Exploring the Accuracy of the Preclinical Alzheimer's Cognitive Composite (PACC) Relative to Amyloid Pathology for the Diagnosis of Alzheimer's Disease: A Systematic Review Protocol

By Sofie Imer Jespersen

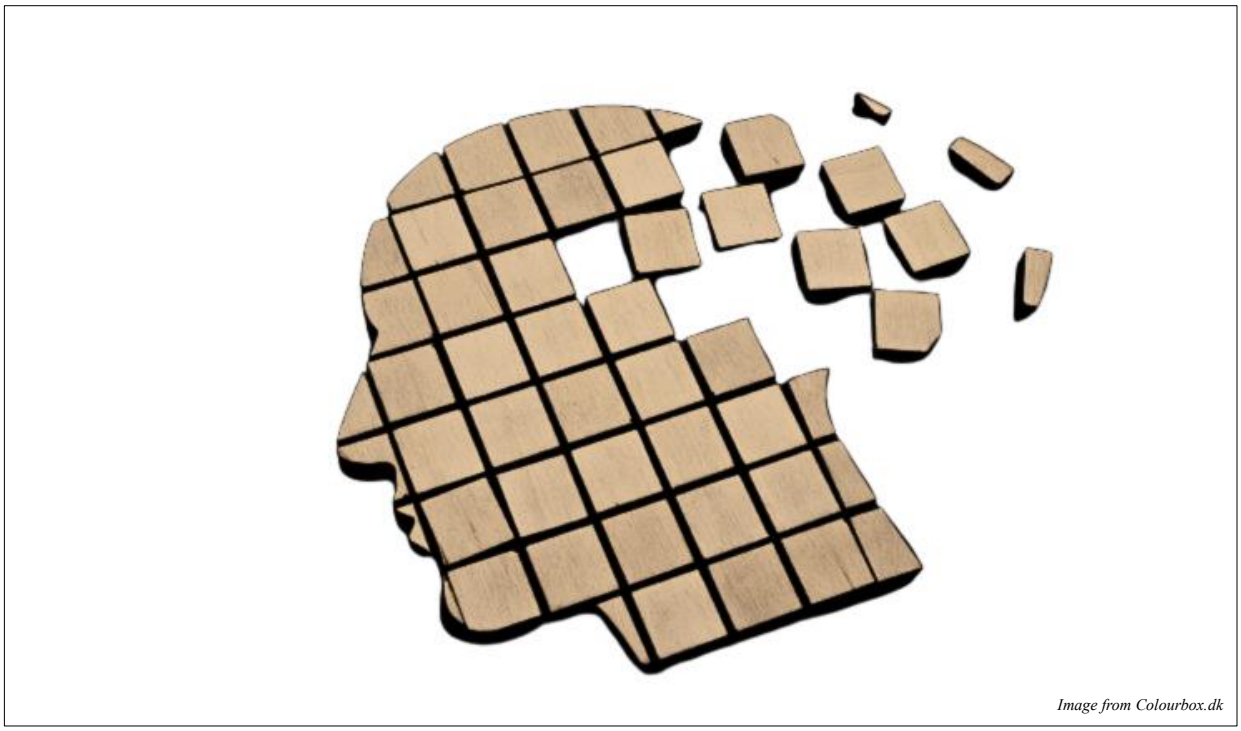

Image from Colourbox.dk

Study number: 20185135

10th semester, Psychology, Aalborg University

Master's Thesis

31-05-2024

Number of characters in total: 128.382

Number of pages of 2400 characters: 53,5

Supervisor: Casper Schmidt

Abstract

Background: Preclinical Alzheimer's Disease (AD) can be defined by an abnormally high level of amyloid pathology (Aβ) in the brain. The need for earlier discovery of the disease increases, as the AD prevalence is rising globally. The Preclinical Alzheimer's Cognitive Composite (PACC) is a compilation of neuropsychological tests specifically designed to detect the earliest symptoms of preclinical AD. **Methods:** The current study consists of a comprehensive systematic review, investigating PACC and its sensitivity across 20 studies. The review is based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and Diagnostic Test Accuracy (DTA) protocol by Joanna Briggs Institute (JBI). The DTA protocol suggests comparing the index test (PACC) to a consistent reference test (Aβ-PET imaging). The sensitivity, accuracy and specificity were investigated across all reviewed studies. **Results:** Present findings add to profound evidence that even very low levels of Aβ-PET levels are associated with subtle cognitive changes as measured by PACC. And a higher level of Aβ leads to a faster rate of cognitive decline. Across different versions of PACC, it was found that especially the Mini Mental State Examination (MMSE) and Digit Symbol Substitution Test (DSST) subtests were highly sensitive towards early cognitive change due to preclinical AD. Additional subtests measuring language, like the Category Fluency Test (CAT), could potentially enhance the overall sensitivity of PACC. Furthermore, PACC can detect cognitive changes in both early and late stages of onset AD, and it can differentiate between preclinical AD, Mild Cognitive Impairment (MCI), and AD, which makes it a great and highly sensitive assessment tool both in clinical trials and clinical settings. **Limitations:** The limitations in this study include potential biases inherent in the selected studies, such as selection and attrition biases, which may affect the generalizability of the findings. Additionally, the reliance on specific subtests within PACC may not capture all aspects of cognitive decline, and the inclusion of additional subtests requires further validation. **Conclusion:** The included studies find that elevated Aβ levels

puts you at higher risk of developing cognitive decline and possibly AD. This also underlines the importance of Aβ as one of the first pathological biomarkers to appear on a trajectory to developing AD. However, further evidence is needed to fully verify the causal relationship between Aβ pathology and cognition, and to validate PACC for clinical settings.

Keywords: Preclinical Alzheimer's Disease, Amyloid-pathology (Aβ), Amyloid-PET, Preclinical Alzheimer's Cognitive Composite (PACC), Cognitive Decline, Diagnostic Test Accuracy, Systematic Review

Table of Contents

Glossary

3MSE = Modified MMSE

11C-PiB = Pittsburgh Compound B

18F-FBP = 18F-Florbetapir

18F-FDG = 18F-Fludeoxyglucose

18F-FLUTE = 18F-Flutemetamol

18F-FTP = 18F-Flortaucipir

A4 = Anti-Amyloid Treatment of Asymptomatic Alzheimer's Disease

Aβ = Amyloid-beta

Aβ+ = High amyloid level

Aβ++ = Very high amyloid level

Aβ- = No amyloid pathology

ACh = Acetylcholine

AD = Alzheimer's Disease

ADAS-Cog = Alzheimer Disease Assessment Scale-Cognitive Subscale

ADCS = Alzheimer's Disease Cooperative Study

ADCS-PI = ADCS-Prevention Instrument

AIBL = Australian Imaging Biomarkers and Lifestyle study

AUB = Aalborg University Library

AVLT = Rey Auditory Verbal Learning Test

AVLT-DR = AVLT-Delayed Recall

AVLT-R = AVLT-Recognition

BD = Block Design from WAIS-R

BNT = Boston Naming Test

BPSO = Behavioural Pattern Separation Task-Object version

CAT = Category Fluency Test

CBB = Cogstate Brief Battery

CBF = Cerebral Blood Flow

CDR = Clinical Dementia Rating Scale

CFI = Cognitive Function Index

CI = Cognitively impaired

CL = Centiloids

COWAT = Controlled Oral Word Association Test

CFA = Confirmatory factor analysis

CSF = Cerebral Spinal Fluid

CVLT-II = California Verbal Learning Test-Second Edition

DCTclock = Digital Clock-drawing Test

DET = Detection Task

DSC = Digit Symbol-Coding

DSST = Digit Symbol Substitution Test

DTA = Diagnostic Test Accuracy

DVR = Distribution Volume Ratio

D-KEFS = Delis-Kaplan Executive Function System

DWR = Delayed Word Recall

FAS = Also known as COWAT: Controlled Oral Word Association Test with the letters F+A+S

FCSRT = Free and Cued Selective Reminding Test

FCSRT-Free = Free and Cued Selective Reminding Test Free Recall

FDA = Food and Drug Administration

FNAME = Face Name Associative Memory Exam

GCC = Global Cognitive Composite

Global-z = Global composite using z-scores

HABS = Harvard Ageing Brain Study

HR = Hazard ratio

IDN = Identification Task

IPACC = Latent PACC

IRT = Item Response Theory

J-ADNI = Japan-ADNI

JBI = Joanna Briggs Institute

LM-DR = Logical Memory Delayed Recall

LM-II = Logical Memory II

LM-IR = Logical Memory Immediate Recall

LMM = Linear Mixed-effects Model and Logistic Mixed-effects Model

M-Age = Mean age

MCI = Mild Cognitive Impairment

MCSA = The Mayo Clinic Study of Aging

MMRM = Mixed method of Repeated Measures

MMSE = Mini-Mental State Examination

MLR = Maximin Likelihood estimation with Robust standard errors

MoCA = Montreal Cognitive Assessment

NA-ADNI = Northern American ADNI

NIA-AA = National Institute on Aging-Alzheimer's Association

NSHD = National Survey of Health and Development

NPV = Negative predictive value

NYU = New York University Paragraph Recall

OCL = One Card Learning Task

ONB = One Back Task

PACC = Preclinical Alzheimer's Cognitive Composite

PACC-R = PACC-Revised

PC = Picture Completion from WMS

PET = Positron Emission Topography

PFDR = P-value of False Discovery Rate

PPV = Positive predictive value

ROC = Receiver Operating Characteristic analysis

ROIs = Regions of Interest

ROS = Reactive Oxygen Species

SCD = Subjective Cognitive Decline

SD = Standard deviation

SUVR = Standard Uptake Value Ratio

TMT = Trail-Making-Test

VR-II = Visual Reproduction-II from WMS

WAIS-R = Wechsler Adult Intelligence Scale-Revised

WMHV = White Matter Hyperintensity Volume

WMS = Wechler Memory Scale

WRAP = Wisconsin Registry for Alzheimer's Prevention

ZAVEN = Z-scores of Attention, Verbal Fluency and Episodic Memory for Nondemented older adults

zPACC = Standardized PACC with z-scores

♀ = Percentage of female participants

1.0 Introduction

As people are getting older globally, there has been an increase in age-related diseases (Winblad et al., 2016). One of the primary risk factors for developing Alzheimer's Disease (AD) is old age, which explains the increase in the prevalence of the disease worldwide (Winblad et al., 2016). AD is the most common cause of dementia, and accounts for an estimated 60-80% of all dementia cases in the U.S. (Alzheimer's Association, 2024). Globally, the number of people with AD is estimated to be 32 million as of 2023 (Gustavsson et al., 2023), and it has further been assessed that the worldwide cost of dementia exceeds more than a trillion US dollars every year. AD imposes not only an economic burden but also significant emotional and social costs on both the individual and their network of support, including family and friends. The toll of the disease extends beyond financial expenses and is impacting the well-being and quality of life for those affected by the disease (Alzheimer's Association, 2024).

The prevalence of AD seems to be higher in women than in men (Alzheimer's Association, 2024). However, while women tend to live longer than men in general, and advancing age is a significant risk factor for AD development, it is not conclusively clear that the prevalence of AD is higher in women (Alzheimer's Association, 2024). Factors such as genetic predisposition, hormonal influences, and healthcare-seeking behaviors may influence the prevalence of AD in both genders (Alzheimer 's Association, 2024). While AD primarily impacts the elderly population, it should not be regarded as a natural consequence of ageing or as an amplification of it (Irwin et al., 2018), rather, the disease is diagnosed clinically based on a variety of symptoms. These symptoms encompass cognitive impairments such as memory loss, challenges in learning, different executive functions, language difficulties, problems with complex attention, perceptual-motor impairments, and social cognition (Irwin et al., 2018; World Health Organization, 2019). Furthermore, the cognitive decline is progressive and relentless, as the disease is currently irreversible (Breedlove & Watson, 2013).

AD is recognized to start years, even decades, before the first symptoms of cognitive decline appear (Alzheimer's Association, 2024). Amyloid pathology build-up and accumulation in the brain is considered to be the first of a sequence of pathologic events to appear prior to the first symptoms of cognitive decline (Donohue et al., 2014; Farrell et al., 2021; Jack et al., 2010; Jack et al., 2018; Mormino et al., 2017; Sperling et al., 2014). Subsequently, the abnormally high level of amyloid pathology in the brain sometimes initiates a cascade of biomarkers, where the amyloid accumulation triggers the development of neurofibrillary tangles (NFTs), which is made up of tau proteins within the neurons, causing them to deteriorate and eventually die (Breedlove & Watson, 2013; Jack et al., 2010). These pathological changes in the brain consequently lead to neuronal death (neurodegeneration) resulting in the progression of cognitive decline (Breedlove & Watson, 2013).

The preclinical stage of AD refers to the period before any signs of cognitive decline becomes apparent, usually determined by an elevated level of amyloid pathology but not yet measurable tau tangles, as tau tangles are directly related to the magnitude of cognitive impairment (Breedlove & Watson, 2013). There is a growing interest in preclinical AD in research (Alzheimer's Association, 2024.; Bransby et al., 2019; Buckley et al., 2017; Donohue et al., 2014; Farrell et al., 2021; Jack et al., 2010; Jack et al., 2018; Lim et al., 2016; Mormino et al., 2017; Sperling et al., 2014; Sperling et al., 2011; Winblad et al., 2016). The number of people with preclinical AD is estimated to be around 315 million people worldwide (Gustavsson et al., 2023). Preclinical AD is a relatively new concept that refers to a condition that is characterized by the presence of AD biomarkers but an absence of specific clinical symptoms (Alzheimer's Association, 2024). Discovering potential AD progression at the preclinical stage presents many benefits. The sooner the discovery of AD biomarkers can be made, the less damage they will do to the brain, as these impairments are permanent. AD biomarkers are usually confirmed by detecting amyloid-beta and tau as shown on positron emission topography (PET) scans, or by lumbar puncture showing the level of biomarkers in the cerebral spinal fluid (CSF)

(Alzheimer’s Association, 2024), as these methods are currently the gold standards for diagnosis. However, these are invasive and costly methods, and they are usually done at stages, where the disease has already progressed in the brain (Alzheimer’s Association, 2024). Besides these methods, early cognitive decline can be measured with a set of various neuropsychological tests. A test battery that is regularly used to determine cognition in preclinical AD is the Preclinical Alzheimer's Cognitive Composite (PACC) (Bransby, 2019; Donohue et al., 2014; Mormino et al., 2017). This battery contains a compilation of the most sensitive tests for early discovery of cognitive decline (Donohue et al., 2014). It has been designed to serve as the primary outcome measure for trials conducted at the asymptomatic phase of AD, and it can detect the earliest cognitive changes due to AD (Donohue et al., 2014). PACC can assess episodic memory, timed executive functions, and global cognition (Donohue et al., 2014).

To this end, the aim of this review is to comprehensively evaluate the sensitivity, accuracy, and specificity of PACC in detecting the earliest signs of preclinical AD. This will be done in relation and comparison to the gold standard of amyloid-PET. Following the guidelines outlined by the Joanna Briggs Institute (JBI) for "Diagnostic Test Accuracy" (DTA) protocols, this investigation seeks to provide a thorough assessment of PACC, by systematically analyzing existing literature and synthesize the findings from relevant studies. This review aims to contribute valuable insights into the diagnostic utility and clinical applicability of the PACC test battery in identifying individuals at risk of developing AD. PACC may provide a more accessible and economically feasible option for early detection of AD, which aligns with the goal of optimizing healthcare resources and making diagnostic procedures more accessible and affordable.

1.1 Research question

This leads to the research question: What is the diagnostic accuracy and sensitivity of the Preclinical Alzheimer's Cognitive Composite (PACC) for detecting preclinical Alzheimer's Disease, compared to amyloid-PET imaging amongst people with elevated amyloid pathology?

1.2 Delimitation of the research object

In framing this systematic investigation, several decisions were made to ensure a focused exploration of the diagnostic accuracy of PACC. Firstly, the scope of the review will encompass all versions of the PACC test battery. This decision allows for a comprehensive assessment of various iterations of the PACC, thereby providing a nuanced understanding of its diagnostic utility.

Secondly, only studies featuring participants who have undergone assessments on two separate occasions will be included in the review, thus necessitating longitudinal studies. This deliberate focus enables an exploration of the PACC's ability to detect subtle cognitive changes associated with the development of amyloid pathology over time. In relation to this, this review will exclusively consider studies that utilize amyloid-PET imaging as the method for measuring amyloid pathology. This decision is based on PET imaging being a widely accepted gold standard for diagnosing AD, and this ensures consistency in the investigation. Consequently, even though amyloid detection via CSF is also a highly accurate measurement method, studies without additional amyloid-PET imaging will be excluded.

Likewise, the review will concentrate its focus on amyloid pathology and will not discuss assessments of tau pathology. This decision stems from the aim of targeting the earliest signs of preclinical AD, which often manifest primarily through amyloid accumulation during initial stages of the disease progression. Also, this helps to maintain a common ground in the literature review, where amyloid pathology remains the standard focus for all studies.

Furthermore, the exclusion of other biological factors contributing to AD, such as genetic predispositions (e.g., ApoE genotype), environmental factors, neuroinflammation, and oxidative stress, was implemented to maintain a clear and unambiguous focus on amyloid pathology as the primary biomarker of interest. While these factors undoubtedly play significant roles in AD pathogenesis and progression, their inclusion in this review would introduce additional complexity and heterogeneity, potentially confounding the analysis of PACC's diagnostic accuracy in relation to amyloid pathology. Additionally, confounding diseases or comorbidities are also being excluded from the review, seeing as they might influence the results of PACC scores and even the progression of AD. The review is focused on preclinical AD, which means that other types of dementia (e.g. Parkinson's Disease, frontotemporal dementia), psychiatric disorders (e.g. depression, schizophrenia), and significant medical comorbidities (cardiovascular diseases, diabetes) will be excluded.

Lastly, a particular emphasis will be placed on investigating the sensitivity of PACC. Recognizing sensitivity is a crucial metric in diagnostic accuracy assessments, and this focus seeks to clarify PACC's capacity to reliably detect cognitive changes indicative of preclinical AD in individuals with elevated amyloid pathology. So, all articles that do not introduce a discussion or any data regarding the sensitivity of PACC in their studies will not be included in this review.

2.0 Theory

The theory section provides a comprehensive overview of AD, covering its progression, biological mechanisms, and diagnostic methods including PET scan and cognitive composites. It explores cognitive decline and neuropsychological assessments, leading to an account of PACC for

early detection. Statistical ways to investigate sensitivity in a composite will also be presented. Lastly, existing literature to highlight the novelty and motivation behind this review will be presented.

2.1 Alzheimer's Disease

The International Classification of Diseases, 11th revision, (ICD-11), classifies AD under Mental, Behavioral or Neurodevelopmental Disorders, as a neurocognitive disorder (6D80), and is described as a disease with a slow but steady decline of cognitive functioning with impairment in domains such as memory, executive functions, attention, language, social cognition and judgement, psychomotor speed, visuoperceptual or visuospatial abilities (World Health Organization, 2019). The disease is currently ranked as the fifth leading cause of death worldwide (Ma et al., 2022). Late-onset AD typically manifests after the age of 65 and is the most common form of AD (Alzheimer's Association, 2024). Whereas, early-onset AD, also known as Early Onset Familial Alzheimer Disease (eFAD), is relatively rare (only 5% of all cases) (World Health Organization, 2019). EFAD has a strong genetic component, typically arising before the age of 65, due to inherited genetic mutations (World Health Organization, 2019). This review focuses only on late-onset AD, which allows for broader insights into the disease's etiology, risk factors, and potential interventions applicable to a larger portion of the population.

2.1.1 The progression of AD

AD can be divided into stages according to its progression: Preclinical AD, Mild Cognitive Impairment (MCI), and dementia due to AD (Alzheimer's Association, 2024), where the dementia stage can be further divided into mild, moderate, and severe dementia (World Health Organization, 2019). These stages often follow the progression of biomarker build-up in the brain, and The National Institute on Ageing and Alzheimer's Association (NIA-AA) proposed a classification system of the disease progression based on these biomarkers (Jack et al., 2018). They look at the disease as more

of a biological construct than a syndromal diagnosis, which takes pathological changes prior to cognitive symptoms into account (Jack et al., 2018). They divide the disease into a classification system of AD biomarkers: Amyloid, tau and neurodegeneration, or abbreviated, [AT(N)] (Jack et al., 2018), where the disease progression is determined by the amount of biomarker pathology. Cognitively normal (CN) individuals without any biomarkers are referred to as A-T-(N)-. Individuals with preclinical AD have one or two elevated biomarkers but no neurodegeneration, which becomes A+T-(N)-. People with AD and signs of dementia would be referred to as A+T+(N)+ (Jack et al., 2018). Here, the addition and subtraction symbols help determine the existence of biomarker pathology. This classification system helps distinguish the different diagnostic stages of the AD progression (Jack et al., 2018), and it was primarily developed to characterize research participants, and many studies use these classifications (Jack et al., 2018).

This system aligns well with the hypothesis of the biomarker cascade previously proposed by Jack and colleagues (2010), who created a model of the proposed biomarker progression in AD. This hypothetical model shows how the dynamic biomarkers of AD appear as a pathological cascade (Jack et al., 2010), where amyloid-beta (Aβ) build-up and the accumulation of neurofibrillary tau tangles (NFTs) in the brain are considered to be the first of a sequence of pathologic events, followed by symptoms of cognitive decline (Jack et al., 2010). However, it is crucial to understand that an elevated Aβ level is not a guarantee for AD development (Ossenkoppele et al., 2015). While Aβ accumulation is a key pathological feature of AD, not all individuals with high Aβ levels go on to develop the disease (Ossenkoppele et al., 2015).

The clinical disease stages include cognitively normal (CN), mild cognitive impairment (MCI) and dementia. And another, more commonly used, categorization method is the Clinical Dementia Rating scale (CDR), which is used both in research and in clinical settings to determine, how far

along participants are on the disease trajectory (0 = no signs of dementia, 1 = MCI, 2 = moderate dementia, 3 = severe dementia) (Berg, 1984).

2.2 Biological background of AD

Biomarkers are variables (physiological, biochemical, and anatomical variables) that can be measured in vivo and that can indicate specific features of disease-related pathological changes (Jack et al. 2010). In AD patients, the cerebral metabolism starts to decline, in contrast to healthy ageing, where the cerebral metabolism seems to be unaffected by age-related changes (Breedlove & Watson, 2013). The biomarker cascade proposed by Jack and colleagues (2010) indicates that as the levels of Aβ increase, they aggregate into structures known as plaques, often referred to as amyloid plaques or senile plaques (Breedlove & Watson, 2013). The accumulation of amyloid plaques interferes with the synaptic communication neuron-to-neuron, whereas the NFTs contribute to the neurodegeneration from within (Alzheimer's Association, 2024).

Aβ plaques are made up of amyloid proteins, and these proteins (Aβ peptides) are considered to be the main components in AD development (Ma et al., 2022). Aβ peptides are formed because of the cleavage of the Amyloid Precursor Protein (APP) by β-secretase and γ-secretase enzymes (Ma et al., 2022). In normal circumstances, Aβ peptides are cleared from the brain efficiently. However, in certain conditions, such as in AD, there is an imbalance between the production and clearance of Aβ peptides, leading to their accumulation and subsequent formation of Aβ plaques (Ma et al., 2022). The Aβ accumulation, particularly the peptide Aβ42, leads to the aggregation of insoluble oligomers (Ma et al., 2022), which refers to small clumps of Aβ molecules that can no longer be dissolved. This accumulation disrupts normal neuronal function and is associated with neurodegenerative diseases like AD, impacting cell-to-cell signaling and synaptic health (Ma et al., 2022).

NFTs consist of abnormal whorls of neurofilaments, including a protein called tau (Breedlove & Watson, 2013). Tau is a protein crucial for maintaining the stability of neural cytoskeletons (Ma et al., 2022), referring to the network that helps transport information between neurons. However, in AD, tau proteins become hyperphosphorylated, which disrupts the structure of a cell from within, eventually causing them to fall apart (Ma et al., 2022). The hyperphosphorylation of tau proteins causes them to become detached from their so-called microtubules, which is the part of the cytoskeleton, where they instead stick to other tau molecules, eventually forming threads or tangles (NIH, 2024). It is suggested, the hyperphosphorylation of tau proteins is caused by the presence of Aβ, which further underlines the biomarker cascade hypothesis (Jack et al., 2010).

In addition to amyloid and tau pathologies, AD progression involves other biological signs of decline. Cerebral blood flow (CBF) decrease is a pathological mechanism that similarly happens in the early stages of the disease (Ma et al., 2022). There are many consequences to less CBF, including cognitive impairment, neurological symptoms, and strokes (Ma et al., 2022). Another characterization of AD pathology is the significant decrease in acetylcholine (ACh) in the brain, because Aβ peptides reduces the synthesis of this neurotransmitter. ACh is a neurotransmitter whose primary function is to maintain consciousness, while it also plays a major role in learning and memory (Ma et al., 2022). Furthermore, the formation of Aβ plaques in the brain consist of aggregated Aβ as well as metal ions, which is a combination directly involved in reactive oxygen species (ROS) production, causing oxidative damage to cellular structures thus contributing to neurodegenerative processes (Ma et al., 2022).

2.2.1 How to measure biomarkers: PET

AD was initially defined as a clinical-pathologic entity, which was only definitely and fully diagnosed at autopsy, and in life it was more of a possible or probable AD (Jack et al., 2018). Today,

the gold standard for diagnosing AD is possible in vivo patients, by positron emission topography (PET) imaging (Young et al., 2020). PET is an imaging technique used to observe metabolic processes in the body (Young et al., 2020). By injecting a small amount of radioactive material (tracer) into the body, a PET scan can measure important functions such as blood flow, oxygen use, and glucose metabolism, providing images and data that reflect how tissues and organs are functioning (Young et al., 2020). These technological advances have made it possible to detect the earliest pathological changes in the asymptomatic preclinical stage of AD, where abnormal biomarker levels are detected (Jack et al., 2018; Sperling et al., 2011). The injected tracer is made to mimic a certain protein, for instance a form of glucose (^{18}F-FDG or Fluorodeoxyglucose), which then travels through the bloodstream and is absorbed by the body's tissues (Young et al., 2020). ^{18}F-FDG was first introduced as a tracer for diagnosing dementia in 1979, and it is useful to get an overview of the brain's overall activity, as glucose is the brain's main source of energy (Young et al., 2020). As a tracer is slightly radioactive, it emits positrons as it decays, and as positrons pass through matter (in the body), they experience the same interactions as electrons, including loss of energy through ionization and excitation of nearby atoms and molecules (Turkington, 2001). The collision between a positron and an electron is called an annihilation, and this is what causes gamma rays, which are then detected by the PET scanner (Turkington, 2001; Young et al., 2020). Different radionuclides exist, which refers to the radioactive component in the injected tracer, and the most used PET radionuclides are the ^{11}C and ^{18}F, because of their short half-life: 20,3 minutes and 110 minutes, respectively (Turkington, 2001).

However, FDG-PET has its limitations, and it is not specific for AD indicators. To detect AD-related hallmarks more specifically, the use of Aβ-specific PET tracers, or ligands, has been developed (Young et al., 2020). Ligands are specific types of tracers that bind to particular proteins or receptors in the brain, for instance Aβ proteins or NFTs (Young et al., 2020). The first Aβ-PET tracer to be

introduced was the ^{11}C-Pittsburgh compound B (PiB), which has a high affinity for Aβ, enabling it to bind to any existing Aβ in the brain (Breedlove & Watson, 2013).

Besides ^{11}C-PiB-PET, the ^{18}F-FLUTE (Flutemetamol) and ^{18}F-FBP (Florbetapir) are similarly PET ligands that target Aβ pathology (Young et al., 2020). These are tagged with radioactive isotopes that does not require a local cyclotron to produce the tracers, as the half-life time is a bit longer (Young et al., 2020). The binding between the tracers and existing Aβ pathology can be seen, as they light up on a screen showing the PET-images from the Aβ-PET scan. Standard Uptake Value Ratio (SUVR) is a quantitative measure used in PET imaging to evaluate the concentration of a tracer within a specific region of the brain relative to a reference region (Young et al., 2020), and this method of measurement is widely used both in clinics and in research. However, the SUVR is usually calculated for each patient, where the regions of interest (ROIs) are divided by a reference region assumed to be free of Aβ-pathology (Young et al., 2020), making this method very specific to a particular individual. This led to the development of a standardized quantitative Aβ imaging measurement using the centiloid (CL) scale (Young et al., 2020). To use the CL scale, imaging data from any Aβ-PET tracer are converted to the CL scale using established conversion factors, which makes it possible for researchers and clinicians to track the progression of Aβ accumulation more accurately in the brain, compare the effectiveness of anti-amyloid therapies, and improve the diagnostic consistency of AD (Young et al., 2020). Today, both measurements can be used in research, as either the mean-CL scale or the mean-SUVR (Young et al., 2020). Sometimes the Distribution Volume Ratio (DVR) is used instead of SUVR to get an even more accurate reflection of the binding potential of the tracer, but it usually requires a blood sample and is a more complex procedure overall (Young et al., 2020).

2.2.2 Brain areas affected by biomarkers

Aβ-PET SUVR is a summary of uptake in the frontal, cingulate, temporal, and parietal regions relative to the cerebellum (Young et al., 2020), as the cerebellum is typically less affected by Aβ deposition in the early stages of AD. The initial stages of Aβ deposition occurs in the association cortices, particularly the frontal, parietal, temporal, and occipital lobes (Koychev et al., 2020). From the association cortices, Aβ spreads to the limbic regions, which includes the entorhinal cortex, hippocampus, and amygdala (Koychev et al., 2020), and these areas are strongly associated with the earliest symptoms of cognitive impairment. As AD progresses, areas in the frontal and temporal lobes begin to show significant Aβ accumulation. These areas are involved in higher cognitive functions and language processing (Koychev et al., 2020), which can lead to impaired judgement and memory loss. In advanced stages of AD, Aβ reaches the parietal lobe, impairing spatial awareness and navigation (Koychev et al., 2020). Lastly, in the later stages of AD, Aβ reaches the occipital lobe responsible for visual processing, which can cause visual disturbances and difficulties in object and facial recognition (Koychev et al., 2020).

2.3 Cognitive decline in AD

As the biomarkers accumulate in the brain, symptoms of cognitive decline progress. However, dementia is a general term used for various symptoms of memory loss, cognitive decline, trouble in communication, and many other related brain complications in performing routine tasks (Ahmad et al., 2023). The development of the biomarkers, Aβ and NFTs, causes a loss in the connection between neurons and results in breaking the transmission of messages between brain regions (Ahmad et al., 2023). Every cluster of neurons has a specific task to perform, for instance thinking, remembering, vision, listening, language (Ahmad et al., 2023). Furthermore, AD is known to cause shrinkage of various brain regions especially associated with thinking, memory, planning, and decision making

(Ahmad et al., 2023). The progression and permanent impairments in the brain underscores the critical need for accurate and timely assessment of cognitive functions. Neuropsychological assessments play an important role in this regard, serving as tools for detecting and monitoring cognitive changes over time.

2.3.1 The Preclinical Alzheimer's Cognitive Composite (PACC)

A study that has aimed for developing tests sensitive to the earliest AD-related changes is the study by Donohue and colleagues (2014), called the Alzheimer's Disease Cooperative Study (ADCS). They suggest the cognitive composite, called the Preclinical Alzheimer's Cognitive Composite (or originally ADCS-PACC) (Donohue et al., 2014). PACC is specifically designed to detect subtle cognitive changes in individuals at risk of developing AD, even before onset of dementia symptoms (Donohue et al., 2014). The composite comprises tests that assess episodic memory, execute function, and processing speed, which are known to be early indicators of cognitive decline in AD (Donohue et al., 2014). The composite was designed to be the primary outcome measure for the Anti-Amyloid Treatment in Asymptomatic Alzheimer's Study (A4 Study), with a specific focus on detecting cognitive changes related to amyloid pathology (Donohue et al., 2014). As the A4 Study was still in early planning stages at the time, the ADCS gathered data from two existing cohorts to test the sensitivity of the ADCS-PACC first: The Alzheimer's Disease Neuroimaging Initiative (ADNI) and The Australian Imaging, Biomarkers, and Lifestyle Study of Ageing (AIBL) (Donohue et al., 2014). All participants were cognitively normal (CN) individuals with no cognitive symptoms, but with a slightly elevated Aβ pathology, determined by an SUVR of 1.5 and above as measured by an ^{11}C-PiB-PET scan (Donohue et al., 2014). They found that Aβ-positive participants showed a significant decline in PACC scores at both 18-, 24-, and 36-month follow-up measurements compared to Aβ-

negative participants (Donohue et al., 2014). This indicated a sensitive relation between PACC scores and Aβ pathology.

The composite is comprised of four tests: The Total Recall score from the Free and Cued Selective Reminding Test (FCSRT), the Delayed Recall score on the Logical Memory II subtest (LM-DR II) from the Wechsler Memory Scale (WMS), the Digit Symbol Substitution Test (DSST) score from Wechsler Adult Intelligence Scale-Revised (WAIS-R), and the Mini Mental State Examination (MMSE) total score (Donohue et al., 2014).

The FCSRT measures verbal episodic memory under conditions that control both attention and cognitive processing (Grober et al., 2008). FCSR starts with a study phase where subjects identify items pictured on a card by matching them to category cues (e.g., matching "grapes" to "fruit") (Grober et al., 2008). After identifying all items, immediate recall of these items is tested (Grober et al., 2008). Items not recalled are re-searched until all 16 items are retrieved (Grober et al., 2008). This is followed by three recall trials, each involving free recall and then cued recall for any missed items (Grober et al., 2008). The combined score of free and cued recall in each trial is termed total recall (Grober et al., 2008).

The LM-DR II from WMS measures episodic memory by assessing encoding, storage, and recall processes (Ahn et al., 2019). The LM-DR consists of three parts in total: LM I (immediate recall), LM II (delayed recall), and LM R (delayed recognition) (Ahn et al., 2019). In LM I, subjects immediately recall details of two short story passages, whereas in LM II, they recall the passages after a 20–30-minute delay (Ahn et al., 2019). Lastly, in LM R, subjects answer yes/no questions about the passages to see how much of the stories they recognize (Ahn et al., 2019). However, in ADCS-PACC only the delayed recall score from LM II is used when calculating the total PACC score (Donohue et al., 2014).

The DSST measures motor speed, attention, and visuoperceptual functions (Jaeger, 2018). The DSST is a paper-and-pencil cognitive test that requires subjects to match symbols to numbers based on a key provided at the top of the page (Jaeger, 2018). The subject then copies the corresponding symbol into spaces below a row of numbers (Jaeger, 2018). The test score is determined by the number of correct symbols matched within the allotted time, typically 90 to 120 seconds (Jaeger, 2018).

The MMSE is a global cognition test, with a total of 30 untimed questions and therefore the possibility to score 30/30 correctly (Folstein et al., 1975). It provides a useful quantified measure of the current cognitive state (Folstein et al., 1975). The MMSE is composed of two parts: The first section focuses solely on vocal responses, assessing orientation, memory, and attention, with a maximum score of 21 (Folstein et al., 1975). The second section evaluates the ability to name objects, follow verbal and written instructions, write a sentence spontaneously, and copy a complex polygon akin to a Bender-Gestalt Figure (a geometrical figure), with a maximum score of 9 (Folstein et al., 1975). Additionally, this test should not take longer than 10 minutes to complete (Folstein et al., 1975), which can help reduce fatigue and lower stress in elderly, potentially demented, participants.

2.4 Reliability, validity, and sensitivity

The PACC composite score is determined from its components using an established normalization method (Donohue et al., 2014). Each of the 4 component scores is divided by the baseline sample standard deviation of that component, to form standardized z-scores, as a z-score is a standardized raw score (Ivanouw, 2006).

Many types of errors can occur, and will occur, when conducting research. But it is important to be aware of this, and how they can affect the overall reliability, validity, and sensitivity of a test (Ivanouw, 2006). This is why it is important to include an error score, where some are positive (overestimations), and some are negative (underestimations) (Ivanouw, 2006). Reliability is the

variance in true scores divided by the variance in observed scores, where high reliability refers to a low variance between the two, and a low reliability refers to a big variance between the two (Ivanouw, 2006). Some errors include random response errors (where participants for instance randomly choose "a" instead of "b"), specific errors (when they misinterpret the question), and transient errors (when their response depends on their mood that day) (Ivanouw, 2006), and both the participant and the clinician can make these errors. Additionally, the clinician can make idiosyncratic errors, where they let their own judgements affect the interpretation of the responses (Ivanouw, 2006). Reliability can be investigated in various ways and corrected accordingly in different statistical ways (Ivanouw, 2006). But it can also be achieved by comparing the scores of interests with a gold standard, which might even speak more to the tests' accuracy than its reliability (Ivanouw, 2006).

Validity concerns whether a particular examination method provides relevant information about the characteristics it is designed to measure (Ivanouw, 2006). It is important to investigate and consider, whether the test measures more or less than the concept, it is intended to measure (Ivanouw, 2006). When validating and evaluating tests for use in a test battery, one should remember to investigate the extent to which the tests overlap, and the extent to which they can complement each other in the process (Ivanouw, 2006). It should be assessed whether the tests in fact support or contradict each other (Ivanouw, 2006). The sensitivity determines, how good the test is at correctly identifying patients who have the disease in question (hence, avoiding false negative results), whereas the specificity determines, how good the test is at excluding patients who do not have the disease in question (avoiding false positive results) (Ivanouw, 2006). However, it can be difficult to determine both the sensitivity and specificity, and therefore it can also be valuable to look at the overall predictive value (Ivanouw, 2006).

Lastly, avoiding Type II and Type II errors is crucial in research. Type I errors happen when the expected effect is found significant, when there is no effect, while Type II errors occur when the expected effect is not found significant, when in reality it is (Šimundić, 2013).

2.4.1 Bias in research

Another important principle within research is to be transparent about potential biases that can and will occur. All scientific papers need to clarify any concerns they might have, when it comes to the quality of work submitted for publication (Šimundić, 2013). All reviews aim to avoid bias (Booth et al. 2022). The more systematic they are, the less likely it is for bias to appear (Booth et al. 2022). Working systematically by following validated guidelines can help with this (Booth et al. 2022).

Bias can occur both intentionally and unintentionally, and they can contribute to misleading results (Šimundić, 2013). Selection bias can occur, when the participants included in a study are not representative of the general population, leading to skewed results that cannot be accurately generalized (Šimundić, 2013). Volunteer bias occurs when individuals who choose to participate in a study might differ from those who do not, potentially affecting the study's outcomes (Šimundić, 2013). Confirmation bias refers to the tendency to seek, interpret, and remember information that confirms preexisting beliefs while ignoring contradicting evidence (Peters, 2022). Attrition bias occurs when participants drop out of a study and does not complete the follow-up assessments, which can affect the study's results (Babic et al., 2019). Recall bias, a form of retrospective bias, occurs when participants' recollections of past experiences are inaccurate or incomplete, potentially leading to faulty conclusions (Ingram, 2023).

2.4.2 How to measure test sensitivity

There are many ways to measure test sensitivity. It can be challenging to summarize the results from studies of different quality, maybe because different methods have been used, or the sample

sizes are different, and then they consequently get different effect sizes (Ivanouw, 2006). In this way, variation in effect sizes is considered to reflect the different influencing variables within each study (Ivanouw, 2006). However, there are alternative ways to measure the effect and sensitivity of a study (Ivanouw, 2006), for instance, beta-coefficients, likelihood-ratios, hazard ratios (HRs), Receiver Operating Characteristics (ROC) analysis, odds-ratios, predictive values, logistic mixed-effects models, linear mixed-effects models, and confirmatory factor analysis (CFA) (Field, 2017).

The beta-coefficient (β-coefficient) is a measurement method that is valuable in detecting the strength of a study, as well as its direction (Field, 2017). The β-coefficient is the average amount that the dependent value increases, when one independent variable increases one standard deviation, (Field, 2017), and is therefore used primarily when comparing variables. A positive β-coefficient indicates that if X increases, Y similarly increases (Field, 2017), whereas if the β-coefficient is negative, it means that if X increases, Y decreases. This can be useful in detecting strength and direction of possible correlations (Field, 2017).

The hazard ratio (HR) is the probability of a particular event occurring in a group compared to a control group over time (Spruance et al., 2004). An HR above 1 suggests an increased risk of disease, where an HZ below 1 suggests a smaller risk (Spruance et al., 2004).

The likelihood ratio is based on the maximum-likelihood theory, which is the idea of finding the best model to explain your data (Field, 2017). The likelihood-ratio can help evaluate two aspects: The potential effectiveness of a specific diagnostic test, and the probability that a patient has a particular disorder or condition (Field, 2017). A higher positive likelihood-ratio indicates that a positive test result is much more likely to be measured in people with the condition compared to those without it (Field, 2017). A higher ratio means the test is better at confirming the presence of a condition (Field, 2017). A slightly similar measurement method is the odds-ratio, which can help compare the odds of

outcome between groups (Field, 2017). The odds-ratio measures effect size by quantifying the relationship between variables (Field, 2017), where a higher ratio indicates a stronger association between exposure and outcome, and a lower ratio indicates a weaker association.

The ROC analysis is used to evaluate the performance of diagnostic tests (Metz, 1978). It plots the true positive rate (sensitivity) against the false positive rate (1 minus specificity) at various threshold settings (Metz, 1978). This method creates a curve based on the specificity and sensitivity scores from the test, where the area under the curve (AUC) can determine how accurate the test is (Metz, 1978). A higher AUC-score indicates that the test has better accuracy in distinguishing between those with and without the condition (Metz, 1978).

A predictive value measures how well a test predicts the presence or absence of a condition (Field, 2017). The positive predictive value (PPV) indicates the probability that subjects with a positive test truly have the condition, whereas the negative predictive value (NPV) indicates the probability that subjects with a negative test truly do not have the condition (Field, 2017).

Logistic mixed-effects models and linear mixed-effects models (both LMM) are methods used to understand data where we collect multiple measurements from the same subjects (Bartlett, 2020). The logistic model helps us analyse outcomes, while considering individual differences, and the linear model looks at continuous outcomes (like test scores) in the same way (Bartlett, 2020). Additionally, the mixed method of repeated measures (MMRM) can help determine how things change over time within the same people, using both fixed effects (overall trends) and random effects (individual differences) (Bartlett, 2020).

Lastly, CFA can help check how well a test item (for instance a subtest) match the skills they are supposed to measure (cognitive domains). If the CFA results show strong connections and good

overall fit, it means the subtests are doing a good job measuring what they are supposed to measure (Field, 2017).

2.5 Motivation behind this review

The motivation for undertaking a systematic review on the early detection of AD is multifaceted, rooted in the escalating need to identify and manage this condition as early as possible. As the global population ages, the prevalence of AD is rising, making the early detection of the disease more critical than ever. This section outlines the key motivations driving this research, synthesizing various factors that underscore its importance.

Firstly, the necessity for detecting AD has never been more pressing. With an increasing number of individuals reaching old age, the incidence of AD continues to climb (Alzheimer's Association, 2024). Early detection is crucial to prevent permanent brain damage and to take advantage of emerging medical treatments designed to slow disease progression. Advances in medicine, such as new drugs, underscore the importance of diagnosing AD at the earliest possible stage (Grant, 2023). However, the early cognitive decline associated with AD is often difficult to differentiate from normal aging, making it challenging to identify and diagnose. Moreover, cognitive changes are sometimes stigmatized, and only 40% of individuals who experience memory problems consult a doctor (Alzheimer & Association, 2024). This highlights a significant gap in early diagnosis and intervention. Additionally, cognitive changes can sometimes be attributed to other treatable conditions, such as tumours or haemorrhages, which need timely medical attention (Mayo Clinic Staff, 2024). An early diagnosis would also allow patients and their families to plan for the future, make their homes safer, and manage expectations effectively (Mayo Clinic Staff, 2024), and to facilitate informed decision-making and help maintain the patient's independence, health, and safety. Furthermore, early-stage diagnosis enables the prescription of drugs that can slow cognitive decline, such as acetylcholinesterase inhibitors, which are vital for maintaining communication between brain

cells (National Health Service, 2023). Most notably, the recent approval by the Food and Drug Administration (FDA) of lecanemab, a treatment for AD, marks a significant advancement (Grant, 2023). Lecanemab has been shown to moderately slow the progression of cognitive and functional decline in the early stages of the disease, reducing cognitive decline by 27% after 18 months compared to a placebo (Grant, 2023). This drug also demonstrated reductions in both Clinical Dementia Rating (CDR) and amyloid proteins, emphasizing the importance of early detection and intervention (Grant, 2023).

Another breakthrough in the early detection of AD was recently achieved by Ashton and colleagues, in January 2024. They developed a blood test capable of detecting p-tau217, a specific form of the tau protein that is an important AD biomarker (Ashton et al., 2024). This approach shows promising abilities to provide quick and accurate diagnoses much earlier in the disease progression, ultimately enabling timely interventions and significantly improving patient outcomes (Ashton et al., 2024).

Despite these advancements, there is still a pressing need for more sensitive tests for preclinical AD to be implemented in clinical settings (Jack et al., 2018). Current neuropsychological tests, such as the MMSE and the Montreal Cognitive Assessment (MoCA), are commonly used to evaluate cognitive domains (Folstein et al., 1975). However, these tests have limitations**.** For example, the MMSE has shown poor test-retest reliability, ceiling effects, and low sensitivity to impairment in cognitively normal older adults (Bransby et al., 2019). The MMSE is also criticized for its insensitivity to detect MCI and early stages of AD, particularly in detecting subtle cognitive deficits in executive function and visuospatial skills (Jia et al., 2021). Mitchell conducted a meta-analysis on the accuracy of MMSE in relation to detecting dementia and MCI (2009). He investigated 39 studies and found, the MMSE is best for confirming dementia in specialist settings and for ruling out dementia in non-specialist settings (Mitchell, 2009). In settings at memory clinics, MMSE had a

pooled sensitivity of 79.8%, a specificity score of 81.3%, and a PPV of 86.3%, which overall indicates the test is a reliable tool in these settings (Mitchell, 2009). In non-clinical community settings, the MMSE had a pooled sensitivity of 85.1%, and a specificity score of 85.5%, and while these scores are high, the PPV value was only 34.5%, which suggests that the MMSE is less reliable for correctly confirming dementia, leading to more false positives (Mitchell, 2009). This indicates that while the MMSE can effectively rule out dementia in these settings, it may not be as useful for confirming a diagnosis without additional testing (Mitchell, 2009). This was also the case in primary care, where similar issues with sensitivity was measured (Mitchell, 2009).

The MoCA, although more sensitive than the MMSE for detecting MCI, also has its limitations. Its sensitivity can vary depending on the population and specific cognitive domains assessed, and it may be too challenging for individuals with more severe cognitive impairment (Jia et al., 2021). As Jia and colleagues (2021) conducted a cross-sectional analysis of 4923 adults aged ≥55 years, they found that the MMSE and the MoCA both were able to identify MCI with varying degrees of prevalence. The MMSE had a prevalence rate of 28.6% for MCI, while the MoCA identified a higher rate of 36.2% (Jia et al., 2021). The study also noted that the MoCA had better sensitivity for detecting MCI, with less ceiling effect and greater detection of cognitive heterogeneity compared to the MMSE (Jia et al., 2021). This indicates that the MoCA is a more effective tool for screening cognitive impairment in this population (Jia et al., 2021).

Recent research suggests using composite cognitive measures to gain more nuanced and sensitive results, which is specifically designed to detect Aβ-related cognitive decline (Hassenstab et al., 2021). Composites are considered more sensitive and specific for early detection of AD compared to MMSE and MoCA (Hassenstab et al., 2021). Hassenstab and colleagues (2021) conducted a study comparing the efficacy of short-form cognitive assessments, like MoCA, with comprehensive long-form cognitive batteries in predicting AD progression and their sensitivity to AD neuroimaging

biomarkers (Hassenstab et al., 2021). In a sensitivity analysis, they found that both MoCA and a Global Cognitive Composite (GCC) show modest sensitivity (AUC = 0.64 for MoCA, and AUC = 0.66 for GCC) (Hassenstab et al., 2021). These results underscore the limitations of current tests and a pressing need to more sensitive and specific tools (Hassenstab et al., 2021).

PACC is one such promising tool specifically designed to detect the earliest signs of cognitive decline due to AD (Donohue et al., 2014). However, there is an overall lack of studies investigating the sensitivity and accuracy of PACC on its own. The need for accessible and reliable tests in clinical settings is evident. Investigating the sensitivity, accuracy, and specificity of PACC could determine its feasibility as a standalone measurement tool in clinical practice. Although PACC may not stand on its own in all diagnostic settings, exploring its potential for implementation into clinical practice is essential. An investigation of its sensitivity independently from other measurements can be useful, as this has not been thoroughly examined previously.

In summary, the motivations for conducting a systematic review on the early detection of preclinical AD are driven by the urgent need to improve diagnosis methods to utilize new treatments and enhance patient outcomes. This review aims to synthesize existing research to gather relevant information on PACC in relation to early disease detection, which in the end can contribute to the ongoing efforts to combat this prevalent disease.

3.0 Methodology

This systematic review is based on an interest in finding the existing connections between the understandings of Aβ pathology in the brain and the progression of cognitive decline in preclinical AD patients. The following section will account for the theory of science and the methodological considerations that establish the foundation for this review. This section will outline how the review follows the 'Diagnostic Test Accuracy' (DTA) protocol and the Preferred Reporting Items for

Systematic Reviews and Meta-Analyses (PRISMA) guidelines for systematic reviews, as established by the Joanna Briggs Institute (JBI). Additionally, the criteria for study eligibility will be outlined, followed by a presentation of the search strategy, including the search strings used in various databases. An overview of the study selection process will also be provided, giving a clear picture of how studies were chosen for inclusion in this systematic review.

3.1 Theory of science

In all science, it is important to emphasize that truth is never something one can achieve complete certainty about (Vengsgaard, 2015). Rather, it is a goal one can approach through critical examination and falsification of hypotheses (Vengsgaard, 2015). This review is grounded in critical rationalism, as the approach of investigating and revising existing hypotheses is done by critically and systematically evaluating and synthesizing the strongest findings (Vengsgaard, 2015). However, in all research it is important to be transparent about the fact, that one can never be entirely objective, as humans naturally are influenced by their assumptions, background, and beliefs (Andersen & Kock, 2015). Especially when a project is based on an underlying interest and potentially prior experience with the subject. It is therefore important to be aware of one's own role in the decisions that may be made during the entire writing process (Anderson & Kock, 2015).

3.2 Study type - A systematic review

A systematic review aims to explore the claims regarding the effectiveness of interventions (Petticrew & Roberts, 2008). A review can help map out areas of uncertainty by comprehensively identifying, appraising, and synthesizing all the relevant research on a specific topic (Petticrew & Roberts, 2008). The demand for evidence-based clinical practice is steadily increasing due to ongoing advancements that expand the array of technologies, medications, and treatments accessible to patients (Aromataris & Pearson, 2014). And a systematic review is precisely based on all accessible

evidence and its data to uncover any patterns, trends, insights across multiple studies (Aromataris & Pearson, 2014). However, it is worth emphasizing that the execution and quality of a systematic review largely depend on the quality of the included studies, thus a systematic review is never better than the studies it includes (Moher et al., 2009). A systematic review is the synthesis of the best available evidence aimed at answering a specific question (Perestelo-Pérez, 2013). During a systematic review, scientific research strategies and guidelines are applied to minimize any existing bias (Perestelo-Pérez, 2013).

3.2.1 Diagnostic test accuracy

The international evidence-based healthcare research organization, JBI, has created a Manual for Evidence Synthesis (Aromataris & Munn, 2020), that is designed to provide guidelines for systematic reviews. In this review, the primary emphasis is on evaluating the sensitivity and accuracy of PACC. Hence, it follows the guidelines outlined in Chapter 9 of the JBI Manual, titled 'Diagnostic Test Accuracy,' and this will be the basis of this review (Aromataris & Munn, 2020; Campbell et al., 2020). As the JBI website has since been updated, thus changing some of the chapters (including Chapter 9), this chapter will be added as an attachment with permission from JBI (See Attachment 1).

Primary studies that investigate the accuracy of diagnostic tests are referred to as diagnostic test accuracy (DTA) studies, and this review will specifically focus on systematically reviewing studies that take test accuracy into account (Campbell et al., 2020). DTA studies compare a diagnostic test of interest (referred to as the ‘index test’) to an established diagnostic test (referred to as the ‘reference test’), which should be the best test available for accurately identifying the presence or absence of the condition of interest (Campbell et al., 2020). In this review, the index test refers to PACC, whereas the reference test refers to amyloid-PET imaging, as it is the current gold standard. A systematic

review of diagnostic test accuracy synthesizes all available evidence to estimate the accuracy of the test in question (Campbell et al., 2020). This can be done in multiple ways. Measures such as sensitivity, specificity, but also predictive values, odds-ratios, likelihood ratios, and ROC analyses are all measurements of accuracy (Campbell et al., 2020).

3.2.2 PRISMA

This study additionally follows the PRISMA guidelines, which are designed to help reviewers maintain a transparent approach when conducting a systematic review (Page et al., 2021). Following a validated checklist also allows for readers of the review to assess the strengths and weaknesses of the investigation (Liberati et al., 2009). The guidelines ensures both quality and consistency throughout the reviewing process, while also allowing other researchers to achieve the same results if they were to apply the same search strategy at the same time (Moher et al., 2009; Page et al., 2021). The PRISMA guidelines, which are based on the PRISMA Statement from 2009 (Page et al., 2021) consist of a 27-items checklist and a four-phase flow diagram. The checklist provides reviewers with a replicable review strategy, which can enhance the reliability of the review (Page et al., 2021). Additionally, the flow diagram provides an overview of the review process, indicating the number of studies excluded and included at various stages of the search (Page et al., 2021). The PRISMA 2020 statement is primarily intended for systematic reviews that assess the effects of health interventions, and the items on the PRISMA checklist are therefore designed to report and evaluate these effects (Page et al., 2021). This review will utilize this checklist as its foundation. However, while the PRISMA checklist provides a comprehensive framework for reporting systematic reviews, not all items may be applicable in this review. For instance, specifying effect measures, risk ratios and mean scores for each study might be a challenge, since not all studies provide these data in a standardized format or may report them in varying ways. Additionally, due to time constraints and the current scope

of the review, a review protocol for this systematic review will not be conducted. For the same reasons, registration of this review prior to its commencement has also not been done. (See the full PRISMA checklist in Attachment 2).

3.3 Study characteristics

To formulate a relevant and precise research question, researchers often utilize the PICOS approach, which encompasses five key components represented by the acronym: Patient/Population, Intervention, Comparison, Outcome, and Study design (Leonardo, 2018; Liberati et al., 2009). A well-structured investigation should incorporate details about all five components (Leonardo, 2018), in both the title and the research question. However, according to Campbell and colleagues (2020), when conducting a systematic review of diagnostic test accuracy, there are additional things to consider. Firstly, the title must explicitly include the phrase "a systematic review protocol" (Campbell et al., 2020). And the title must additionally include each of the elements of the PIRD acronym: Population, Index test, Reference test, and Diagnosis of interest (Campbell et al., 2020), which concludes the title: "Exploring the accuracy of the Preclinical Alzheimer's Cognitive Composite (PACC) relative to amyloid pathology for the diagnosis of Alzheimer's Disease: A systematic review protocol".

3.3.1 Table 1: PICOS

A PICOS chart has been made for this review and can be seen below in Table 1:

Participants	Participants / patients with elevated amyloid pathology Cognitively normal (CN) individuals Participants in any preclinical AD stage
Intervention	Amyloid-PET has been used to determine amyloid pathology PACC has been used to determine cognitive decline

Comparison	Participants with elevated amyloid pathology and a CN control group
Outcome	The outcome revolves around detecting any cognitive changes in participants with elevated amyloid pathology, contrary to CN individuals. Outcome must include both a change of amyloid build-up as well as a change in PACC score in the follow-up study
Study design	Study types with n ≥ 10 participants Longitudinal studies with 2 or more measurements of PACC scores Data on sensitivity and/or accuracy of PACC

Table 1: PICOS

3.4 Study eligibility

Studies meeting specific criteria will be included in the review. The criteria are based on the PICOS and PIRD approaches, as well as the research question and the intervention of interest. Studies published between 2014 and 2024 are included, as PACC was not constructed until 2014 (Donohue et al., 2014).

The focus on participants along the AD continuum, including cognitively normal adults with abnormal Aβ-levels, is of interest in this review due to the current investigation of trying to better understand the earliest stages of the disease and identifying individuals who may be at risk of developing Alzheimer's-related cognitive decline. Accordingly, this review does not focus on animal studies, abnormal Aβ pathology resulting from other known causes, or individuals already diagnosed with AD. Furthermore, studies where participants' cognitive status relied on self-administered assessments, such as questionnaires or their own subjective cognitive decline (SCD), instead of Aβ-PET imaging, were not included in this review. Exclusions also apply to participants undergoing cognitive training, dietary restrictions, or specific physical training programs, as well as those

involved in placebo-drug trials. This last decision is based on the potential manipulation of results at post-intervention, which may not reflect the natural progression of the disease, hence potentially influencing the PACC scores.

The utilization of PACC to measure cognition is necessary to conduct a diagnostic test accuracy systematic review on the intervention method. Additionally, studies included must contain available data on the sensitivity and/or accuracy of PACC.

The focus on Aβ-PET imaging stems from its status as the current gold standard for AD diagnosis. Consequently, articles that do not specify the type of imaging used are excluded from consideration. Also, Aβ pathology determined by other means, such as CSF or plasma analysis, does not align with the focus of this review. Lastly, included studies must be peer-reviewed, as this improves the validity of the review.

3.4.1 Table 2: Inclusion and exclusion criteria

Inclusion criteria	• Publications between 2014-2024 • Amyloid-PET to determine amyloid pathology • PACC test battery to determine cognition • Patients along the AD continuum (abnormal amyloid pathology) • Peer-reviewed publications • Longitudinal study designs • Data on PACC sensitivity/accuracy
Exclusion criteria	• Patients with major comorbid diseases • Duplicate publications • No full text available • Other languages than English, Danish, Swedish and Norwegian • Studies with n < 10 participants • Reviews, books, conference abstracts, theses, posters, manuscripts

Table 2: Eligibility criteria

3.5 The search

A systematic search was conducted across three academic databases, namely PsycInfo, PubMed, and Embase on March 20th, 2024. The search strategy covered the terms “Alzheimer’s Disease” or “amyloid”, and “Preclinical Alzheimer’s Cognitive Composite”, and was not limited to title, abstract, and keywords. The search was instead made in “any field” or for the terms to appear anywhere in the full text as well as in the index terms for each database, encompassing APA index terms, MeSH, and Emtree. Furthermore, the search was focused on studies published between 2014 and 2024 to gather all relevant information since the introduction of PACC.

The search was divided into two parts, where one part focused on all articles containing either “amyloid” or “Alzheimer’s Disease” and the corresponding index terms for these, whereas the other part focused on all articles containing “Preclinical Alzheimer’s Cognitive Composite” or “PACC” as this is the focus of the review. Boolean operators such as “AND”, “OR”, and “*” were used to organize the search (see an example in Figure 1). The search was done in collaboration with librarians at Aalborg University Library (AUB) to ensure the quality of the search. (See the whole search in Attachment 3).

3.5.1 Figure 1: Example of a search in PubMed

Search	Actions	Details	Query	Results	Time
#10	•••	›	Search: **(("Preclinical Alzheimer's Cognitive Composite") OR (PACC)) AND (((("Alzheimer Disease"[Mesh] OR "Amyloid beta-Peptides" [Mesh]) OR ("Amyloid"[Mesh] OR "Plaque, Amyloid"[Mesh] OR "Amyloid Precursor Protein Secretases"[Mesh] OR "Amyloid beta-Protein Precursor"[Mesh])) OR (amyloid)) OR (Amyloid*)) OR (Alzheimer*))** Filters: **from 2014 - 2024**	99	08:26:19
#9	•••	›	Search: **(("Preclinical Alzheimer's Cognitive Composite") OR (PACC)) AND (((("Alzheimer Disease"[Mesh] OR "Amyloid beta-Peptides" [Mesh]) OR ("Amyloid"[Mesh] OR "Plaque, Amyloid"[Mesh] OR "Amyloid Precursor Protein Secretases"[Mesh] OR "Amyloid beta-Protein Precursor"[Mesh])) OR (amyloid)) OR (Amyloid*)) OR (Alzheimer*))**	100	08:26:11
#8	•••	›	Search: **("Preclinical Alzheimer's Cognitive Composite") OR (PACC)**	888	08:23:33
#7	•••	›	Search: **((("Alzheimer Disease"[Mesh] OR "Amyloid beta-Peptides" [Mesh]) OR ("Amyloid"[Mesh] OR "Plaque, Amyloid"[Mesh] OR "Amyloid Precursor Protein Secretases"[Mesh] OR "Amyloid beta-Protein Precursor"[Mesh])) OR (amyloid)) OR (Amyloid*)) OR (Alzheimer*)**	300,108	08:23:09
#6	•••	›	Search: **("Alzheimer Disease"[Mesh] OR "Amyloid beta-Peptides" [Mesh]) OR ("Amyloid"[Mesh] OR "Plaque, Amyloid"[Mesh] OR "Amyloid Precursor Protein Secretases"[Mesh] OR "Amyloid beta-Protein Precursor"[Mesh])** Sort by: **Most Recent**	167,493	08:21:57
#5	•••	›	Search: **amyloid**	124,032	08:18:20
#4	•••	›	Search: **"Preclinical Alzheimer's Cognitive Composite"**	41	08:17:41
#3	•••	›	Search: **PACC**	864	08:17:12
#2	•••	›	Search: **Amyloid***	142,442	08:17:04
#1	•••	›	Search: **Alzheimer***	228,262	08:16:54

Figure 1: Example from a search in PubMed

3.6 Study selection and screening process

The selection process utilized the online tool, Rayyan, which supports users in conducting systematic reviews (Ouzzani et al. 2016). The initial search resulted in a total of 392 articles from all three databases before screening. Firstly, duplicates were manually excluded, as Rayyan had automatically identified 230 possible duplicates. This resulted in the exclusion of 127 articles. Afterwards, the remaining 265 references were screened based only on abstract and title according to the eligibility criteria. This screening resulted in 171 articles. A second screening included a more

thorough look into the methodology section of each article to investigate, whether they provided any information and data regarding the sensitivity and specificity of PACC, as this was the focal point of this review. This screening resulted in 30 articles. This was done by searching systematically for the keywords: “Sensitivity”, “specificity”, “accuracy”, “ROC / receiver operating characteristic / curve”, “value”, “odds-ratio”, “likelihood ratio”, “validity”, “variance”, “effect size”, “predictive value”, and “AUC“. Additionally, to determine whether PACC was a focal point in any sensitivity analyses or discussions that might not include these keywords, all sentences regarding “PACC” were investigated as well. This led to exclusions of articles that did not include any discussions or analyses of PACC, but instead had used PACC exclusively as a measurement tool in their analyses. Many articles were further excluded in this screening due to being abstracts, podium presentations, manuscripts, or unpublished preprints. The last screening included a full-text reading of the 30 articles, which led to 20 included articles in this review. The articles met all inclusion criteria, as well as contained data on PACC sensitivity and accuracy. The entire search process is illustrated in the flow diagram below (see Figure 2). All steps of this screening follow the principles of PRISMA.

4.0 Results

This section will present the findings from a systematic review of 20 studies meeting the inclusion criteria. The process of the study selection is detailed in a flow diagram (see Figure 2), followed by a quality assessment of the included studies. A comprehensive table summarizes the key characteristics and findings of each study. Finally, the results are presented, where the main findings across the studies will be highlighted.

4.1 Figure 2: Flow diagram of the selection process

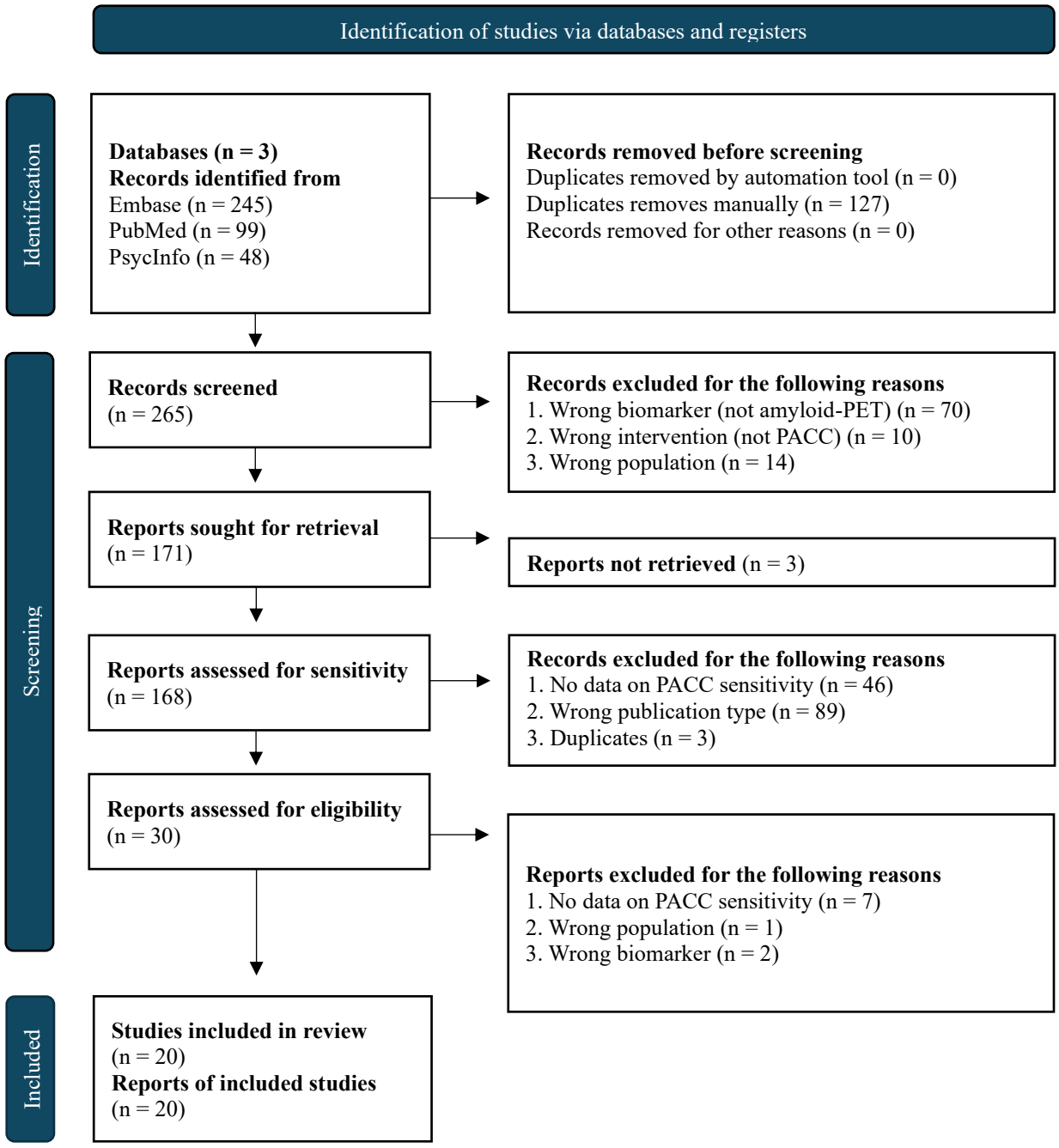

Figure 2: PRISMA Flow diagram (Page et al. 2021)

4.2 Quality assessment

The JBI Critical Appraisal Tool for Diagnostic Test Accuracy Studies is a validated instrument developed by JBI, and it is specifically designed to assess the methodological quality of diagnostic test accuracy studies and to identify potential sources of bias in study design, conduct, and analysis (Whiting et al., 2011). The assessment tool is also known as QUADAS-2, which stands for Quality Assessment of Diagnostic Accuracy Studies (Whiting et al., 2011). The tool is focused on evaluating the reliability and validity of DTA studies, which helps to ensure high-quality evidence regarding clinical decisions in healthcare. It consists of 10 structured questions covering key aspects such as patient selection, index test interpretation, reference standard, blinding, and data analysis methods (see Table 3). (See full QUADAS-2 in Attachment 4).

4.2.1 Table 3: QUADAS-2 - Questions

Question number	Question
1	Was a consecutive or random sample of patients enrolled?
2	Was a case control design avoided?
3	Did the study avoid inappropriate exclusions?
4	Were the index test results interpreted without knowledge of the results of the reference standard?
5	If a threshold was used, was it pre-specified?
6	Is the reference standard likely to correctly classify the target condition?
7	Were the reference standard results interpreted without knowledge of the results of the index test?
8	Was there an appropriate interval between index test and reference standard?
9	Did all patients receive the same reference standard?
10	Were all patients included in the analysis?

Table 3: QUADAS-2 – Questions (Whiting et al. 2011.

The QUADAS-2 assessment tool was used to evaluate the quality of each included study in this review, where the answer to each question could be either “yes”, “no”, “unclear”, or “not applicable”

(Whiting et al., 2011). The results of the quality assessment for each included article are presented in Table 6, and it includes 20 articles as previously presented in the flow diagram. This detailed review of the quality of the included studies provides insights into the strengths and weaknesses of each study and its contribution to the overall evidence base. As QUADAS-2 does not introduce a numerical scoring total, the total estimate is based on the total amount of times the answer, "Yes" is the answer, whereas "No" and "U" (unclear) both equals a score of 0. A numerical value of quality assessment can help to make a quick and clear quality check for each included article. (See Table 4).

4.2.2 Table 4: QUADAS-2 - Results

	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q8**	**Q9**	**Q10**	**Total**
S1	Yes	Yes	Yes	**No**	Yes	Yes	Yes	**U**	**No**	Yes	**7/10**
S2	Yes	Yes	**No**	**No**	Yes	Yes	Yes	Yes	Yes	**No**	**7/10**
S3	Yes	Yes	Yes	**No**	Yes	Yes	**U**	**No**	Yes	Yes	**7/10**
S4	Yes	Yes	**No**	**No**	Yes	Yes	Yes	Yes	Yes	**No**	**7/10**
S5	Yes	Yes	Yes	**No**	Yes	Yes	Yes	**U**	**No**	**U**	**6/10**
S6	Yes	Yes	Yes	**No**	Yes	Yes	Yes	Yes	Yes	Yes	**9/10**
S7	Yes	Yes	Yes	**No**	Yes	Yes	**U**	Yes	Yes	Yes	**8/10**
S8	Yes	Yes	**U**	**No**	Yes	Yes	Yes	Yes	Yes	**U**	**7/10**
S9	Yes	Yes	Yes	**No**	Yes	Yes	Yes	**U**	Yes	Yes	**8/10**
S10	Yes	Yes	**U**	**No**	Yes	Yes	Yes	Yes	Yes	Yes	**8/10**
S11	Yes	Yes	Yes	**No**	Yes	Yes	Yes	**U**	**No**	Yes	**7/10**
S12	Yes	Yes	Yes	**No**	Yes	Yes	Yes	Yes	**No**	Yes	**8/10**
S13	Yes	Yes	**U**	**No**	Yes	Yes	Yes	**U**	**No**	**U**	**5/10**
S14	Yes	Yes	Yes	**No**	Yes	Yes	Yes	Yes	Yes	Yes	**9/10**
S15	Yes	Yes	Yes	**No**	Yes	Yes	Yes	Yes	**No**	Yes	**8/10**
S16	Yes	Yes	**No**	**No**	Yes	Yes	Yes	**U**	**No**	**No**	**5/10**
S17	Yes	Yes	**No**	**No**	Yes	Yes	Yes	**U**	Yes	Yes	**7/10**

S18	Yes	Yes	Yes	**No**	**No**	Yes	Yes	**U**	**No**	Yes	**6/10**
S19	Yes	Yes	Yes	**No**	**No**	Yes	Yes	**U**	**U**	Yes	**6/10**
S20	Yes	Yes	Yes	**No**	Yes	Yes	Yes	**U**	Yes	Yes	**8/10**

Abbreviations: Q = Question, **S =** Study**, U =** Unclear

Table 4: QUADAS-2 – Results

4.3 The table of the reviewed articles

The following table provides a detailed summary of the 20 articles included in this systematic review (see Table 5), each of which evaluates the diagnostic accuracy and sensitivity of the PACC for detecting preclinical AD, compared to Aβ-PET imaging among individuals with elevated Aβ pathology. These articles were selected based on the thorough quality assessment (QUADAS-2), as shown above, ensuring that only high-quality studies were included. The table includes key information for each study, such as the author(s), study type (longitudinal studies and their follow-up periods), sample size (including gender and age distribution), reference test information (PET-imaging type and the threshold used), index test information (PACC version and its included subtests), baseline data required in the study, sensitivity analysis method(s) used, their main results, and a brief overview of their limitations. Additionally, the quality score from QUADAS-2 has been added to summarize the quality of each study. This comprehensive overview allows for a systematic comparison of the studies, highlighting crucial aspects and differences that inform the discussion of the diagnostic utility of PACC in preclinical AD detection.

	Author, year, country	Study type	Sample	Aβ-PET (Reference)	PACC version (Index)	Baseline data	Sensitivity analysis method	Results	Limitations
1	(Bransby et al., 2019) Australia	Longitudinal study 18-months follow-up 36-months follow-up	AIBL dataset Age = 60+ *n* = 66 CN (total) *n* = 34 Aβ+ ♀ = 52.9% *n* = 32 Aβ++ ♀ = 40.6%	^{11}C-PiB-PET / ^{18}F-FBP-PET / ^{18}FLUTE-PET SUVR cutoff = 1.4. 1.4-1.9 = Aβ+ >1.9= Aβ++	1. ADCS-PACC (MMSE, CVLT-II, LM II, DSST) 2. ADCS-PACC (without MMSE) 3. ZAVEN (CVLT-II, LM-II, DSC, D-KFES verbal fluency)	CDR = 0.	Effect sizes measured	Aβ+ showed slower rate of cognitive decline than Aβ++. Exclusion of MMSE did not improve sensitivity. Higher Aβ = worse cognitive performances. Variation in Aβ is connected to rate of decline. ADCS-PACC (*d* = 0.85). ADCS-PACC no MMSE (*d* = 0.62) ZAVEN (*d* = 0.72)	Small sample size. Aβ accumulation is more dynamic than allowed for in this study design. Quality score = 7/10
2	(Buckley et al., 2017) USA	Longitudinal study 3-year follow-up	HABS dataset *n* = 237 CN Age = 63-90 ♀ = 60%.	^{11}C-PiB-PET DVR cutoff = >1.2.	PACC: 1. LM-DR (WMS) 2. MMSE 3. DSST (WAIS) 4. FCSRT (free + cued)	CDR = 0 MMSE = >25	Maximum likelihood estimation Between network connectivity, Aβ, and PACC.	Aβ is linked to brain network connectivity (default, salience, control networks). Lower PACC z-score = higher PiB-DVR (Aβ) A quadratic model presented that higher Aβ could predict poorer PACC scores over time (acceleration in decline in the third year of follow-up).	There is a potential for attrition bias, as some participants did not complete all follow-up assessments. Quality score = 7/10
3	(Demnitz-King et al., 2023) France	Longitudinal study Every week for 18-months	Age-Well dataset *n* = 135 Age = 69.3 (SD = 3.8) ♀ = 61%	^{18}F-FBP-PET SUVR = 1.3 (SD = 0.2)	PACC5 (LM-DR from WMS, CVLT-II, WAIS Coding, category fluency	CDR = 0.	LMM analysis with interaction	No significant interaction between PACC5-scores between groups (meditation training + no intervention) PACC5 vs. Aβ = PFDR = 0.2. Episodic memory vs. Aβ = PFDR = 0.2. Executive function vs. Aβ = PFDR = 0.44. Attention vs. Aβ = PFDR = 0.8.	It is not possible to determine any potential lasting effects of either intervention since cognition was assessed immediately after the interventions ended. Quality score = 7/10
4	(Donohue et al., 2017a)	Longitudinal study	ADNI dataset *n* = 445 CN	^{11}C-PiB-PET / ^{18}F-FBP-PET	ADNI-PACC: (ADAS-Cog, DWR, LM-DR,	CDR = 0.	Logistic Mixed-Effects	Participants with elevated amyloid had significantly worse PACC-scores at 4-year follow-up (mean difference, 1.51	Data lost due to loss of follow-up for many participants.

	USA, Canada	6-months follow-up 1-year follow-up Every year follow-up	(total). Age = M=74. ♀ = 52% Normal Aβ, *n* = 243. Elevated Aβ, *n* = 202. (SUVR = >1.1)	SUVR = 1.10.	MMSE, TMT)		Model Likelihood ratio test	points; 95% CI, [0.94-2.08], p < .001). Worse scores on MMSE: Mean difference, 0.56 points; 95% CI, 0.32-0.80; P < .001), No significant difference for LM-DR: (mean difference, 0.73 story units; 95% CI, −0.02 to 1.48; P = .056).	Quality score = 7/10
5	(Donohue et al., 2017b) USA, Canada, Japan, Australia	Longitudinal study 6-, 12-, 18-, 24-, and 36-month follow-up.	4 datasets: **NA-ADNI** (*n* = 97) **J-ADNI** (*n* = 58) **AIBL** (*n* = 164) **ADCS-PI** (*n* = 918)	^{18}F-FBP-PET SUVR = 1.11.	1. NA-ADNI: (MMSE, ADAS-Cog, LM, DSST). 2. J-ADNI: (Same as NA-ADNI). 3. AIBL: (MMSE, CVLT, LM, DSST). 4. ADCS-PI: (3MSE, FCRST, NYU, DSST).	CDR = 0.	MMRM Cross-validation	MMSE and DSST have good face validity. Sensitivity was not increased by down-weighting or removing these. Results from cross-validation was limited due to small sample sizes. Minimum detectable effect sizes (d): Without optimization = 51% Logistic regression weights = 60% Minimized d weights = 58% The original PACC had the highest sensitivity.	Small sample sizes, except for ADCS-PI. Age and gender distribution is not mentioned in the paper. Quality score = 6/10
6	(Farrell et al., 2021) USA	Longitudinal study HABS: 4.21 years follow-up. AIBL: 18-months follow-up. ADNI: 2.97 year-follow-up.	**HABS** (*n* = 342), M-Age = 71.7 (8.0) ♀ = 60%. **AIBL** (*n* = 157), M-Age = 72.5 (6.72) ♀ = 55%. **ADNI** (*n* = 356) M-Age = 74.6 (6.5) ♀ = 56%.	^{11}C-PiB-PET / ^{18}F-FBP-PET 15-18.5 CL across all three samples.	PACC5 (MMSE, LM, DSST / TMT, FCSRT / CVLT / ADAS-Cog).	CDR = 0	LMM	Optimal Aβ-PET cutoff determined to be 15-18.5 Centiloid). Below this Aβ-PET threshold, cognitive decline was not significantly associated with Aβ-PET tracer retention. While above the threshold, cognitive decline tracked with Aβ-PET retention.	Individual Aβ variations may affect these findings. Variations between datasets. Quality score = 9/10

7	(Farrell et al., 2022) USA	Longitudinal study Follow-up at Year 3 Year 5 Year 8	HABS dataset *n* = 112 CN M-Age = 72.0 (SD=6) ♀ = 60%	^{11}C-PiB-PET / ^{18}F-FTP-PET CL threshold = CL40	PACC5 (FCSRT total, LM-DR from WMS, DSST, category fluency, MMSE)	CDR = 0 MMSE = ≥27 CL40 PACC = 0.16 (0.59)	Linear mixed-effects PiB-slopes	Participants with high Aβ already at baseline (>CL40), showed most cognitive decline in PACC5. Steeper PiB slopes overall (higher Aβ level) were associated with declining processing speed (as measured by DSST), executive function, and memory. DSST is highly sensitive to processing speed and memory Tau became a strong driver of decline.	DSST and its sensitivity towards multiple domains (memory, processing speed, executive function) make it less specific to pre-AD. Quality score = 8/10.
8	(Hampton et al., 2022) USA Australia Canada	Data from longitudinal studies: Exploratory. HABS + ADNI = yearly follow-up. AIBL = 18 months follow-up.	**ADNI**: (*n* = 509) M-Age = 74.2 (5.8) ♀ = 52% **HABS**: (*n* = 345) M-Age = 71.5 (7.9) ♀ = 60% **AIBL**: (*n* = 1176) M-Age = 70.7 (6.7) ♀ = 58% **A4**: (*n* = 4492) M-Age = 71.3 (4.7) ♀ = 59%.	^{11}C-PiB-PET / ^{18}F-FBP-PET Cutoffs for high Aβ: ADNI = >1.11 SUVR AIBL = >1.40 SUVR HABS = >1.185 DVR	IPACC: Proposed harmonized version. zPACC: standardized version (Sum of all PACC-versions): **ADNI**: MMSE, LM, TMT, Category fluency, ADAS-Cog. **HABS**: MMSE, LM, DSST, Category fluency, FCSRT. **AIBL**: MMSE, LM, DSST, Category fluency, CVLT-II. **A4**: MMSE, LM, DSST, FCSRT.	CDR = 0 MMSE = 24-30/25-30/26-30. zPACC median scores: ADNI=-0.08 HABS=0.11 AIBL=0.07 A4=0.04	IRT CFA MLR LMM HR Validation analyses	lPACC scores slightly outperformed zPACC in predicting AD-progression. Bigger effect size for IPACC compared to zPACC. IPACC more flexible across multiple studies and their cohort-specific differences. **LMM:** lPACC: t(df:6,978) = −10.43. zPACC: t(df:6,978) = −9.89. **HR:** lPACC: HR(95% CI) = 0.491 (0.386–0.626). zPACC: HR(95% CI) = 0.424 (0.330–0.546).	Not a longitudinal study, but an exploration of longitudinal studies. No longitudinal data in A4. Quality score = 7/10.
9	(Insel et al., 2020) USA	Longitudinal study	Data from A4 study. *n* = 4432 CN Age = 65-85	^{18}F-FBP-PET (SUVR = 1.1)	PACC (MMSE, FCSRT, LM-DR from	MMSE = 25-30. CDR = 0.	LMM	Decrease in PACC scores was associated with increasing Aβ. Old age indicated a steeper cognitive decline in Aβ+ patients (*p* = 0.02).	Investigating subthreshold Aβ levels makes it difficult to compare

			M-Age = 71.3. ♀ = 59,4%		WMS, DSST)			Females performed better in PACC at the equivalent level of Aβ as males, but cognitive decline showed a parallel pattern in both sexes in increasing Aβ. Aβ accumulation was linked to cognitive decline even at subthreshold levels (SUVR). Som subtests (FCSRT and LM-DR) showed greater sensitivity towards early cognitive decline.	to other studies using the defined threshold. There is a main focus on baseline (cross-sectional) and not on follow-up scores. Quality score = 8/10.
10	(Insel et al., 2021) USA	Longitudinal study M-time between follow-ups = 2.2 years (SD = 0.8).	Data from ADNI. ***n* = 127** **CN, Aβ-** M-Age = 70.1 (5.8) ♀ = 59.8%. ***n* = 100** **CN, Aβ+** M-Age = 72.9 (6.6) ♀ = 58%. ***n* = 70** **MCI, Aβ+** M-Age = 72.0 (6.9) ♀ = 41.4%. ***n* = 38** **AD, Aβ+** M-Age = 74.5 (7.2) ♀ = 47.4%.	^{18}F-FBP-PET (SUVR = 1.1)	PACC: (MMSE, LM-DR from WMS, ADAS-Cog, TMT)	CDR = 0.	LMM	The study demonstrates PACC performance declines before Aβ positivity (in cohort CN, Aβ-). MMSE showed a 0.2 SD drop six years before Aβ+. PACC decreased four years before. PACC outperformed several other predictors (such as CSF, tau-PET, general cognition tests), when it came to predicting cognitive outcome and decline. PACC is more accurate than individual biomarker or cognition measures alone.	Cross-sectional evaluation of Aβ status, not looking at the longitudinal data for the initial analysis. There are differences in disease trajectories among the cohorts (CN, MCI, AD). Quality score = 8/10.
11	(Jutten et al., 2022) USA	Longitudinal study 3-month follow-up.	Data from HABS. *n* = 114 CN. Age = 77.6 (±5.0) ♀ = 61%.	^{11}C-PiB-PET / ^{18}F-FBP-PET DVR: 1.21 ± 0.23.	PACC5: (MMSE, LM-DR from WMS, DSST, FCSRT, category fluency task). Computerized C3:	MMSE 29 ± 1.2. CDR = 0.	LMM ROC	BPSO showed strong predictive value (AUC = 0.90). FNAME also have a good predictive value (AUC = 0.80). PACC5 have a moderate predictive value compared to subtests within C3 (AUC = 0.75).	Sensitivity and specificity assessments for PACC5 are not provided in this paper.

					(FNAME, BPSO, CBB: Including DET, IDN, OCL, and ONB).				Quality score = 7/10.
12	(Lim et al., 2016) Australia	Longitudinal study. 18-, 36-, 54-, and 72-month follow-up.	AIBL dataset. *n* = 423 CN. *n* = 326 Aβ- M-Age = 68.27 (5.95) ♀ = 54.9%. *n* = 33 Aβ+ M-Age = 73.06 (7.13) ♀ = 66.7%. *n* = 64 Aβ++ M-Age = 73.19 (7.41) ♀ = 46.9%.	^{11}C-PiB-PET SUVR for: Aβ- = <1.5 Aβ+ = 1.5-1.9 Aβ++ = >1.9 ^{18}F-FBP-PET Aβ- = <1.10 Aβ+ = 1.10-1.29 Aβ++ = ≥ 1.29 ^{18}FLUTE-PET Aβ- = <0.61 Aβ+ = 0.61-0.82 Aβ++ = ≥ 0.82	**ADCS-PACC** (CVLT-II, LM-DR, DSST, MMSE). **ADCS-PACC (no MMSE)** (CVLT-II, LM-DR, DSST). **EM composite** (CVLT-II, LM-DR, Rey Complex Figure). **ZAVEN** (DSST, FAS, CVLT-II, LM-DR).	MMSE = 28-30. CDR = 0.	Effect size LMM Test-retest reliability	Aβ+ and Aβ++ CN participants showed faster cognitive decline than Aβ- across all composites. Sensitivity of composites, (Aβ+ vs. Aβ++): **ZAVEN** (d = 1.07). **ADCS-PACC no MMSE** (d = 1.01). **EM composite** (d = 0.64). **ADCS-PACC** (d = 0.64). (Aβ+ vs. Aβ-): **EM composite** (d = 0.53) **ZAVEN** (d = 0.50). **ADCS-PACC no MMSE** (d = 0.43). **ADCS-PACC** (d = 0.26). High test-retest reliability for all composites: EM composite: $r = 0.93$, $p < 0.001$ ADCS-PACC: $r = 0.92$, $p < 0.001$ ADCS-PACC no MMSE: $r = 0.94$, $p < 0.001$ ZAVEN: $r = 0.96$, $p < 0.001$	Relatively small sample size in Aβ positivity conditions. Comparisons between different composites assumes equal weighting of each test. This may not be accurate. Different SUVRs. Quality score = 8/10.
13	(Lu et al., 2019) United Kingdom	Longitudinal study.	NSHD dataset *n* = 502 CN in total. *n* = 74 Aβ+ M-Age = 70.6 (0.66) ♀ = 46%. *n* = 332 Aβ-	^{18}F-FBP-PET SUVR >0.6104.	PACC: (MMSE, LM from WMS, DSST, FNAME).	CDR = 0.	Standardized β-coefficient (95% CI)	Females generally performed better than males on PACC subtests. **Association between Aβ positivity and cognition** show significant declines in PACC scores for Aβ-positive individuals, (β = -0.17, CI [-0.32, -0.02], p = <0.05). **Higher WMHV** is associated with poorer PACC performance (β = -0.10, 95% CI [-0.20, -0.01], p = <0.05). **MMSE**: β = -0.24, CI [-0.46, -0.02],	A part (n = 57) of participants had missing biomarker data. Relatively small effect size (β-value). Relatively small sample size with many Aβ-negative. Quality score = 5/10.

			M-Age = 70.6 (0.70) ♀ = 51%.					$p<0.05$. **DSST**: β = -035, CI [-0.51, -0.19], $p<0.01$. **LM-IR**: β = -0.31, 95% CI [-0.56, -0.06], $p<0.05$. **LM-DL**: β = -0.20, 95% CI [-0.44, 0.05], not statistically significant.	
14	(Mormino et al., 2017) USA	Longitudinal study Annual follow-ups. Focus on year 3 and 5.	HABS dataset. *n* = 277 CN in total. *n* = 206, Aβ- M-Age = 72.9 (±6.0) ♀ = 59%. *n* = 71, Aβ+ M-Age = 75.2 (±5.7) ♀ = 61%.	^{11}C-PiB-PET SUVR = 1.20.	1. PACC (MMSE, LM-DR from WMS, DSC from WAIS-R, FCSRT-Free). 2. PACC no MMSE 3. PACC no FCSRT-Free 4. PACC no LM 5. PACC no DSC	CDR = 0. MMSE ≥ 25.	LMM MMRM HR Effect size	Greater risk of progression to CDR 0.5 in Aβ+ compared with Aβ- (hazard ratio = 1.84, $p = 0.021$). **PACC**: High sensitivity. 3-year follow-up = Large effect. 5-year follow-up = Large effect. **PACC no MMSE**: Early sensitivity. 3-year follow-up = Large effect. 5-year follow-up = Large effect. **PACC no FCSRT**: Lower sensitivity. 3-year follow-up = Decreased effect. 5-year follow-up = Decreased effect. **PACC no LM**: Variable sensitivity. 3-year follow-up = Small effect 5-year follow-up = Large effect. **PACC no DSC**: Not consistent 3-year follow-up = Variable 5-year follow-up = Variable	Highly educated dataset. Moderate and varying sample sizes. Quality score = 9/10.
15	(Papp et al., 2017) USA	Longitudinal study Annual follow-ups. Up to 5 years.	HABS dataset. *n* = 279 CN in total *n* = 70 Aβ+ M-Age = 74.99 ± 5.74 ♀ = 61%. *n* = 209 Aβ- M-Age = 72.88 ± 6.02	^{11}C-PiB-PET+ ^{18}F-FDG-PET SUVR = 1.20·	PACC5 (MMSE, LM-DR, DSC, FCSRT, + Category fluency test / CAT)	CDR = 0. MMSE = normal performance	Effect size: (coefficient (β)) LMM	Removing CAT from the PACC resulted in a 20% reduction in effect size of Aβ-related cognitive decline at 3-year follow up, and a 12% reduction at 5-year follow-up. CAT-component makes PACC more sensitive. PACC with CAT = large effect size, high sensitivity: (β = -0.09, SE = 0.02, $p < 0.001$). PACC without CAT = 12% reduction, reduced sensitivity: (β = -0.08, SE = 0.03, $p < 0.01$).	Sample is highly educated and predominantly Caucasian (82%). The focus on semantic memory (CAT) as addition improved sensitivity. But what about other cognitive domains? Potential practice effects using same

			♀ = 59%'.					PACC without FCSRT = 21% reduction in effect size, lower sensitivity: (β = -0.06, SE = 0.02, $p < 0.001$).	tests annually. Quality score = 8/10.
16	(Rentz et al., 2021) USA	Longitudinal study Annual follow-ups. Focus on 18-month and 36-month follow-up.	HABS dataset. *n* = 264 CN in total. *n* = 103 Aβ- M-Age = 77.89 ± 6,42 ♀ = 65.0%. *n* = 40 Aβ+ M-Age = 79.61 ± 5.67 ♀ = 50.0%. Additionally, MCI+AD: *n* = 36 CI. M-Age = 77.09 ± 8.13. ♀ = 54.3%.	^{11}C-PiB-PET+ ^{18}F-FTP-PET (*n* = 143 CN) Cutoff = 1.185	PACC (MMSE, LM, DR, DSST, FCSRT) DCTclock as potential addition.	CDR = 0.	Effect size ROC Correlation	<u>Diagnostic discriminability:</u> PACC had greater discriminative ability between CN and MCI (AUC = 0.95). Effect size, Cohens' d = 2.42. DCTclock showed good discrimination between CN and MCI (AUC = 0.86). Effect size, Cohen's d = 1.55. <u>Amyloid status discrimination:</u> DCTclock had better discrimination between Aβ+ and Aβ- (AUC = 0.72, d = 0.76), compared to PACC (AUC = 0.63, d = 0.30). <u>Correlation with amyloid</u> DCTclock correlation with amyloid = $r = -0.241$, $p < 0.01$. PACC correlation with amyloid = $r = -0.100$, $p = 0.295$.	Only a small subset of the total number of participants underwent PET-imaging (143 out of 264). But they were still part of the analyses. Digital versus traditional test methods may be difficult to compare. Quality score = 5/10.
17	(Ruthirakuhan et al., 2024) USA Canada Australia Japan	Longitudinal study.	A4 Study dataset. *n* = 5.061 CN. Age = 65-85. *n* = 3.115 Aβ-, M-Age = 71.0 (4.5). ♀ = 59%. *n* = 1.309 Aβ+, M-Age = 72.1 (4.9) ♀ = 59%.	^{18}F-FBP-PET SUVR threshold of <u>≥</u> 1.15. (Defined Aβ+)	PACC: (MMSE, LM-DR, FCSRT, DSST) CFI	CDR = 0. MMSE = 25 or higher.	CFA	PACC and CFI both showed good measurement invariance. PACC was effective in differentiating between individuals with and without amyloid plaques (Aβ+ vs. Aβ-) (d = 0.39, $p < 0.001$). CFI was equally as effective in differentiating between Aβ+ and Aβ- participants (d = 0.39, $p < 0.001$).	Too small sample sizes were excluded (excluding eg. American Indian, Alaskan native, Hawaiian) limiting the inclusivity. Analysis is based on baseline screening data, and longitudinal cognitive changes were not assessed. Quality score = 7/10.

18	(Sato et al., 2021) USA Canada Japan Australia	Longitudinal study.	A4 dataset, *n* = 3233 CN M-Age = 70.2 ♀ = 59.5%. ADNI dataset, *n* = 86 CN M-Age = 75.6 ♀ = 44.2%. J-ADNI dataset, *n* = 50 CN M-Age = 67.4 ♀ = 44.0%.	^{18}F-FBP-PET M-SUVR in A4-study: 1.15 SUVR in ADNI study: 1.11-1.22 SUVR in J-ADNI study ≥ 1.15	PACC: (MMSE, LM-DR, DSST, FCSRT) ADNI-PACC: (MMSE, LM-DR, DSST, ADAS-Cog)	CDR = 0. MMSE = 27-30.	ROC (Optimization procedure before and after new SUVR-thresholds)	Models without considering APOE showed a higher expected maximum AUC improvement compared to models with APOE. **ADNI cohort**: With APOE: Mean AUC improvement = 0.033 (AUC improved from 0.724 to 0.774). Without APOE: Mean AUC improvement = 0.075 (AUC improved from 0.61 to 0.69). **J-ADNI cohort:** With APOE: Mean AUC improvement = 0.009 (AUC from 0.65 remained at 0.65). Without APOE: Mean AUC improvement = 0.019 (AUC improved from 0.61 to 0.64).	Smaller sample size in J-ADNI compared to ADNI. Datasets also have different amyloid positivity criteria. The study used baseline data only, not assessing the longitudinal changes. Quality score = 6/10.
19	(Sperling et al., 2014) USA Canada Australia Japan	Longitudinal study.	A4 study. *n* = 4486 CN in total *n* = 3160 Aβ-, M-Age = 70.95 (4.53) ♀ = 60%. *n* = 1323 Aβ+, M-Age = 72.10 (4.89) ♀ = 59%.	^{18}F-FBP-PET M-SUVR = 1.33 (0.18) for Aβ+.	PACC: (MMSE, DSST, LM-DR, FCSRT) +CFI	CDR = 0. MMSE = 25-30.	Effect size	PACC overall: d= -0.18, 95% CI [-0.21, -0.15], p<.001 MMSE: d = -0.05, 95% CI [-0.08, -0.02], p<.01 FCSRT-Free: d = -0.14, 95% CI [-0.17, -0.11], p<.001 LM-DR: d = -0.10, 95% CI [-0.13, -0.07], p<.001 DSST: d = -0.13, 95% CI [-0.16, -0.10], p<.001	Threshold variability across different populations. This study bases the analysis on baseline data and not longitudinal data. Quality-score = 6/10.
20	(Stricker et al., 2023) USA	Longitudinal study. Every 15-month	MCSA study *n* = 614 CN *n* = 428 Aβ- M-Age =	^{11}C-PiB-PET+ SUVR ≥ 1.48 (CL = 22).	Mayo-PACC: (AVLT, TMT, animal fluency). PACC-R:	CDR = 0.	Slope differences (effect sizes)	All composites showed sensitivity to Aβ+-related longitudinal cognitive decline, with Aβ+ individuals showing greater decline over time than Aβ-.	Significant practice effects found in ADCS-PACC. There was a higher

		follow-up. M-follow-up of 7 years.	72.47 (4.81) ♀ = 45.6%. *n* = 186 Aβ+ M-Age = 73.79 (5.03) ♀ = 45.2%.		(AVLT, AVLT-DR, AVLT-R, DSST, Category fluency). ADCS-PACC: (AVLT-DR, LM-II, DSST, MMSE). Global-z: (AVLT-DR, LM-II, DSST, TMT, Category fluency, BNT, WR-II, PC, BD).		Test-retest reliability LMM	Standardized slope differences in rate of decline between Aβ+ and Aβ+-: Mayo-PACC: d = -0.132 (95% CI: -0.18 to -0.084, $p < 0.05$). PACC-R: d = -0.134 (95% CI: -0.179 to -0.089, $p < 0.05$). ADCS-PACC: d = -0.154 (95% CI: -0.201 to -0.106, $p < 0.05$). Global-z: d = -0.133 (95% CI: -0.176 to -0.09, $p < 0.05$). All tests had high test-retest reliability ($\geq$ 0.80).	proportion of APOE carriers in Aβ+ group. Potential for overfitting. Quality-score = 8/10.

Table 5: Overview of the reviewed articles

Abbreviations:

AIBL = Australian Imaging Biomarkers and Lifestyle study, **CN** = Cognitively normal, **SD** = Standard deviation, **Aβ** = Amyloid-beta, **^{11}C-PiB** = Pittsburgh Compound B, **^{18}F-FBP** = ^{18}F-Florbetapir, **^{18}F-FLUTE** = 18F-Flutemetamol, **SUVR** = Standard Uptake Value Ratio, **Aβ+** = High amyloid level, **Aβ++** = Very high amyloid level, **M-Age** = Mean age, **ADCS** = Alzheimer's Disease Cooperative Study, **PACC** = Preclinical Alzheimer Cognitive Composite, **ZAVEN** = Z-scores of Attention, Verbal Fluency and Episodic Memory for Nondemented older adults, **MMSE** = Mini-Mental State Examination, , ♀ = Percentage of female participants, **PET** = Positron Emission Topography, **FCSRT** = Free and Cued Selective Reminding Test, **DSST** = Digit Symbol Substitution Test, **WAIS-R** = Wechsler Adult Intelligence Scale-Revised, **WMS** = Wechler Memory Scale, **LM-DR** = Logical Memory Delayed Recall, **CVLT-II** = California Verbal Learning Test-Second Edition, **LM-II** = Logical Memory II, **LM-DR** = Logical Memory Delayed Recall, **DSC** = Digit Symbol-Coding, **D-KEFS** = Delis-Kaplan Executive Function System, **HABS** = Harvard Ageing Brain Study, **CDR** = Clinical Dementia Rating Scale, **DVR** = Distribution Volume Ratio, **LMM** = Linear Mixed effects Model, **PFDR** = P-value of False Discovery Rate, **ADNI** = Alzheimer's Disease Neuroimaging Initiative, **ADAS-Cog** = Alzheimer Disease Assessment Scale-Cognitive Subscale, **DWR** = Delayed Word Recall, **TMT** = Trail-Making-Test, **NA-ADNI** = Northern American ADNI, **J-ADNI** = Japan-ADNI, **ADCS-PI** = ADCS-Prevention Instrument, **3MSE** = Modified MMSE, **NYU** = New York University Paragraph Recall, **MMRM** = Mixed method of Repeated Measures, **WRAP** = Wisconsin Registry for Alzheimer's Prevention, **IRT** = Item Response Theory, **CFA** = Confirmatory factor analysis, **MLR** = Maximin Likelihood estimation with Robust standard errors, **A4** = Anti-Amyloid Treatment of Asymptomatic Alzheimer's Disease, **IPACC** = Latent PACC, **zPACC** = Standardized PACC with z-scores, **HR** = Hazard ratio, **Aβ-** = No Amyloid-beta, **MCI** = Mild Cognitive Impairment, **AD** = Alzheimer's Disease, **ROC** = Receiver Operating Characteristic analysis, **FNAME** = Face Name Associative Memory Exam, **BPSO** = Behavioral Pattern Separation Task-Object version, **CBB** = Cogstate Brief Battery, **DET** = Detection Task, **IDN** = Identification Task, **OCL** = One Card Learning Task, **ONB** = One Back Task, **FAS** = Also known as COWAT: Controlled Oral Word Association Test with the letters F+A+S, **NSHD** = National Survey of Health and Development, **WMHV** = White Matter Hyperintensity Volume, **LM-IR** = Logical Memory Immediate Recall, **MMRM** = Mixed Model of Repeated Measures, **FCSRT-Free** = Free and Cued Selective Reminding Test Free Recall, **^{18}F-FDG** = ^{18}F-Fludeoxyglucose, **DCTclock** = Digital Clock-drawing Test, **^{18}F-FTP** = ^{18}F-Flortaucipir, **CI** = Cognitively impaired, **CFI** = Cognitive Function Index, **MCSA** = The Mayo Clinic Study of Aging, **PACC-R** = PACC-Revised, **Global-z** = ??, **AVLT** = Rey Auditory Verbal Learning Test, **AVLT-DR =** AVLT-Delayed Recall, **AVLT-R** = AVLT-Recognition, **BNT** = Boston Naming Test, **VR-II** = Visual Reproduction-II from WMS, **PC** = Picture Completion from WMS, **BD** = Block Design from WAIS.

4.4 Main findings in reviewed articles

The systematic review synthesized findings from 20 longitudinal studies that evaluated the diagnostic accuracy and sensitivity of PACC compared to Aβ-PET imaging for detecting preclinical AD among individuals with elevated Aβ pathology. As dictated by the inclusion criteria, all studies had a longitudinal design with follow-up periods ranging from 3 months to 7 years. The studies focused on cognitively normal (CN) individuals, with varying sample sizes and Aβ-statuses. Despite differences in how they investigated Aβ pathology related to PACC scores, all studies examined the relationship between the two. Aβ-PET imaging was used in all studies, as required by the inclusion criteria, though the types of PET imaging and thresholds for Aβ-positivity varied. The variability in the use of Aβ-PET imaging tracers included, ^{11}C-PiB-PET, ^{18}F-FBP-PET, and ^{18}F-FLUTE-PET, as well as varying thresholds and cutoff-values in SUVRs, DVRs and CLs for defining Aβ-positivity. PET tracers and SUVR thresholds used to determine Aβ positivity varied both across and within studies with multiple datasets. These variations will be discussed in the context of their limitations. Across all studies, a version of PACC was used as a measurement tool, demonstrating varying degrees of sensitivity.

The included studies also seem to agree to old age being a main factor for developing cognitive decline, and this can be seen in all samples recruiting participants above the age of sixty, with a mean age across all studies at 76,93 years. Gender differentiation also showed that females performed better on PACC when at the same level of Aβ pathology as males. However, as Aβ levels increased, both sexes exhibited similar rates of cognitive decline (9, 13).

4.4.1 PET imaging and PACC

As the inclusion criteria stated, Aβ-PET imaging and PACC measurements were required to be part of the included studies and their results. This was also true for all 20 studies. Across the studies,

different versions of PACC were used, with some including or excluding specific subtests such as MMSE, LM-DR, and DSST. Similarly, different tracers and thresholds were used in Aβ-PET imaging in determining Aβ positivity. Despite the variations, all studies did at some point compare the scores from PACC (index test) with the level of amyloid (reference test). These results consistently demonstrated that amyloid accumulation led to cognitive decline. In some studies, (1, 7, and 12), it was even underlined that a higher Aβ level (Aβ++) showed a faster rate of cognitive decline compared to individuals with only slightly elevated Aβ pathology (Aβ+).

4.4.2 Sensitivity of PACC

Across the reviewed studies, PACC demonstrated varying degrees of sensitivity and specificity in detecting cognitive decline associated with elevated Aβ levels. However, the sensitivity of PACC was generally high, and higher Aβ levels were demonstrated to affect cognitive decline throughout all reviewed articles. One study (10) even claimed, PACC could potentially detect performance decline four years before any measurable Aβ pathology. Studies (1, 5, and 15) demonstrated that the inclusion or exclusion of certain subtests affected the sensitivity of PACC. For instance, the original PACC generally showed higher sensitivity compared to modified versions that excluded MMSE or DSST (5, 12). However, sensitivity increased with the addition of the subtest, CAT (15). Other studies (9, 19) found certain subtests to individually show more sensitivity, such as the FCSRT and LM-DR.

PACC demonstrated effectiveness in predicting changes in cognitive function over time related to Aβ in many studies (2, 3, 8, and 15), and PACC's validity was supported by showing significant associations between baseline PACC scores and baseline connectivity in brain networks. PACC was found to be a valuable tool for early detection of cognitive decline, showing comparable performance to Aβ-PET imaging (2, 6, 7, 14, and 16). High sensitivity was shown for PACC at 3-year and 5-year

follow-ups (14). Furthermore, it was found that PACC had great discriminative ability between CN individuals and individuals with MCI (16).

4.4.3 Statistical methods

Throughout the reviewed studies, different statistical methods were employed to determine the sensitivity and accuracy of PACC. Including logistic mixed-effects models, linear mixed-effects models (LMM), and ROC analyses. Studies (4, 11, 17, and 18) showed how different statistical methods influenced the sensitivity reports. The studies using LMM (3, 4, 6, 7, 8, 9, 10, 11, 12, 14, 15, and 20) showed that elevated Aβ was associated consistently with worse PACC scores across all variations of the composite. There was also found a moderate predictive value for PACC in the ROC analyses done by a few studies (11, 16). Furthermore, CFA showed that PACC was great for differentiating between participants with and without Aβ pathology (17), the effect size was higher for PACC with the addition of CAT (15), and β-coefficients showed significant associations between Aβ positivity and decline in PACC scores (13).

5.0 Discussion

This discussion will firstly synthesize the results and evidence from the reviewed articles, comparing the findings from the 20 studies. It will address considerations and limitations related to the results, methodology, and quality assessment. The research in this review focuses on evaluating the diagnostic accuracy and sensitivity of PACC in detecting preclinical AD, compared to Aβ-PET imaging, among individuals with elevated Aβ pathology. This discussion will explore how these findings compliment and contradict each other, as well as how their differences in methodology can cause certain limitations. The table discussion has been divided into themes to maintain a clear

overview of the discussion flow. The methodology of this review will similarly be discussed, detailing the screening process, the impact of the inclusion and exclusion criteria, and the challenges encountered during the synthesis. A summary of the main strengths and weaknesses in the use of PACC and Aβ-PET imaging will be presented, and a discussion of the potential directions for future research will conclude this review.

5.1 Theme 1: Methodological considerations

Sample sizes and demographics (age, gender, education level, and ethnicity) varied significantly across studies. Smaller sample sizes (studies, 1, 3, 5, 7, 12, 18), can reduce statistical power and increase the risk of type II errors. Demographic differences can affect the generalizability of findings and introduce biases. For instance, highly educated samples might exhibit less cognitive decline, underestimating the sensitivity of PACC in more diverse populations. This can lead to educational bias. It can be difficult to determine, whether this has affected the results of any studies. But it should be considered that studies (2, 6, 7, 8, 11, 14, 15, 16) all investigated the HABS (Harvard Aging Brain Study) dataset, which has a high percentage of highly educated individuals, as mentioned in study 14 and 15. Many studies also investigated multiple datasets in a single study, but within each dataset there might be varying requirements and methodologies. By using different tools or protocols to measure the same variables, the results in the data might instead be due to measurement differences rather than true differences in the phenomena being studied. These results should be interpreted with caution.

Follow-up durations between studies also varied, which refers the frequencies of assessments and measurements. This could have an impact on the results from each study. Shorter follow-up periods may miss long-term changes in cognitive decline, while longer intervals between assessments might overlook subtle changes. Furthermore, it may be challenging to compare the results from studies with significantly different follow-up periods.

Another important notion is that studies varied in their use of a specific reference Aβ-PET imaging protocol. This variability in the use of Aβ-PET imaging tracers, as well as varying thresholds and cutoff-values for defining Aβ positivity, might lead to inconsistencies in identifying individuals as Aβ positive or negative, making direct comparisons between studies more difficult. For example, a study with a higher SUVR cutoff might classify fewer individuals as Aβ positive, potentially underestimating the sensitivity of PACC (for instance, 1, 8, 12, 20). The use of different PET tracers (^{11}C-PiB, ^{18}F-FBP-PET, ^{18}F-FLUTE-PET) across studies, and sometimes within the same study, can lead to inconsistent sensitivity and specificity results. This might also affect the follow-up scores, seeing as the initial reference scores have different requirements for cut-off thresholds and therefore different baseline scores. These variations all highlight the need for standardization in Aβ-PET imaging protocols.

Despite variations in methodologies and cohort characteristics, several key themes and findings emerge across the studies. Many studies (2, 4, 7, 9, 10, 12, 13, 20) found that a higher level of Aβ is associated with more significant cognitive decline, as measured by PACC scores. This association was consistently observed regardless of differences in cohort sizes and follow-up periods, underscoring the robustness of this relationship across diverse study designs. These findings highlight the critical role of Aβ in cognitive decline.

5.1.1 The quality of reviewed articles

QUADAS-2 was used to assess the quality of the studies included in this review. All studies had enrolled a consecutive sample of patients, as the criteria was an age above 60+ as well as cognitively normal individuals scoring 0 on the CDR scale. All studies similarly avoided a case control design, where patients with a condition (AD), are compared to those without the condition. This approach helps to avoid retrospective bias, such as selection bias and recall bias.

Most studies (12 studies: 1, 3, 5, 6, 7, 9, 11, 12, 14, 15, 18, 19, 20) avoided inappropriate exclusions. A few studies (3 studies: 8, 10, 13) did not specify this, which was characterized as unknown. And lastly, a few studies (4 studies: 2, 4, 16, 17) made seemingly inappropriate exclusions. These four studies excluded data based on participants not completing all follow-up assessments (2, 4), as well as data not being based on the same reference test (16), and lastly because of a few sample sizes being too small (17). Excluding data due to missing follow-up information can be considered an inappropriate exclusion because it introduces attrition bias. This type of exclusion can skew the study results if the participants who are lost to follow-up differ systematically from those who remain. In this case, if individuals with more severe cognitive decline are more likely to miss follow-up assessments, the study may underestimate the true extent of cognitive decline associated with preclinical AD as a condition. Also, the exclusion of some data due to different reference tests can affect the results in a way, where valuable information could be lost. Lastly, excluding data and samples because they are too small can be considered inappropriate, because it limits the inclusivity and diversity of the study. For instance, this study (17) misses out on representing American Indian, Alaskan Native, and Hawaiian populations in their study, which might overlook important nuances.

No studies interpreted the index test (PACC) without the knowledge of the reference test (PET), which explains why all studies scored a "no" on this question. As the knowledge of Aβ pathology was the basis of undergoing a neuropsychological investigation, it was impossible to avoid this. Knowledge of the reference test while interpreting the index test results can introduce confirmation bias, where researchers might consciously or unconsciously interpret the PACC results in a way that is consistent with the PET results. However, it is unknown if there was used blinding in the analysis, where different researchers dealt with different aspects of the study, without knowledge of the parts of the study that might affect their interpretation of the results. The lack of blinding can lead to lack

of objectivity, idiosyncratic errors, the overestimation of PACC accuracy, and compromised data integrity.

In relation to whether a threshold was used and if it was pre-specified, most studies could answer “yes”, except for two studies (18, 19). One study investigates SUVR thresholds, so they measured many different thresholds (18). Another study had varying thresholds across multiple participant groups (19). This variability can make it difficult to compare the results with other studies’ results.

All studies could rely on the reference standard being likely to correctly classify the target condition, as all studies had their participants undergo Aβ-PET imaging prior to being part of the studies as a requirement. Similarly, most studies (except 3, 7) could interpret the reference standard results without knowledge of the results of the index test, as the PACC assessment was done afterwards. Two studies (3, 7) never clarified this nor the time interval between the reference test (PET) and the index test (PACC). Additionally, many studies (1, 3, 5, 9, 11, 13, 16, 17, 18, 19, 20) never clarified what the interval between the reference and the index test was, making it difficult to determine, whether it was an appropriate time interval.

The participants in half of the studies received the exact same reference standard (2, 3, 4, 6, 7, 8, 9, 10, 14, 17, 20). The rest still used Aβ-PET imaging but with varying tracers. Lastly, most studies included all participants in the analysis, except for those studies that made inappropriate exclusions.

5.2 Theme 2: PACC versions and subtests

Another challenge in comparing the result from the reviewed articles is the use of varying versions of PACC and its subtests. Some studies looked at the sensitivity of PACC as a combined composite score (1, 2, 3, 6, 8, 11, 12, 16, 17, 18, 20), where other studies looked at PACC scores both as total scores as well as scores for the individual subtests (4, 5, 7, 9, 10, 13, 14, 15, 19). In the studies

looking at the overall composite score, most agree that the original PACC with all its subtests (MMSE, DSST, LM-DR, and FCSRT) is superior when compared to alternative composites (ADCS-PACC no MMSE, zPACC, DCTclock). Except for two studies (12, 20), suggesting the overall PACC composite might be less sensitive compared to composites, such as ZAVEN, EM composite, PACC-R, and Global-z. However, in the studies looking at both the overall PACC score as well as the sensitivity of the subtests, it was found that especially the subtests, MMSE and DSST, are sensitive tests within PACC (4, 5, 7, 10, 13, 14, 19), as well as FCSRT and LM-DR (9, 10, 13, 14, 19). One study (15) even found the addition of the subtest CAT made PACC even more sensitive. While the inclusion of CAT enhanced sensitivity (15), the exclusion of MMSE or DSST reduced sensitivity (5, 12). This variability underscores the importance of consistent PACC versions in research and clinical practice.

While the four subtests within PACC offers a balanced assessment across various cognitive domains, including memory, executive function, and visuospatial skills, CAT offers insights into verbal fluency as well as executive function in relation to language, which is not a primary focus in the original PACC composite. The inclusion of CAT enhances PACC by providing a more robust assessment of language function.

5.3 Theme 3: Sensitivity of PACC

The reviewed studies employed various statistical methods to analyse PACC's sensitivity and accuracy: Logistic mixed-effects model (4), linear mixed-effects models (LMM) (3, 6, 7, 8, 9, 10, 11, 12, 14, 15, 20), MMRM (5, 14), ROC analysis (11, 16, 18), effect size (1, 12, 14, 15, 16, 19, 20), CFA (8, 17), β-coefficient (13, 15), and HR (8, 14).

These methods influenced the findings, with logistic mixed-effects and linear mixed-effect models consistently showing that elevated Aβ was associated with worse PACC scores and cognitive decline. This consistency underscores the reliability of the statistical method as well as its

effectiveness. The ROC analyses were helpful in highlighting the discriminatory power of certain subtests, such as FCSRT and LM-DR. It was found that PACC5 had a moderate predictive value compared to a computerized composite (11), and that PACC showed greater discriminative ability between CN and MCI individuals (16). When measuring sensitivity by using effect sizes, larger effect sizes were found in studies, where MMSE and DSST were included (14, 19, 20). Additionally, adding CAT increased the effect size of PACC by 20% at the 3-year follow-up (15). However, two studies (12, 16) found the effect size to be bigger for alternative composites (ZAVEN, EM composite, and DCTclock). In the ZAVEN composite, FAS and CVLT-II is included, and in the EM composite, Rey Complex Figure is included (12). FAS measures verbal fluency and executive functions, which might help detect subtle changes in language. CVLT-II measures verbal learning and memory, providing more details into aspects of memory function. Rey Complex Figure measures visuospatial skills, memory, attention, and executive functioning. While the original PACC composite is already robust in assessing memory (LM-DR and FCSRT), general cognitive function (MMSE), and attention and executive function (DSST), some alternative subtests might be missing. ZAVEN's inclusion of FAS and CVLT-II adds depth to the evaluation of verbal fluency (language) and detailed memory processes, potentially making it more sensitive to early cognitive changes in preclinical AD. This increased sensitivity might explain the larger effect sizes observed in ZAVEN compared to the original PACC. The measurements of verbal memory in CVLT-II and visuospatial skills in Rey Complex Figure might also contribute to a broader understanding of early cognitive decline.

The CFA was used to validate the constructs behind zPACC and IPACC (8), which indicated that the factor structure of PACC was consistent and accurately representing the underlying cognitive abilities. Another study similarly confirmed the robustness of PACC in different populations (17). β-coefficient analyses show that DSST, LM-IR, and MMSE are highly sensitive to cognitive decline, making them valuable for early detection (13). HRs provide insights into the risk of cognitive decline.

And it was found that IPACC and zPACC had higher HRs, indicating that these composites, including subtests like MMSE and DSST, were all sensitive in predicting cognitive decline (8, 14).

Despite the diversity in statistical methodology, the subtests (MMSE, DSST, FCSRT, LM-DR) remained consistently sensitive across all studies and across all statistical analyses. This underlines their robustness and reliability.

5.4 Methodology discussion for this review

This section will delve into the methodological choices and limitations encountered throughout this systematic review. The eligibility criteria and the rationale behind including and excluding certain studies will be discussed, as well as the potential biases and errors inherent in the search strategy and data collection. Additionally, the use of the QUADAS-2 tool for quality assessment will be evaluated, highlighting its strengths and weaknesses.

5.4.1 Eligibility criteria

The inclusion and exclusion criteria were designed to ensure the relevance and quality of the studies reviewed. However, these stringent criteria may have led to the exclusion of potentially valuable studies. For instance, studies that did not explicitly measure sensitivity or specificity of PACC but provided relevant data on cognitive decline might have been excluded. In the screening phase, several studies were excluded due to lacking data or discussions on PACC sensitivity. Sensitivity in a study was determined by a search for certain keywords: “Sensitivity”, “specificity”, “accuracy”, “ROC / receiver operating characteristic / curve”, “value”, “odds-ratio”, “likelihood ratio”, “validity”, “variance”, “effect size”, “predictive value”, and “AUC“. As well as a search for “PACC”, to check if any mentions of PACC was related to a discussion of its accuracy. However, this led to exclusions of studies that might have investigated PACC accuracy in different ways. For instance, many excluded studies used PACC as something they could compare their results of another

intervention to, to determine any effect of their intervention in question. Here, PACC becomes the measurement tool in an investigation. So, while these studies use PACC to measure cognitive decline in a longitudinal study design, their primary aim is not to validate PACC itself, but to use it as a tool to understand the impact of different risk factors. In general, there was a great lack of sensitivity analyses across the studies. Many studies only briefly discussed the accuracy of PACC, which might lead to bias in the data analysis. The data from included studies on PACC sensitivity is already very varying, making it challenging to compare the results. However, if every article that used PACC as a measurement method had been included as well, it could have potentially overwhelmed the scope and focus of this paper. Priorities are important in systematic reviews, where one must include studies that are directly relevant to the research question, which helps gain clear and useful answers in the end.

The decision to exclude studies that did not clarify the reference test was based on a methodological approach designed to ensure the consistency and reliability of the findings. This choice aimed to include only studies that provided clear and precise details on the use of Aβ-PET imaging, which is crucial for accurately determining Aβ pathology in participants. Although this approach might lead to the exclusion of potentially valuable studies and introduce selection bias, it was necessary to maintain a consistent foundation for comparison. Despite the availability of alternative methods to measure Aβ pathology, such as CSF or plasma samples, focusing on Aβ-PET imaging ensured a more comparable baseline for pathology measurements. This focus also adhered to the DTA protocol for the reference test requirements. The exclusion of studies investigating tau pathology followed similar reasoning, aiming to maintain methodological consistency and reliability in the review.

Originally, multiple studies conducting research on the same datasets were meant to be excluded. However, during the full-text screening, it was found that these studies contributed to a better

understanding of the research question in different ways. Despite using the same datasets or groups of participants, each study provided unique insights and addressed different aspects of the diagnostic accuracy and sensitivity of PACC. This allowed for a comprehensive analysis of the dataset, where various perspectives were presented. This also underlines the consistency and reliability of the findings, as different studies arrived at similar conclusions. However, including multiple studies using the same datasets can also lead to redundancy, where similar findings are repeated, thus ending up overemphasizing certain results. Multiple analyses of the same dataset might confirm the same hypotheses, leading to confirmation bias. Besides this, several biases can occur when using datasets. Firstly, the fact that participants were volunteers might introduce volunteer bias, as those who choose to participate could have different characteristics from those who do not. Secondly, the datasets included limited variation in ethnicity, which can restrict the diversity represented in these studies, potentially impacting the generalizability of the findings. Additionally, the studies often included highly educated individuals, which may not reflect the broader population, leading to educational bias. Lastly, the heterogeneity among the datasets, stemming from variations in study design, population characteristics, measurement techniques, and interventions, can subsequently pose limitations as it complicates the ability to draw consistent and definitive conclusions from the results.

The exclusion due to participants being of the "wrong population" were based on participants with co-existing illnesses (for instance, vascular disease, heart problems, familial AD, diagnosis of AD, brain atrophy or individuals receiving some kind of medication). These criteria may limit the inclusivity and generalizability of the findings, potentially excluding important data that could offer a more nuanced understanding of PACC's diagnostic accuracy across different populations and their backgrounds. However, some of these conditions can independently affect cognitive function, making it difficult to isolate the effects of Aβ pathology in cognitive decline measurements. Also, even though atrophy is a known consequence of AD, it typically manifests in the later stages of the disease and

does not align with the focus on preclinical AD in this review. Furthermore, participants in placebo-drug trials were excluded from this study, even though the placebo groups are essentially composed of CN individuals. However, these groups may still experience placebo effects or other influences related to the trial protocols, which can confound the study results. The aim of the review is to understand the natural progression of cognitive decline related to amyloid pathology, and including participants from placebo-drug trials could affect the natural course of the disease.

5.4.2 Errors within the search strategy and data collection

Errors can arise both within the reviewed articles and in the review process itself. Random errors, such as data entry mistakes or variability in measurement techniques, can affect the accuracy of study results. Specific errors, such as misinterpretation of results or selective reporting, can introduce bias. It is crucial to be transparent about these uncertainties to introduce reliability. In this case, as there is only one reviewer, the potential errors are even more likely. All studies were screening and read by a single reviewer, which might lead to idiosyncratic errors as well as selection bias. A single reviewer may consciously or unconsciously select studies that confirm their pre-existing beliefs or hypotheses, thus skewing the results of the review. Additionally, a single person's interpretation is inherently narrower than a collective one. The lack of blinding across the different aspect of the review similarly introduces bias in both the selection of studies and the interpretation.

The systematic search strategy and standardized data extraction ensured consistency across the studies. Following approved guidelines such as the PRISMA guidelines and the DTA protocol from JBI, a systematic and reliable search helps to enhance the quality of the review. However, the lack of a review protocol and prior registration due to time constraints is a methodological limitation. Some parts of the PRISMA guidelines were not followed, including reporting effect measures, risk ratios, and mean differences for each study outcome, as many studies did not provide this information and

the overall time frame did not allow for these steps. This could affect the transparency and reproducibility of the review, potentially limiting the completeness of the reported findings. Furthermore, the DTA protocol and the PRISMA guidelines for the abstract provided conflicting information, with the main difference being whether to use an abstract for a systematic review or an abstract for a DTA review protocol. A combination was chosen.

Collaborating with librarians at Aalborg University Library additionally helped enhance the quality of the search. While the search strategy was comprehensive, covering multiple databases and using broad search terms, it was limited to studies accessible through the three chosen databases: PubMed, Embase, and PsycInfo. This could have excluded relevant studies published in other databases.

5.4.3 Assessing QUADAS-2

The QUADAS-2 tool was used to assess the quality of included studies, as this tool is recommended by JBI when conducting DTA reviews. However, there are some limitations to this assessment tool. Firstly, the tool did not have a numerical score value to determine the overall quality score of each study. This was done manually to better compare the results of the quality scores, which resulted in the weighting of “no” and “unknown” as equally 0, and “yes” as 1. This might be an oversimplification of the quality assessment. The lack of numerical scoring in QUADAS-2 and the manual scoring of quality done by a single reviewer, might introduce subjectivity and potential bias in evaluating the study quality.

Additionally, the questions regarding the quality assessment might have some limitations. The question regarding whether a consecutive or random sample of patients was enrolled ensures systematic sampling but does not address volunteer bias or the representativeness of the sample. Although avoiding a case-control design helps reduce retrospective bias, this question does not

eliminate the potential for selection bias in other study designs. The question about avoiding inappropriate exclusions is crucial, yet it may miss other exclusions based on factors that the reviewer does not deem to be “inappropriate”. The question on whether the index test results were interpreted without knowledge of the reference standard ensures objectivity but overlooks potential biases from the test administrators themselves. Ensuring thresholds are pre-specified is important for consistency, yet this question does not assess the difference in thresholds used across different studies. The question assuming the reference standard correctly classifies the target condition might not account for the varying accuracy of the reference standard, which can affect study outcomes. Ensuring that reference standard results were interpreted without knowledge of the index test results promotes objectivity, but in a DTA review, this question is less relevant because the reference test is typically conducted prior to the index test. The question about whether there was an appropriate interval between the index test and reference standard, does not define what an “appropriate interval” is, and this question also neglects the impact of differing disease progression rates among participants. While the question about all patients receiving the same reference standard promotes consistency, it does not account for the possibility that different researchers might interpret the reference standard differently depending on their training, and the question also does not consider the variations within the same reference standard (18F-FBP-PET / 11C-PiB-PET /18F-FLUTE-PET). Lastly, including all patients in the analysis does not state, whether this was before or after any inappropriate exclusions, which might lead to biased or incomplete results.

5.5 PACC compared to Aβ-PET imaging

In this study, the diagnostic accuracy and sensitivity of PACC was evaluated relative to Aβ pathology as measured by Aβ-PET imaging for the early detection of AD. Both PACC and Aβ-PET imaging serve as crucial tools in identifying preclinical stages of AD, yet they offer distinct advantages and face specific limitations.

PACC, a composite of four neuropsychological tests, has demonstrated high sensitivity in detecting early cognitive changes linked to elevated Aβ pathology. The composite includes the subtests, MMSE, DSST, FCSRT, and LM-DR, with the MMSE and DSST being particularly sensitive in identifying subtle cognitive decline before significant symptoms manifest. PACC also has an ability to detect early cognitive decline across different stages of AD, from preclinical to MCI and finally AD. This underscores its potential utility in both clinical trials and routine clinical settings. Additionally, PACC is a non-invasive and cost-effective tool, making it accessible for widespread use. However, its effectiveness could be further enhanced by incorporating a language-focused subtest, such as the FAS or CAT test, to provide a more comprehensive evaluation of cognitive decline. Lastly, PACC presents results that clearly demonstrate which cognitive functions are declining.

In contrast, Aβ-PET imaging directly measures Aβ plaques in the brain, which are among the earliest biomarkers of AD. This specific imaging technique involves injecting radioactive tracers that bind to Aβ deposits, allowing for the visualization and quantification of Aβ pathology. Aβ-PET is considered the gold standard for detecting Aβ pathology due to its high specificity and ability to provide detailed images of Aβ distribution in the brain. However, while Aβ-PET can identify the locations of Aβ plaques, it is less sensitive in determining which cognitive functions might be affected. Aβ-PET imaging is also an invasive and costly procedure, which may limit its accessibility and widespread use. Furthermore, Aβ-PET imaging requires specialized equipment and expertise, which may not be available in all clinical settings. Just as importantly, an elevated Aβ level is not a guarantee for AD development. It is just a likely outcome.

When comparing PACC to Aβ-PET, several strengths and weaknesses emerge for each method. PACC excels in its non-invasive nature, ease of administration, and cost-effectiveness, making it suitable for regular monitoring and early detection in various settings. Its sensitivity in identifying

early cognitive changes makes it a valuable tool for detecting preclinical AD. However, PACC relies on cognitive testing, which can be influenced by factors such as education, language proficiency, and other individual differences, potentially affecting its accuracy. On the other hand, Aβ-PET offers direct measurement, and direct confirmation, of any existing Aβ pathology. This provides a clear and objective indication of Aβ deposition in the brain. This specificity makes it a powerful diagnostic tool for confirming the presence of Aβ plaques.

In conclusion, both PACC and Aβ-PET imaging play critical roles in detecting preclinical AD, each with unique advantages and challenges. Perhaps a conjoined approach, utilizing both tools, could offer the most comprehensive assessment of preclinical AD, combining the strengths of each method to enhance early detection and management of the disease. Together, these tools offer complementary insights, enhancing our ability to detect and manage AD at its earliest stages. Both are highly sensitive measurement methods for preclinical AD detection.

5.6 Further research

Further exploration of PACC's potential in clinical settings is essential. Future research should address the challenges of implementing PACC in primary care, where specialized training for test administrators and streamlined scoring methods are needed to make PACC more accessible in general practice. New research is already being conducted on the blood-based biomarker 'p-tau217' and the drug lecanemab, which slows down the progression of AD. Integrating PACC with these emerging diagnostic methods and early drug interventions could significantly slow disease progression and reduce the costs associated with expensive PET imaging. This combination would facilitate easier and earlier diagnosis of preclinical AD, making it more accessible to broader populations.

6.0 Conclusion

This systematic review aimed to evaluate the diagnostic accuracy and sensitivity of the Preclinical Alzheimer's Cognitive Composite (PACC) relative to amyloid pathology (Aβ) as determined by amyloid-PET imaging for the early detection of Alzheimer's Disease (AD). The increasing prevalence of AD and the critical need for early detection were primary motivators for this review. Additionally, groundbreaking research on the AD-slowing drug, lecanemab, and emerging methods for measuring Aβ in blood, further highlights the timeliness and relevance of this study.

Twenty studies were included in this systematic review, adhering to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and the Diagnostic Test Accuracy (DTA) protocol by the Joanna Briggs Institute. The findings indicate that PACC is a highly sensitive tool for detecting early cognitive changes associated with elevated Aβ pathology. Specifically, the Mini Mental State Examination (MMSE) and Digit Symbol Substitution Test (DSST) subtests demonstrated high sensitivity in identifying cognitive decline in both the early and late stages of preclinical AD. Studies consistently showed that even low levels of Aβ pathology detected via Aβ-PET imaging are linked with subtle cognitive decline measurable by PACC. This correlation underlines Aβ as one of the first pathological biomarkers appearing in the disease trajectory towards AD. PACC's ability to differentiate between preclinical AD, Mild Cognitive Impairment (MCI), and AD suggests its broad utility across different stages of cognitive decline. This makes it a valuable tool in both clinical trials and clinical settings. The review supports the implementation of PACC as a non-invasive, cost-effective tool for early detection of AD. Given its high sensitivity, PACC can be a valuable addition to clinical practice for identifying individuals at risk of cognitive decline before significant symptoms manifest. This early detection is crucial for timely intervention, potentially slowing the progression of the disease and improving patient outcomes.

Despite different versions of PACC demonstrating varying degrees of sensitivity and specificity, higher levels of Aβ were consistently associated with cognitive decline across all studies in the review. This consistent relationship between elevated Aβ levels and cognitive decline highlights the diagnostic robustness of PACC. However, the inclusion of a language subtest, such as the Controlled Oral Word Association Test (FAS) or Category Fluency Test (CAT), might enhance PACC's overall sensitivity. This is supported by findings showing that the inclusion of CAT enhanced PACC and that a larger effect size was found for the ZAVEN composite compared to PACC, indicating that language measurements might be lacking in the original PACC composite.

Overall, this review underscores the importance of early detection in AD and highlights the potential of PACC as a sensitive and accurate tool for identifying early cognitive changes. By advancing our diagnostic capabilities, PACC can play a central role in the early identification and management of AD, ultimately improving patient outcomes and contributing to the optimization of healthcare resources.

7.0 Reference list

***** = Reviewed articles

Ahmad, F., Javed, M., Athar, M., & Shahzadi, S. (2023). Determination of affected brain regions at various stages of Alzheimer's disease. *Neuroscience Research*, *192*, 77–82. https://doi.org/10.1016/j.neures.2023.01.010

Ahn, Y. D., Yi, D., Joung, H., Seo, E. H., Lee, Y. H., Byun, M. S., Lee, J. H., Jeon, S. Y., Lee, J. Y., Sohn, B. K., & Lee, D. Y. (2019). Normative data for the logical memory subtest of the wechsler memory scale-iv in middle-aged and elderly korean people. *Psychiatry Investigation*, *16*(11), 793–799. https://doi.org/10.30773/pi.2019.0061

Alzheimer's Association. (2024). *Alzheimer's Association 2024 Alzheimer's Disease Facts and Figures*. alzheimers-facts-and-figures.pdf

Aromataris, E., & Munn, Z. (2020). *JBI Manual for Evidence Synthesis JBI, 2020*. https://doi.org/10.46658/JBIMES-24-01

Aromataris, E., & Pearson, A. (2014). The systematic review: an overview. *AJN The American Journal of Nursing*, *114*(3), 53–58. https://doi.org/10.1097/01.NAJ.0000444496.24228.2c

Ashton, N. J., Brum, W. S., Molfetta, G. Di, Benedet, A. L., Arslan, B., Jonaitis, E., Langhough, R. E., Cody, K., Wilson, R., Carlsson, C. M., Vanmechelen, E., Montoliu-Gaya, L., Lantero-Rodriguez, J., Rahmouni, N., Tissot, C., Stevenson, J., Servaes, S., Therriault, J., Pascoal, T., … Zetterberg, H. (2024). Diagnostic Accuracy of a Plasma Phosphorylated Tau 217 Immunoassay for Alzheimer Disease Pathology. *JAMA Neurology*, *81*(3), 255–263. https://doi.org/10.1001/jamaneurol.2023.5319

Babic, A., Tokalic, R., Amílcar Silva Cunha, J., Novak, I., Suto, J., Vidak, M., Miosic, I., Vuka, I., Poklepovic Pericic, T., & Puljak, L. (2019). Assessments of attrition bias in Cochrane systematic reviews are highly inconsistent and thus hindering trial comparability. *BMC Medical Research Methodology*, *19*(1). https://doi.org/10.1186/s12874-019-0717-9

Bartlett, J. (2020, 30th of December). *Mixed model repeated measures (MMRM) in Stata, SAS, and R.* The Stats Geek. https://thestatsgeek.com/2020/12/30/mixed-model-repeated-measures-mmrm-in-stata-sas-and-r/

Berg, L. (1984). Clinical dementia rating. *The British Journal of Psychiatry*, *145*(3), 339. https://doi.org/10.1192/S0007125000118082

Booth A., Sutton, A., Clowes, M. & James, M. M. (2022). Systematic approaches to a successful literature review. (3rd ed.). SAGE publications.

* Bransby, L., Lim, Y. Y., Ames, D., Fowler, C., Roberston, J., Harrington, K., Snyder, P. J., Villemagne, V. L., Salvado, O., Masters, C. L., & Maruff, P. (2019). Sensitivity of a Preclinical Alzheimer's Cognitive Composite (PACC) to amyloid β load in preclinical Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, *41*(6), 591–600. https://doi.org/10.1080/13803395.2019.1593949

* Buckley, R. F., Schultz, A. P., Hedden, T., Papp, K. V, Hanseeuw, B. J., Marshall, G., Sepulcre, J., Smith, E. E., Rentz, D. M., & Johnson, K. A. (2017). Functional network integrity presages cognitive decline in preclinical Alzheimer disease. *Neurology*, *89*(1), 29–37. https://doi.org/10.1212/WNL.0000000000004059

Campbell, J. M., Kulgar, M., Ding, S., Carmody, D. P., Hakonsen, S. J., Jadotte, Y. T., & Ws, C. (2020). Diagnostic test accuracy systematic reviews. *JBI Manual for Evidence Synthesis*, *1*. https://doi.org/10.1097/XEB.0000000000000061

* Demnitz-King, H., Requier, F., Whitfield, T., Schlosser, M., Gonneaud, J., Ware, C., Barnhofer, T., Coll-Padros, N., Dautricourt, S., Delarue, M., Klimecki, O. M., Paly, L., Salmon, E., Schild, A. K., Wirth, M., Frison, E., Lutz, A., Chételat, G., Collette, F., & Marchant, N. L. (2023). Effects of Meditation Training and Non-Native Language Training on Cognition in Older Adults: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Network Open*, *6*(7), E2317848. https://doi.org/10.1001/jamanetworkopen.2023.17848

* Donohue, M. C., Sperling, R. A., Petersen, R., Sun, C. K., Weiner, M., & Aisen, P. S. (2017). Association between elevated brain amyloid and subsequent cognitive decline among cognitively normal persons. *JAMA - Journal of the American Medical Association*, *317*(22), 2305–2316. https://doi.org/10.1001/jama.2017.6669

Donohue, M. C., Sperling, R. A., Salmon, D. P., Rentz, D. M., Raman, R., Thomas, R. G., Weiner, M., & Aisen, P. S. (2014). The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. *JAMA Neurology*, *71*(8), 961–970. https://doi.org/10.1001/jamaneurol.2014.803

* Donohue, M. C., Sun, C. K., Raman, R., Insel, P. S., & Aisen, P. S. (2017). Cross-validation of optimized composites for preclinical Alzheimer's disease. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, *3*(1), 123–129. https://doi.org/10.1016/j.trci.2016.12.001

* Farrell, M. E., Jiang, S., Schultz, A. P., Properzi, M. J., Price, J. C., Becker, J. A., Jacobs, H. I. L., Hanseeuw, B. J., Rentz, D. M., & Villemagne, V. L. (2021). Defining the lowest threshold for amyloid-PET to predict future cognitive decline and amyloid accumulation. *Neurology*, *96*(4), e619–e631. https://doi.org/10.1212/WNL.0000000000011214

* Farrell, M. E., Papp, K. V., Buckley, R. F., Jacobs, H. I. L., Schultz, A. P., Properzi, M. J., Vannini, P., Hanseeuw, B. J., Rentz, D. M., Johnson, K. A., & Sperling, R. A. (2022). Association of

Emerging β-Amyloid and Tau Pathology with Early Cognitive Changes in Clinically Normal Older Adults. *Neurology*, *98*(15), E1512–E1524. https://doi.org/10.1212/WNL.0000000000200137

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

Grant, A. (2023). FDA grants accelerated approval for alzheimer's disease treatment. *US Food and Drug Administration: Rockville, MD, USA*. https://www.fda.gov/news-events/press-announcements/fda-grants-accelerated-approval-alzheimers-disease-treatment

Grober, E., Hall, C. B., Lipton, R. B., Zonderman, A. B., Resnick, S. M., & Kawas, C. (2008). Memory impairment, executive dysfunction, and intellectual decline in preclinical Alzheimer's disease. *Journal of the International Neuropsychological Society*, *14*(2), 266–278. https://doi.org/10.1017/S1355617708080302

Gustavsson, A., Norton, N., Fast, T., Frölich, L., Georges, J., Holzapfel, D., Kirabali, T., Krolak-Salmon, P., Rossini, P. M., & Ferretti, M. T. (2023). Global estimates on the number of persons across the Alzheimer's disease continuum. *Alzheimer's & Dementia*, *19*(2), 658–670. https://doi.org/10.1002/alz.12694

* Hampton, O. L., Mukherjee, S., Properzi, M. J., Schultz, A. P., Crane, P. K., Gibbons, L. E., Hohman, T. J., Maruff, P., Lim, Y. Y., Amariglio, R. E., Papp, K. V., Johnson, K. A., Rentz, D. M., Sperling, R. A., & Buckley, R. F. (2022). Harmonizing the Preclinical Alzheimer Cognitive Composite for Multicohort Studies. *Neuropsychology*, *37*(4), 436–449. https://doi.org/10.1037/neu0000833

Hassenstab, J., Nicosia, J., LaRose, M., Aschenbrenner, A. J., Gordon, B. A., Benzinger, T. L. S., Xiong, C., & Morris, J. C. (2021). Is comprehensiveness critical? Comparing short and long format cognitive assessments in preclinical Alzheimer disease. *Alzheimer's Research & Therapy*, *13*, 1–14. https://doi.org/10.1186/s13195-021-00894-5

Ingram, O. (2023, 10th of July). *What is recall bias – causes and examples*. ResearchProspect. https://www.researchprospect.com/what-is-recall-bias/

* Insel, P. S., Donohue, M. C., Berron, D., Hansson, O., & Mattsson-Carlgren, N. (2021). Time between milestone events in the Alzheimer's disease amyloid cascade. *NeuroImage*, *227*. https://doi.org/10.1016/j.neuroimage.2020.117676

* Insel, P. S., Donohue, M. C., Sperling, R., Hansson, O., & Mattsson-Carlgren, N. (2020). The A4 study: β-amyloid and cognition in 4432 cognitively unimpaired adults. *Annals of Clinical and Translational Neurology*, *7*(5), 776–785. https://doi.org/10.1002/acn3.51048

Irwin, K., Sexton, C., Daniel, T., Lawlor, B., & Naci, L. (2018). Healthy aging and dementia: Two roads diverging in midlife? *Frontiers in Aging Neuroscience*, *10*, 275. https://doi.org/10.3389/fnagi.2018.00275

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C., & Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, *9*(1), 119–128. https://doi.org/10.1016/S1474-4422(09)70299-6

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., & Karlawish, J. (2018). NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

Jaeger, J. (2018). Digit symbol substitution test. *Journal of Clinical Psychopharmacology*, *38*(5), 513–519. https://doi.org/10.1097/JCP.0000000000000941

Jia, X., Wang, Z., Huang, F., Su, C., Du, W., Jiang, H., Wang, H., Wang, J., Wang, F., Su, W., Xiao, H., Wang, Y., & Zhang, B. (2021). A comparison of the Mini-Mental State Examination (MMSE) with the Montreal Cognitive Assessment (MoCA) for mild cognitive impairment screening in Chinese middle-aged and older population: a cross-sectional study. *BMC Psychiatry*, *21*(1). https://doi.org/10.1186/s12888-021-03495-6

* Jutten, R. J., Rentz, D. M., Fu, J. F., Mayblyum, D. V., Amariglio, R. E., Buckley, R. F., Properzi, M. J., Maruff, P., Stark, C. E., Yassa, M. A., Johnson, K. A., Sperling, R. A., & Papp, K. V. (2022). Monthly At-Home Computerized Cognitive Testing to Detect Diminished Practice Effects in Preclinical Alzheimer's Disease. *Frontiers in Aging Neuroscience*, *13*. https://doi.org/10.3389/fnagi.2021.800126

Koychev, I., Hofer, M., & Friedman, N. (2020). Correlation of Alzheimer disease neuropathologic staging with amyloid and tau scintigraphic imaging biomarkers. *Journal of Nuclear Medicine*, *61*(10), 1413–1418. https://doi.org/10.2967/jnumed.119.230458

Leonardo, R. (2018). PICO: model for clinical questions. *Evid Based Med Pract*, *3*(115), 2. https://doi.org/10.4172/2471-9919.1000115

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of Internal Medicine*, *151*(4), W-65. https://doi.org/10.1016/j.jclinepi.2009.06.006

* Lim, Y. Y., Snyder, P. J., Pietrzak, R. H., Ukiqi, A., Villemagne, V. L., Ames, D., Salvado, O., Bourgeat, P., Martins, R. N., & Masters, C. L. (2016). Sensitivity of composite scores to amyloid burden in preclinical Alzheimer's disease: Introducing the Z-scores of Attention, Verbal fluency, and Episodic memory for Nondemented older adults composite score. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *2*, 19–26. https://doi.org/10.1016/j.dadm.2015.11.003

* Lu, K., Nicholas, J. M., Collins, J. D., James, S. N., Parker, T. D., Lane, C. A., Keshavan, A., Keuss, S. E., Buchanan, S. M., Murray-Smith, H., Cash, D. M., Sudre, C. H., Malone, I. B., Coath, W., Wong, A., Henley, S. M. D., Crutch, S. J., Fox, N. C., Richards, M., & Schott, J. M. (2019). Cognition at age 70: Life course predictors and associations with brain pathologies. *Neurology*, *93*(23), E2144–E2156. https://doi.org/10.1212/WNL.0000000000008534

Ma, C., Hong, F., & Yang, S. (2022). Amyloidosis in Alzheimer's disease: Pathogeny, etiology, and related therapeutic directions. *Molecules*, *27*(4), 1210. https://doi.org/10.3390/molecules27041210

Mayo Clinic Staff. (2024, 13th of March). *Diagnosing Alzheimer's: How Alzheimer's is diagnosed. https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers/art-20048075*

Metz, C. E. (1978). *Basic Principles of ROC Analysis*. https://doi.org/10.1016/s0001-2998(78)80014-2

Mitchell, A. J. (2009). A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *Journal of Psychiatric Research*, *43*(4), 411–431. https://doi.org/10.1016/j.jpsychires.2008.04.014

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

* Mormino, E. C., Papp, K. V, Rentz, D. M., Donohue, M. C., Amariglio, R., Quiroz, Y. T., Chhatwal, J., Marshall, G. A., Donovan, N., & Jackson, J. (2017). Early and late change on the preclinical Alzheimer's cognitive composite in clinically normal older individuals with elevated amyloid β. *Alzheimer's & Dementia*, *13*(9), 1004–1012. https://doi.org/10.1016/j.jalz.2017.01.018

National Institute on Aging. (2024, 11th of January). *What happens in the brain in Alzheimer's Disease?* https://www.nia.nih.gov/health/alzheimers-causes-and-risk-factors/what-happens-brain-alzheimers-disease

National Health Service. (2023, 24th of July). *What are the treatments for dementia? https://www.nhs.uk/conditions/dementia/about-dementia/treatment/*

Ossenkoppele, R., Jansen, W. J., Rabinovici, G. D., Knol, D. L., van der Flier, W. M., van Berckel, B. N. M., Scheltens, P., Visser, P. J. elle, Verfaillie, S. C. J., Zwan, M. D., Adriaanse, S. M., Lammertsma, A. A., Barkhof, F., Jagust, W. J., Miller, B. L., Rosen, H. J., Landau, S. M., Villemagne, V. L., Rowe, C. C., … Brooks, D. J. (2015). Prevalence of amyloid PET positivity in dementia syndromes: a meta-analysis. *JAMA*, *313*(19), 1939–1949. https://doi.org/10.1001/jama.2015.4669

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic reviews, 5,* 1-10. https://doi.org/10.1186/s13643-016-0384-4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*, *372*. https://doi.org/10.1136/bmj.n71

* Papp, K. V., Rentz, D. M., Orlovsky, I., Sperling, R. A., & Mormino, E. C. (2017). Optimizing the preclinical Alzheimer's cognitive composite with semantic processing: The PACC5. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, *3*(4), 668–677. https://doi.org/10.1016/j.trci.2017.10.004

Pavisic, I. M., Lu, K., Keuss, S. E., James, S.-N., Lane, C. A., Parker, T. D., Keshavan, A., Buchanan, S. M., Murray-Smith, H., & Cash, D. M. (2021). Subjective cognitive complaints at age 70: associations with amyloid and mental health. *Journal of Neurology, Neurosurgery & Psychiatry*, *92*(11), 1215–1221. https://doi.org/10.1136/jnnp-2020-325620

Perestelo-Pérez, L. (2013). Standards on how to develop and report systematic reviews in Psychology and Health. *International Journal of Clinical and Health Psychology*, *13*(1), 49–57. https://doi.org/10.1016/S1697-2600(13)70007-3

Peters, U. (2022). What is the function of confirmation bias? *Erkenntnis*, *87*(3), 1351–1376. https://doi.org/10.1007/s10670-020-00252-1

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons. https://doi.org/10.1002/9780470754887

* Rentz, D. M., Papp, K. V., Mayblyum, D. V., Sanchez, J. S., Klein, H., Souillard-Mandar, W., Sperling, R. A., & Johnson, K. A. (2021). Association of Digital Clock Drawing With PET Amyloid and Tau Pathology in Normal Older Adults. *Neurology*, *96*(14), E1844–E1854. https://doi.org/10.1212/WNL.0000000000011697

* Ruthirakuhan, M., Wood Alexander, M., Cogo-Moreira, H., Robinson, T., Amariglio, R., Buckley, R. F., Sperling, R. A., Swardfager, W., Black, S. E., & Rabin, J. S. (2024). Investigating the Factor Structure of the Preclinical Alzheimer Cognitive Composite and Cognitive Function Index across Racial/Ethnic, Sex, and Aβ Status Groups in the A4 Study. *Journal of Prevention of Alzheimer's Disease*, *11*(1), 48–55. https://doi.org/10.14283/jpad.2023.98

* Sato, K., Mano, T., Ihara, R., Suzuki, K., Niimi, Y., Toda, T., Iwatsubo, T., & Iwata, A. (2021). Cohort-Specific Optimization of Models Predicting Preclinical Alzheimer's Disease, to Enhance Screening Performance in the Middle of Preclinical Alzheimer's Disease Clinical Studies. *Journal of Prevention of Alzheimer's Disease*, *8*(4), 503–512. https://doi.org/10.14283/jpad.2021.39

Šimundić, A. M. (2013). Bias in research. *Biochemia Medica*, *23*(1), 12–15. https://doi.org/10.11613/BM.2013.003

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack Jr, C. R., Kaye, J., & Montine, T. J. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 280–292. https://doi.org/10.1016/j.jalz.2011.03.003

Sperling, R., Mormino, E., & Johnson, K. (2014). The evolution of preclinical Alzheimer's disease: implications for prevention trials. *Neuron*, *84*(3), 608–622. https://doi.org/10.1016/j.neuron.2014.10.038

* Sperling, R. A., Donohue, M. C., Raman, R., Sun, C.-K., Yaari, R., Holdridge, K., Siemers, E., Johnson, K. A., Aisen, P. S., A4 Study Team. (2020). Association of factors with elevated

amyloid burden in clinically normal older individuals. *JAMA Neurology,* 77(6), 735-745. https://doi.org/10.1001/jamaneurol.2020.0387

Spruance, S. L., Reid, J. E., Grace, M., & Samore, M. (2004). Hazard ratio in clinical trials. *Antimicrobial Agents and Chemotherapy*, *48*(8), 2787–2792. https://doi.org/10.1128/AAC.48.8.2787-2792.2004

* Stricker, N. H., Twohy, E. L., Albertson, S. M., Karstens, A. J., Kremers, W. K., Machulda, M. M., Fields, J. A., Jack, C. R., Knopman, D. S., Mielke, M. M., & Petersen, R. C. (2023). Mayo-PACC: A parsimonious preclinical Alzheimer's disease cognitive composite comprised of public-domain measures to facilitate clinical translation. *Alzheimer's and Dementia*, *19*(6), 2575–2584. https://doi.org/10.1002/alz.12895

Turkington, T. G. (2001). Introduction to PET instrumentation. *Journal of Nuclear Medicine Technology*, *29*(1), 4–11. 4.full.pdf (snmjournals.org)

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M. G., Sterne, J. A. C., Bossuyt, P. M. M., & Group*, Q.-2. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, *155*(8), 529–536. http://doi.org/10.7326/0003-4819-155-8-201110180-00009

World Health Organization. (2019). *International statistical classification of diseases and related health problems* (11th ed.). https://icd.who.int/

Winblad, B., Amouyel, P., Andrieu, S., Ballard, C., Brayne, C., Brodaty, H., Cedazo-Minguez, A., Dubois, B., Edvardsson, D., Feldman, H., Fratiglioni, L., Frisoni, G. B., Gauthier, S., Georges, J., Graff, C., Iqbal, K., Jessen, F., Johansson, G., Jönsson, L., … Zetterberg, H. (2016). Defeating Alzheimer's disease and other dementias: a priority for European science and society. *The Lancet Neurology*, *15*(5), 455–532. https://doi.org/10.1016/S1474-4422(16)00062-4

Young, P. N. E., Estarellas, M., Coomans, E., Srikrishna, M., Beaumont, H., Maass, A., Venkataraman, A. V, Lissaman, R., Jiménez, D., Betts, M. J., McGlinchey, E., Berron, D., O'Connor, A., Fox, N. C., Pereira, J. B., Jagust, W., Carter, S. F., Paterson, R. W., & Schöll, M. (2020). Imaging biomarkers in neurodegeneration: current and future practices. *Alzheimer's Research & Therapy*, *12*(1), 49. https://doi.org/10.1186/s13195-020-00612-7

Pages in total = 1687

Reused Curriculum

Andersen, H. & Koch, L. (2015). Hermeneutik og fænomenologi. In M. H. Jacobsen, K. Lippert-Rasmussen & P. Nedergaard (Eds.). *Videnskabsteori i statskundskab, sociologi og forvaltning* (3rd ed.). Pp. 205-250. Hans Reitzel.

Breedlove, S. M., & Watson, N. V. (2013). *Biological psychology: An introduction to behavioral, cognitive, and clinical neuroscience*. Sinauer Associates. [Chapters 7+17]

Field, A. (2017). *Discovering Statistics Using IBM SPSS Statistics*. (5th ed). SAGE. [Chapters 3+18+19]

Ivanouw, J. (2006). Psykometri. In P. Elsass, J. Ivanouw, E. L. Mortensen, S. Poulsen, & B. Rosenbaum (Eds.), Assessmentmetoder: Håndbog for psykologer og psykiatere (pp. 97-158). Dansk Psykologisk Forlag.

Vengsgaard, K. (2012). Kritisk rationalisme. In M. H. Jacobsen, K. Lippert-Rasmussen & P. Nedergaard (Eds.), *Videnskabsteori-i Statskundskab, Sociologi Og Forvaltning*, *2*, 139–173.

Pages in total = 374