26%
ADD versus ARCD dementia	Kapaki 2005	N = 53: 33 ADD; 20 ACRD	≤ 562 pg/ml ELISA, Innotest, Innogenetics, Ghent, Belgium	No	85%	80%

ADD: probable or possible Alzheimer's disease dementia; ARCD: alcohol-related cognitive disorder; CAA: cerebral amyloid angiopathy; CBGD: corticobasal-ganglionic degeneration; CJD: Creutzfeldt-Jakob disease; DLB: dementia with Lewy bodies; EO: early onset; FTD: frontotemporal dementia; FTLD: frontotemporal lobe degeneration; LO: late onset; N: a number of participants included in the analysis in the review; non-ADD: two or more other subtype dementias; NPH: normal pressure hydrocephalus; PDD: Parkinson's disease dementia; PPA: primary progressive aphasia; PSP: progressive supranuclear palsy; SASE: subcortical arterial sclerotic; SD: semantic dementia; VaD: vascular dementia; WMC: white matter changes

HISTORY

Protocol first published: Issue 1, 2014
Review first published: Issue 2, 2021

CONTRIBUTIONS OF AUTHORS

All authors contributed to the drafting of the review.

DECLARATIONS OF INTEREST

None known.

SOURCES OF SUPPORT

Internal sources

- None, Other

External sources

- None, Other
- NIHR, UK

This review was supported by the National Institute for Health Research (NIHR), via Cochrane Infrastructure funding to the Cochrane Dementia and Cognitive Improvement group. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, National Health Service or the Department of Health

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

In the protocol, we planned to separately examine those studies that included 30% patients below the age of 65. Not all studies reported the proportion of participants aged under 65, so we focussed on those with a proportion of more than 30%, or studies where the mean age of ADD participants was below 66 years. In the protocol we had not planned to investigate the test accuracy of CSF ABeta42 between ADD and FTD subtypes. However, different FTD subtypes have different presentations, and some are pathologically closer to ADD (primary progressive aphasias) than FTD. Furthermore, many studies also included progressive supranuclear palsy and corticobasal syndrome under FTD, and the pathology of these disorders are distinct from that of more classical behavioural variant FTD. Given this significant heterogeneity in the FTD sample enrolled by studies, we performed subgroup analyses of FTD subtype where sufficient data permitted.

INDEX TERMS

Medical Subject Headings (MeSH)

Alcoholism [complications]; Alzheimer Disease [blood] [cerebrospinal fluid] [*diagnosis]; Amyloid beta-Peptides [*blood] [*cerebrospinal fluid]; Bias; Biomarkers [blood] [cerebrospinal fluid]; Cognitive Dysfunction [blood] [cerebrospinal fluid] [diagnosis] [etiology]; Confidence Intervals; Creutzfeldt-Jakob Syndrome [blood] [cerebrospinal fluid] [diagnosis]; Dementia, Vascular [blood] [cerebrospinal fluid] [diagnosis]; Diagnosis, Differential; Frontotemporal Dementia [blood] [cerebrospinal fluid] [diagnosis]; Hydrocephalus, Normal Pressure [blood] [cerebrospinal fluid] [diagnosis]; Lewy Body Disease [blood] [cerebrospinal fluid] [diagnosis]; Likelihood Functions; Peptide Fragments [*blood] [*cerebrospinal fluid]; Sensitivity and Specificity

MeSH check words

Humans