LUND UNIVERSITY

The future of clinical trials for Alzheimer's disease. A blood-based biomarker perspective

Cullen, Nicholas

2022

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Cullen, N. (2022). *The future of clinical trials for Alzheimer's disease. A blood-based biomarker perspective*. [Doctoral Thesis (compilation), Department of Clinical Sciences, Malmö]. Lund University, Faculty of Medicine.

Total number of authors:
1

Creative Commons License:
CC BY

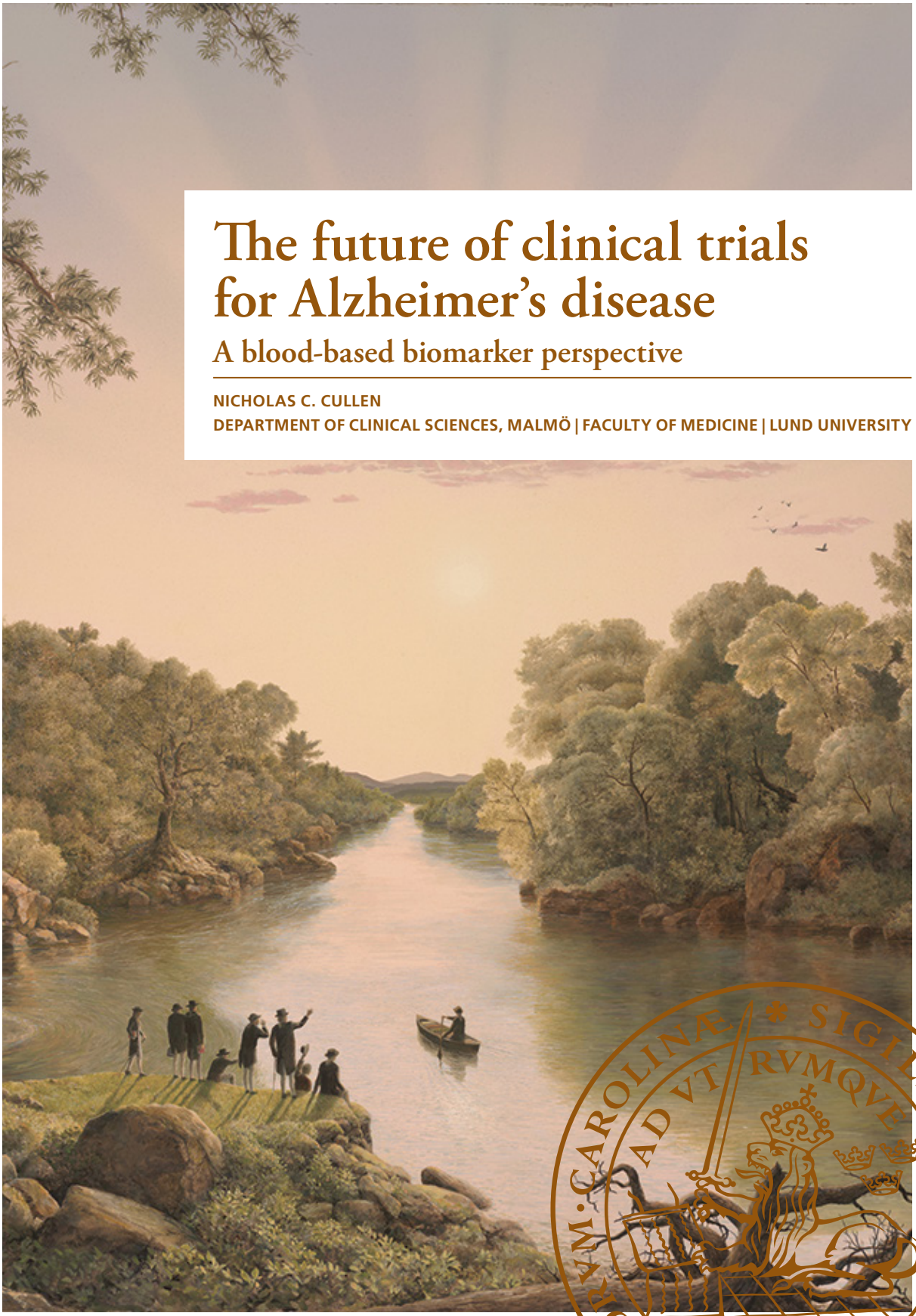
General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



The future of clinical trials for Alzheimer's disease

A blood-based biomarker perspective

NICHOLAS C. CULLEN
DEPARTMENT OF CLINICAL SCIENCES, MALMÖ | FACULTY OF MEDICINE | LUND UNIVERSITY

The future of clinical trials for Alzheimer's disease

A blood-based biomarker perspective

The future of clinical trials for Alzheimer's disease

A blood-based biomarker perspective

by Nicholas C. Cullen

LUND
UNIVERSITY

DOCTORAL DISSERTATION

By due permission of the Faculty of Medicine, Lund University, Sweden.

To be defended at Belfragesalen, on 30 September 2022 at 13:00

Faculty opponent
Masud Husain, University of Oxford

Organization LUND UNIVERSITY	**Document name** DOCTORAL DISSERTATION
Department of Clinical Sciences, Malmö University, Sölvegatan 19 – BMC G12, 221 84 Lund, Sweden	**Date of issue** 2022-09-30
Author Nicholas C. Cullen	Sponsoring organization

Title and subtitle

The future of clinical trials for Alzeimer's disease
A blood-based biomarker perspective

Abstract

Objectives: The primary objective was to investigate the utility of blood-based biomarkers of amyloid, tau, and neurodegeneration for (i) screening, (ii) enrichment, and (iii) tracking response to treatment in clinical trials of Alzheimer's disease.

Methods: Longitudinal, participant-level data used in these studies was drawn from the Swedish BioFINDER study and the ADNI study. Participants were classified as cognitively unimpaired, mild cognitive impairment, or Alzheimer's disease dementia. For screening, logistic regression was used to predict amyloid PET status in CU individuals from plasma Aβ42/Aβ40, *APOE* status, and age. For enrichment, Linear mixed effects models were used to predict longitudinal cognitive decline and future risk of AD dementia in CU individuals or in MCI individuals from a basic model (age, sex, education, *APOE* status) and varying combinations of blood-based biomarkers (plasma Aβ42/Aβ40, plasma pTau181, plasma pTau217, plasma NfL). For treatment response, plasma NfL was measured longitudinally in MCI or AD patients and properties such as slope, inter-subject variability, and intra-subject variability were calculated. Plasma NfL was then compared with MRI and cognition.

Results: The amyloid PET screening model had an AUC of 0.87, with a significant independent effect for plasma Aβ42/Aβ40 and *APOE* status, but not age. This model was estimated to reduce total cost of recruiting 500 amyloid-positive CU participants by 31 – 42%, depending on the relative cost of amyloid scanning to plasma measurement. For enrichment, plasma pTau181 and pTau217 had the largest effect on predicting cognitive decline in CU and MCI participants, with Aβ42/Aβ40 and NfL having significant effects in some scenarios. Using these biomarkers in a clinical trial could reduce the required sample size of a clinical trial in CU participants by up to 70%. Finally, plasma NfL was shown to have worse theoretical performance as a trial progression marker compared to MRI-based measures, primarily due to its high within-subject variability. NfL compared better to cognitive measures as endpoints.

Discussion: The future of AD clinical trials will likely leverage plasma biomarkers for initial screening. Their utility for enrichment and tracking treatment response still needs to be evaluated in the context of other biomarkers measured in CSF, MRI, or PET. The plasma ATN biomarkers evaluated here all appear to be independently useful, but there is strong potential for more plasma biomarkers to be added to such a panel.

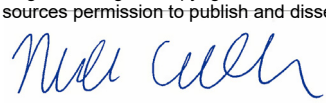
Key words

Alzheimer's disease, biomarkers, amyloid, tau, neurodegeneration

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language**
ISSN and key title 1652-8220		**ISBN** 978-91-8021-290-8
Recipient's notes	**Number of pages** 57	Price
	Security classification	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature  Date 2022-08-15

The future of clinical trials for Alzheimer's disease

A blood-based biomarker perspective

by Nicholas C. Cullen

LUND
UNIVERSITY

Cover photo by Erick Ingraham

Copyright pp 1-57 Nicholas C. Cullen (2022)

Paper 1 © The authors

Paper 2 © Springer Nature

Paper 3 © Springer Nature

Paper 4 © John Wiley & Sons

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2022:128

ISBN 978-91-8021-290-8
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2022

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

Table of Contents

Abbreviations

AD	Alzheimer's disease
ADNI	The Alzheimer's Disease Neuroimaging Initiative study
AIC	Akaike Information Criteria
APOE	Apolipoprotein E (typically, refers to the e4 allele)
AUC	Area under the curve of the receiver operating characteristic
Aβ	Amyloid-beta protein
Aβ42/Aβ40	The ratio of the 42-amino acid long beta-amyloid peptide to the 40-amino acid long beta-amyloid peptide
β	Beta coefficient from regression model
BioFINDER	The Biomarkers for Identifying Neurodegenerative Disorders Early and Reliably study
CDR	Clinical Dementia Rating – Global score
CDR-SB	Clinical Dementia Rating – Sum of Boxes
CN	Cognitively Normal
CSF	Cerebrospinal fluid
CU	Cognitively Unimpaired
CV	Coefficient of Variation
LME	Linear mixed effects modelling
MAE	Mean Absolute Error
MCI	Mild Cognitive Impairment
MMSE	Mini-Mental State Examination
MRI	Magnetic Resonance Imaging
NfL	Neurofilament light
OR	Odds ratio
PACC	Preclinical Alzheimer's Cognitive Composite
PET	Positron emission tomography
pTau181	Tau phosphorylated at 181
pTau217	Tau phosphorylated at 217
SCD	Subjective Cognitive Decline
SUVR	Standardized Uptake Value Ratio

List of Publications

I. "Plasma Aβ42/Aβ40 and *APOE* for amyloid PET pre-screening in secondary prevention trials of Alzheimer's disease." *In Revision.*

II. "Individualized prognosis of cognitive decline and dementia in mild cognitive impairment based on plasma biomarker combinations." *Nature Aging*

III. "Plasma biomarkers of Alzheimer's disease improve prediction of cognitive decline in cognitively unimpaired elderly populations." *Nature Communications*

IV. "Comparing progression biomarkers in clinical trials of early Alzheimer's disease." *Annals of Clinical and Translational Neurology*

Abstract

Objectives: The primary objective was to investigate the utility of blood-based biomarkers of amyloid, tau, and neurodegeneration for (i) screening, (ii) enrichment, and (iii) tracking response to treatment in clinical trials of Alzheimer's disease.

Methods: Longitudinal, participant-level data used in these studies was drawn from the Swedish BioFINDER study and the ADNI study. Participants were classified as cognitively unimpaired, mild cognitive impairment, or Alzheimer's disease dementia. For screening, logistic regression was used to predict amyloid PET status in CU individuals from plasma Aβ42/Aβ40, *APOE* status, and age. For enrichment, Linear mixed effects models were used to predict longitudinal cognitive decline and future risk of AD dementia in CU individuals or in MCI individuals from a basic model (age, sex, education, *APOE* status) and varying combinations of blood-based biomarkers (plasma Aβ42/Aβ40, plasma pTau181, plasma pTau217, plasma NfL). For treatment response, plasma NfL was measured longitudinally in MCI or AD patients and properties such as slope, inter-subject variability, and intra-subject variability were calculated. Plasma NfL was then compared with MRI and cognition.

Results: The amyloid PET screening model had an AUC of 0.87, with a significant independent effect for plasma Aβ42/Aβ40 and *APOE* status, but not age. This model was estimated to reduce total cost of recruiting 500 amyloid-positive CU participants by 31 – 42%, depending on the relative cost of amyloid scanning to plasma measurement. For enrichment, plasma pTau181 and pTau217 had the largest effect on predicting cognitive decline in CU and MCI participants, with Aβ42/Aβ40 and NfL having significant effects in some scenarios. Using these biomarkers in a clinical trial could reduce the required sample size of a clinical trial in CU participants by up to 70%. Finally, plasma NfL was shown to have worse theoretical performance as a trial progression marker compared to MRI-based measures, primarily due to its high within-subject variability. NfL compared better to cognitive measures as endpoints.

Discussion: The future of AD clinical trials will likely leverage plasma biomarkers for initial screening. Their utility for enrichment and tracking treatment response still needs to be evaluated in the context of other biomarkers measured in CSF, MRI, or PET. The plasma ATN biomarkers evaluated here all appear to be independently useful, but there is strong potential for more plasma biomarkers to be added to such a panel.

Background

The first generation of modern clinical trials of Alzheimer's disease (AD) relied on clinical diagnosis as the primary inclusion criteria (J. Cummings 2018). A clinical diagnosis is assigned by a neurologist who interacts with the patient, performs cognitive and functional assessment, and interviews caregivers or relatives to understand if there has been a significant decline in any relevant domains. Notably, a clinical diagnosis in the era of early AD clinical trials was not based on any objective measure of underlying biological processes related to AD – namely, abnormal amyloid and tau accumulation (Reitz 2012). This is not to assign blame, however; none of those biological processes could be measured back then.

Because clinical diagnosis of AD is notoriously difficult and often inaccurate – around 20% of diagnoses are incorrect in a specialized memory clinical setting and up to 50% may be incorrect in generalized care settings – a significant portion of participants who were included in the first generation of AD trials did not actually have AD (Anderson et al. 2017). This phenomenon likely explains in part (but not fully) why none of these early trials succeeded in halting or slowing the progression of the disease.

With time, the idea that AD was better defined by its underlying biological processes – namely, the accumulation of amyloid and subsequently tau proteins in the brain, followed by neurodegeneration – rather than clinical symptoms began to be appreciated. This led to a biological definition of AD in which abnormal amyloid accumulation was specifically noted as a signature requirement of an AD diagnosis (Dubois et al. 2007; McKhann et al. 2011).

Naturally, defining a disease by its biological characteristics necessitates a method to measure these processes through objective measures or tests. Such objective tests which measure underlying disease processing are conventionally called *biomarkers* (Hansson 2021). For AD, biomarkers of amyloid and tau (and other related processes such as neurodegeneration and neuroinflammation) were first developed and validated in CSF and with PET (Zetterberg and Blennow 2013; Jagust et al. 2015).

CSF and PET biomarkers in AD quickly gained widespread use in specialized care settings and have become the standard for inclusion into clinical trials of AD and for detecting a biological response to treatment (Sevigny et al. 2016; Anderson et al. 2017). As it stands today, it is exceedingly rare to see a late-stage clinical trial

which does not require confirmation of abnormal amyloid and/or tau status through CSF or PET. It is equally rare that such a clinical trial does not collect these measures during the trial as well.

However, while CSF and PET biomarkers have become firmly established as accurate methods for identifying underlying AD pathology and are thereby essential to clinical trials of AD, they are not without their drawbacks. These biomarker modalities are expensive to collect, invasive for the patient, and are largely inaccessible outside of specialized care settings (Zetterberg 2017).

Solving the obvious clash between the fact that CSF and PET biomarkers have become essential to run AD clinical trials and the fact that these biomarkers having a large economic burden therefore constitutes an important area of research. Making progress on this issue serves as the motivation and background for the present thesis work.

Thankfully, a promising solution for this dilemma has begun to take hold: blood-based biomarkers of amyloid, tau, and neurodegeneration. In the last few years, blood-based biomarkers have started to become more mature and widespread (Leuzy et al. 2022). The results are starting to pour in which show that these biomarkers have significant utility for AD clinical care and, by extension, clinical trials.

Introduction

Blood-based biomarkers have been proposed as a way to reduce the burden of CSF collection and PET scanning in the context of AD clinical trials (Hampel et al. 2018). Still, it is unlikely that CSF or PET biomarkers will be replaced by blood-based biomarkers in terms of what is used to determine whether an individual meets the biological inclusion criteria for a clinical trial. Assuming that CSF or PET biomarkers are likely to be required for inclusion into AD clinical trials for the near future, then, blood-based biomarkers may be used in three ways: for screening, for enrichment, or for tracking response to treatment.

Screening involves identifying individuals who are highly likely to meet the trial's inclusion criteria. In the case of AD clinical trials, the most important inclusion criteria which has been established in recent years is requirement of abnormal amyloid levels in the brain as evidenced by an amyloid PET scan (preferably) or by measurement of CSF Aβ42/Aβ40 levels (Sevigny et al. 2016).

The reason blood-based biomarkers are needed for screening is because using CSF or PET as required inclusion criteria leads to an expensive and drawn-out recruitment process. Only about half of individuals at the MCI stage of the disease are estimated to have abnormal levels of amyloid in the brain, and the prevalence for abnormal amyloid accumulation in the brain is even lower in for CU individuals, making it difficult to avoid a large number of negative (i.e., unnecessary) PET scans (Jansen et al. 2015).

Still, CSF or PET are likely to continue to be required for trial inclusion because they provide the most accurate way to measure AD pathology in the brain. The failure of AD clinical trials of the past two decades has shown how important it is to confirm that participants have AD pathology for inclusion into a trial (J. L. Cummings, Morstorf, and Zhong 2014).

Blood-based biomarkers, on the other hand, are both inexpensive and highly accessible. This makes them a perfect complement – rather than replacement – for CSF or PET biomarkers when used for inclusion into AD clinical trials. A major question, however, is how exactly blood-based biomarkers would be used for screening and which biomarker(s) are best suited for the task depending on the inclusion criteria (e.g., PET versus CSF; amyloid versus tau) and on the population (e.g., CU versus MCI).

One proposed workflow is to use blood-based biomarkers to identify individuals who are likely to be amyloid positive on a PET scan, for instance (Keshavan et al. 2021). The individuals who are at high risk for amyloid PET positivity would then be invited further to receive an amyloid PET scan. If the amyloid PET scan is positive, then the individual could be included in the trial. While screening in this way would result in added costs due to measuring blood-based biomarkers, if used effectively it would also result in lower costs due to a sharp reduction in the number of negative (i.e., wasted) amyloid PET scans which would be taken.

Understanding the predicted risk threshold for blood-based biomarker screening which would result in the largest cost saving is therefore a major open question. The optimal screening strategy (e.g., whether to have a low risk threshold and invite everyone for PET screening who is in the top 75% of risk levels, or whether to be strict and invite for PET screening only those who are in the top 25% of risk levels) can be determined by weighing the performance of blood-based biomarkers to accurately identify amyloid PET positive individuals with the cost ratio between blood-based biomarker measurement and amyloid PET measurement. It is exactly these questions which were addressed in Paper 1, which features an investigation of the combined use of plasma Aβ42/Aβ40, *APOE* status, and age to screen for amyloid PET positivity in prevention trials of AD.

While screening involves using inexpensive and/or accessible biomarkers to identify individuals who are likely to meet the inclusion criteria of AD clinical trials, it is no sure thing that individuals who meet the inclusion criteria will actually decline cognitively during the trial. The complex and progressive nature of AD-related cognitive decline is important to consider because a clinical trial full of participants who all remain completely stable during the trial will have zero chance of demonstrating a significant treatment effect (Veitch et al. 2018).

The issue of including non-progressors in AD clinical trials becomes even more important to consider as trials move to earlier stages of the disease where cognitive trajectories are more nuanced and drawn out (Brookmeyer et al. 2018). There are two main options to address this issue: increase the length of the trial so that more participants are likely to progress or select participants using criteria which goes even further beyond the standard inclusion criteria and is based on an individual's likelihood of showing cognitive decline during the trial.

Increasing the length of AD clinical trials continues to be explored, as trials targeting the earliest stages of the disease can now run up to four or six years (Insel et al. 2020). However, even this timeline does not ensure that enough participants will progress during the trial. Therefore, using biomarkers to identify individuals likely to show cognitive decline is an important area of research which can solve the issue of non-progressors. This task is typically called enrichment, and it has been an open question as to whether newly established blood-based biomarkers perform well here (Freidlin and Korn 2014; Holland, McEvoy, Desikan, et al. 2012).

It was clear quite early on in the development of blood-based biomarkers that they were not on the same level as the same biomarkers measured in CSF and PET when it comes to the strength of association with near-term (i.e., two to four years) cognitive decline (Betthauser et al. 2019). However, it has been mostly unexplored as to whether a combination of blood-based biomarkers could improve on the prediction of cognitive decline compared to individual biomarkers. Paper 2 and Paper 3 are aimed at answering exactly this question – does a combination of blood-based biomarkers outperform individual blood-based biomarkers for predicting cognitive decline? – along with trying to gain insight into what exactly that best combination would be. Paper 2 addressed these questions in patients with MCI, while Paper 3 featured elderly CU individuals.

Finally, measuring AD-related biomarkers in blood would also be beneficial when it comes to tracking response to treatment during the actual trial. Clinical trials in AD of today are focused on detecting not just a reduction in cognitive decline (the primary endpoint), but also a reduction in AD pathology and neurodegeneration (Howard and Liu 2020; Selkoe 2019). A reduction in AD-related pathology can be a strong signal in both early- and late-stage clinical trials that the treatment is effective. Deciding on the appropriate biomarkers which can measure the response to treatment across the different pathological axes of AD is an open question (Insel et al. 2015).

Tracking response to treatment involves identifying reduction in longitudinal biomarker trajectories in the treatment group with the hope of relating this reduction in biomarker to a reduction in clinical endpoint to establish proxy endpoints. Measuring biomarkers longitudinally, however, would be expensive and invasive if CSF or PET modalities were used. This claim is evidenced by the fact that recent AD trials typically collected CSF or PET biomarkers longitudinally only in a small subset of participants (Thambisetty et al. 2021).

Nonetheless, new data is starting to come out that suggests blood-based biomarkers of AD indeed show reductions in response to treatment (Swanson et al. 2021). Identifying the blood-based biomarkers which are most appropriate to measure during the trial, however, requires an analysis of the longitudinal properties of the biomarker in an observational context. Properties of interest which can affect the utility of a biomarker as a longitudinal or proxy endpoint include (1) the change over time in the disease group compared to normal aging, (2) the variability of the biomarker change over time across different individuals, and (3) the variability in the biomarker levels at each time point around the overall trajectory within the same individual. These properties can respectively be called the slope, the inter-subject variability, and the intra-subject variability; these properties are all important to consider when deciding which biomarkers to collect during a trial (Huang et al. 2017; M. C. Donohue and Aisen 2012).

An important question to be answered in the blood-based biomarker era is how these properties measured in blood-based biomarkers compare to biomarkers measured in other modalities. This question was addressed by Paper 4, where the focus was specifically on biomarkers of neurodegeneration because the main blood-based biomarker of neurodegeneration (plasma NfL) had been developed furthest at that point in time.

In all, the blood-based biomarker revolution brought about the development of multiple biomarkers for AD-related processes such as amyloid and tau accumulation, along with general neurodegeneration. The blood-based biomarkers of focus in the work presented here are plasma Aβ42/Aβ40 to reflect amyloid accumulation, plasma p-tau181 and plasma p-tau217 to reflect tau accumulation, and NfL to reflect general neurodegeneration (West et al. 2021; Janelidze et al. 2020; Thijssen et al. 2020; Mattsson et al. 2017). It is important to note that these biomarkers are thought reflect the associated underlying AD pathology to varying degrees of closeness, although this is not a focus of the current work. Validating blood-based biomarkers for screening, enrichment, and treatment response involved building clinical prediction models of AD-related outcomes and then developing and testing strategies by which these blood-based biomarker models can make clinical trials more efficient. It required data from longitudinal, observational cohorts with harmonized, multi-modal biomarker collection.

Aims

The primary aim for each study was as follows:

1. To assess the utility of plasma Aβ42/Aβ40, in combination with *APOE* and age, to screen for abnormal amyloid PET status in CU individuals
2. To investigate the performance of a combination of blood-based biomarkers for predicting cognitive decline in CU individuals
3. To investigate the performance of a combination of blood-based biomarkers for predicting cognitive decline in MCI individuals
4. To measure properties of longitudinal trajectories for plasma NfL and to compare these properties to other neurodegeneration biomarkers

There were also more general aims of these studies as a whole. These aims can be defined as follows:

1. To understand the degree to which blood-based biomarkers provide independent information from each other
2. To compare blood-based biomarkers to more basic models consisting of only demographics, baseline cognition, and *APOE*
3. To compare blood-based biomarkers to more invasive models consisting of similar biomarkers measured in CSF
4. To determine whether blood-based biomarkers would be effective for screening, enrichment, and treatment response in AD trials

Methods

Participants

The studies presented in this thesis rely on participants from two longitudinal, observational cohorts: the ADNI study (NCT00106899) and the Swedish BioFINDER-1 study (NCT01208675). Both cohorts share many similarities, particularly in the fact that they both recruit and follow participants across the AD spectrum – from subjective cognitive decline (SCD), mild cognitive impairment (MCI), and AD dementia – along with cognitively unimpaired (CU), healthy elderly individuals. The ADNI study was launched in 2003 as a public-private partnership (Petersen et al. 2010). The Swedish BioFINDER-1 study began in 2005 (Palmqvist et al. 2015). Both studies are still ongoing and were approved by local institutional review boards. Written informed consent was received from all participants.

In Paper 1, CU participants from the Swedish BioFINDER-1 study were included. The CU participants consisted of (i) individuals who were cognitively normal (CN) with no objective evidence of cognitive impairment at baseline and (ii) individuals with subjective cognitive decline (SCD) who were referred to the memory clinic for investigation but deemed not to have any cognitive impairment after undergoing an extensive neuropsychological battery. The inclusion criteria for these participants included being at least 60 years of age, under 80 years of age if SCD, have no objective cognitive impairment, have an MMSE score of at least 28 for CN participants and at least 24 for SCD participants at the screening visit, be fluent in Swedish, and not fulfil the criteria for MCI or dementia according to the DSM-5. Exclusion criteria included having an unstable illness that would make it difficult to participate in the study or have current alcohol or substance abuse.

In Paper 2, MCI participants from the Swedish BioFINDER-1 study were included in the analysis. These MCI participants were recruited and evaluated in a memory clinic setting after referral from primary care. Inclusion criteria involved being between 60 and 80 years old, fulfilling consensus criteria for MCI, an MMSE score of at least 24, and no or minimal impact of daily living activities, while not fulfilling criteria for dementia. Exclusion criteria was cognitive impairment that could better be attributed to another non-neurodegenerative condition, severe somatic disease, and alcohol or substance abuse. The analysis in Paper 2 was also validated using MCI participants from the ADNI study.

In Paper 3, the same criteria as defined for Paper 1 was used to identify participants from the Swedish BioFINDER-1 study.

In Paper 4, participants from the ADNI study were included in analysis. Inclusion criteria for this analysis involved being between 55 and 90 years of age, completing at least six years of education, fluent in Spanish or English, and no significant neurological disease. All participants included in this analysis had a CDR score between 0 – 1 and had amyloid PET or CSF Aβ42 measurement available.

Cognitive & Clinical Assessment

Paper 1 did not feature any cognitive assessment in the analysis beyond what was used for determining participant inclusion and exclusion as defined above.

The main outcomes in Paper 2 involved the Mini-Mental State Examination (MMSE), which is a global cognitive measure scored from 0 – 30, and clinical conversion to AD dementia (Chapman et al. 2016). Both of these outcomes were evaluated four years and two years after baseline. Clinical status of dementia due to AD was evaluated according to the DSM-5 criteria in the Swedish BioFINDER-1 study and the ADNI study used the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association criteria for probable AD (Dubois et al. 2007).

Paper 3 featured the Pre-Alzheimer's Cognitive Composite (PACC) as the primary cognitive outcome because this cognitive scale was developed to be more sensitive to the earliest cognitive changes in CU individuals (Michael C Donohue et al. 2014). The PACC score used in this analysis consisted of a weighted sum of the z-score values of four cognitive assessments: the MMSE, delayed word recall from the Alzheimer's Disease Assessment Scale – Cognitive Subscale (weighted double), animal fluency, and trail-making B tests. The secondary cognitive assessment was the MMSE score on its own. Conversion to AD dementia or all-cause dementia was similarly assessed as described above for Paper 2 and used as a secondary outcome.

The analysis in Paper 4 involved two cognitive scales evaluated longitudinally. The scales included the Clinical Dementia Rating – Sum of Boxes (CDRSB) and the modified PACC score. The CDRSB score reflects clinically relevant symptoms throughout AD progression and the PACC score attempts to detect these symptoms earlier on (Aisen 2015). The PACC score used here was composed of the MMSE, logical memory delayed recall, trail-making test B, and the delayed word recall from the ADAS-Cog scale. Since Paper 4 was focused on cognitive endpoints in clinical trials, it was important to include two of the most commonly used cognitive endpoints in actual clinical trials.

Imaging Biomarkers

In Paper 1, amyloid PET scans were collected at baseline and used at the primary outcome for analysis. The amyloid PET scan used was 18F-flutemetamol PET conducted on a Philips Gemini TF 16 scanner. A global neocortical compositive standardized uptake ratio (SUVR) was calculated using cerebellar cortex as reference region, with an abnormal amyloid PET scan was defined as SUVR > 0.742 as defined previously (Palmqvist et al. 2014; Janelidze et al. 2017).

No imaging biomarkers were assessed as outcomes in Paper 2 or Paper 3. In Paper 4, structural MRI scans were collected longitudinally and used as an outcome in the analysis. The scans were acquired using a 3T scanner with a standardized protocol across sites. Regional volume and cortical thickness measurements were derived using the 2010 Desikan-Killany atlas and the longitudinal pipeline from FreeSurfer v5.1 software (Fischl 2011). From these regional structural brain estimates, a temporal composite was derived from consisted of the area-normalized bilateral cortical thickness in entorhinal, fusiform, inferior temporal, and middle temporal regions (Jack et al. 2015). The bi-laterally averaged hippocampal volume was also used a structural MRI outcome.

Amyloid PET scans were also used in Paper 4 to defined amyloid-positive individuals using 18-F Florbetapir PET scans, with abnormality being defined from SUVR values derived from a cortical ROI (Carotenuto et al. 2020).

Fluid Biomarkers

In Paper 1, plasma Aβ42/Aβ40 was measured at baseline using an IP-MS method described previously (Janelidze et al. 2021). The average intra-assay coefficient of variation (CV) was 0.72% and the average inter-assay CV was 3.46%. Analysis was done at the Bateman laboratory at Washington University in St. Louis. Additionally, *APOE* genotype was measured at baseline and treated as a binary variable indicated the presence of at least one e4 allele or not.

In Paper 2, a wider range of plasma biomarkers were measured. For the Swedish BioFINDER-1 cohort, this included plasma Aβ42/Aβ40 as measured using an Elecsys immunoassay on a Cobas e601 analyzer, as well as using a mass spectrometry-based plasma Aβ42/Aβ40 assay from Araclon Biotech (Palmqvist et al. 2019). Plasma pTau181 was measured on a Meso Scale Discovery platform using an assay developed by Eli Lilly (Janelidze et al. 2020). Plasma NfL was analyzed using a Simoa-based assay (Mattsson et al. 2019). CSF biomarkers were also measured in this cohort. CSF Aβ42 was and pTau181 were measured using Elecsys

assays from Roche Diagnostics and CSF NfL was measured using an ELISA method from Uman Diagnostics.

Additionally, plasma biomarkers were measured in Paper 2 for participants from the ADNI study. Plasma Aβ42/Aβ40 was measured using an IP-MS method, and pTau181 was measured using a Simoa HD-X Analyzer from Quanterix using an assay developed in the Clinical Neurochemistry Laboratory at the University of Gothenburg, Sweden (Karikari et al. 2020). Plasma NfL in ADNI study participants was measured using the same Simoa-based assay as described above. All biomarkers used in Paper 2 were natural log transformed and any binary cutoffs were derived using Youden's index to maximize separation between amyloid-negative CU participants and amyloid-positive AD dementia patients.

In Paper 3, the biomarkers were measured using the same method as was done for participants in Paper 2. The main difference was that plasma pTau217 was also measured and available for the analysis in Paper 3. Plasma pTau217 levels were measured on a Meso-Scale Discovery platform using an assay developed by Eli Lilly (Palmqvist et al. 2020).

In Paper 4, plasma NfL was the only fluid biomarker included in the analysis. Here, longitudinal plasma NfL levels were measured at the Clinical Neurochemistry Laboratory at University of Gothenburg, Sweden using an in-hour ultrasensitive ELISA on the Simoa platform from Quanterix (Mattsson et al. 2019). The assay had limits of quantification of 6.7 ng/L and 1620 ng/L, with an intra-assay CV of 6.2% and an inter-assay CV of 9.0%.

Statistical Analysis

A variety of statistical methods were used across the studies presented here, although the methods were generally consistent based on whether the outcome was cross-sectional versus longitudinal and continuous versus binary.

For cross-sectional, continuous outcomes, linear regression was used. Linear regression models were fit with combinations of biomarkers as independent variables. Performance metrics of interest for linear regression include R^2 and AIC. For longitudinal, continuous outcomes, linear mixed effects modelling was used (M. C. Donohue and Aisen 2012). Linear mixed effects models were employed in Paper 2 and Paper 3 to analyze the association between baseline levels of plasma biomarkers and longitudinal change in cognition. The models featured random intercepts and random slopes, and baseline levels of the outcome were always included as a covariate. Performance metrics of interest here included R^2 and AIC. Similar linear mixed effects models were used in Paper 4 to model longitudinal biomarker trajectories. Multiple metrics of model fit relevant to statistical power

analysis were extracted and compared across biomarkers. These metrics included the residual error, inter-subject variability, and intra-subject variability.

For cross-sectional, binary outcomes, logistic regression was used. Such a logistic regression model was used in Paper 1 to predict abnormal amyloid PET status from plasma Aβ42/Aβ40, *APOE*, and age. The metrics used to evaluate this model were AUC and AIC. For longitudinal, binary outcomes, Cox regression was used. Cox regression requires an outcome variable representing whether or not each individual converting to AD dementia (or any type of dementia) while in the study, along with another variable specifying how long after baseline the individual converted to AD dementia (or any type of dementia) or how long the individual has been in the study (if they have not converted). Alternatively, logistic regression can be used instead of Cox regression if the longitudinal, binary outcome is viewed at a specific point in time – here, two years or four years after baseline. In this case, only participants who have developed the positive binary outcome value or who have been followed in the study up to the given time cutoff while not developing the positive binary outcome value can be included.

All statistical analysis in these papers was performed using the R programming language. All statistical tests were two-sided with an alpha of 0.05. Correction for multiple comparisons was not relevant for any analyses.

Results

Participant characteristics

The ADNI and BioFINDER studies made up the core of the participants used in the studies presented here. Inclusion into the studies was based primarily on meeting diagnostic and biomarker criteria, completing enough follow-up visits (usually, 2 or 4 years), and availability of relevant biomarker data.

In Paper 1, the group consisted of 180 CU participants from the BioFINDER study, of which 80 individuals (44.4%) were SCD participants. A total of 52 participants (28.9%) had abnormal amyloid PET scans at baseline. The percentage of amyloid PET positive participants in the SCD group (32 of 80; 40%) was significantly greater than in the CN group (20 of 100; 20%; P = 0.005). The average age of participants in this study was 73.0 ± 5.3 years, 111 participants (61.7%) were female, and the average years of education was 11.9 ± 3.3 years.

In Paper 2, the participants consisted of 148 MCI individuals from the BioFINDER study who had available plasma and CSF biomarker measurements. The mean age in this group was 71.4 years, 36.5% were female, and the mean education was 11.2 years. The mean MMSE score in this group was 27.2 (1.7) at baseline and decreased to 21.8 (5.2) on average four years after baseline. Moreover, 59.8% of participants in this group developed AD dementia within four years after baseline.

The analysis in Paper 2 also included 86 MCI individuals from the ADNI study for whom the full set of plasma biomarker measurements (Aβ42/Aβ40, P-tau181, and NfL) were available. In this cohort, the mean age was 71.5 years, 51.2% were female, and the mean education was 16.4 years. The mean MMSE score was 28.3 (1.7) at baseline and dropped only slightly to 27.6 (2.9) at four years after baseline. Similarly, only 10.8% of participants in this group developed AD dementia within four years after baseline.

There was also a separate group of 425 MCI participants from the ADNI cohort which was used for validating fitted statistical models. this group had only plasma P-tau181 and plasma NfL measurements available. This group had similar baseline characteristics as the smaller group of ADNI participants but showed more decline: the mean MMSE score dropped from 28.2 (1.7) at baseline to 26.6 (4.1) four years after baseline, and 33.1% of participants developed AD dementia within four years of baseline.

In Paper 3, the participants consisted of 435 CU individuals, of which 167 (38.4%) had SCD while all other patients were CN. The average age in the overall CU group was 72.58 (5.45), with 44.6% being female and having 12.17 (3.66) years of education on average. The average MMSE score at baseline was 28.8 (1.2) and the average PACC score at baseline was 0.01 (0.74). These participants were followed for an average of 4.75 (1.66) years, during which the average four-year change in MMSE was -1.02 (2.93) points and the average four-year change in PACC was -0.33 (0.85) points. Moreover, 28 participants (6.4%) developed AD dementia and 39 participants (9.0%) developed any form of dementia.

The primary difference in how SCD versus CN participants were recruited is that SCD participants were referred to the memory clinic but did not meet the criteria for MCI after undergoing cognitive testing, while CN participants were recruited from the community and not followed in a memory clinic setting. Therefore, looking at differences in demographics, biomarker levels, and outcomes in the CN versus SCD groups was highly relevant for participants involved in Paper 3.

For demographics, it was found that the SCD group was significantly younger (73.62 in CN versus 70.92 in SCD; $P < 0.0001$) and had a significantly higher proportion of *APOE* e4 carriers (27% in CN versus 45% in SCD; $P = 0.0001$). There was no significant difference in education or gender representation between the two groups.

In terms of cognitive and clinical outcomes, the SCD group had lower MMSE scores at baseline (-29.03 versus 28.44; $P < 0.0001$) and lower PACC scores at baseline (0.17 versus -0.23; $P < 0.0001$), although there was no significant difference in four-year change in MMSE ($P = 0.414$) or four-year change in PACC ($P = 0.102$) between the two groups. Moreover, the SCD group was significantly more likely to develop AD dementia (1.1% in CN versus 15.0% in SCD; $P < 0.0001$) or any type of dementia ($P < 0.0001$).

For biomarker levels, there were no significant differences between CN and SCD participants in terms of plasma Aβ42/40 or CSF Aβ42/40, nor in terms of plasma pTau217 or CSF pTau181. However, the SCD group had significantly lower plasma NfL levels (7.63 versus 7.54; $P = 0.036$) yet significantly higher CSF NfL levels (6.75 versus 6.86; $P = 0.027$).

In Paper 4, the participants consisted of individuals from the ADNI study who were classified into three separate groups based on baseline amyloid status as defined by amyloid PET (or CSF Aβ42 if amyloid PET was not available) as well as baseline cognitive status as defined by the CDR Global cognitive score. This study was focused on analyzing longitudinal biomarkers of plasma, MRI, and cognition, but there was no requirement to have completed a given number of visits.

The first group ("controls") consisted of 330 individuals defined by a negative amyloid status (i.e., normal levels of amyloid) and a CDR score of 0. In this group, the average age was 72.7 (6.1), 50% were female, and the education level was 16.2

(2.5) years. The average MMSE at baseline was 29.2 (1.1) and the average PACC score at baseline was 0.05 (2.6). This group had a total of 631 MRI observations, 750 plasma NfL observations, and 1,794 cognition observations.

The second group ("preclinical AD") consisted of 218 individuals defined by a positive amyloid status and a CDR score of 0. In this group, the average age was 73.3 (6.1), 61.9% were female, and the education level was 16.4 (2.5) years. The average MMSE at baseline was 29.0 (1.2) and the average PACC score at baseline was -0.35 (2.8). This group had a total of 343 MRI observations, 430 plasma NfL observations, and 1,117 cognition observations.

The third and final group ("mild AD") consisted of 697 individuals defined by a positive amyloid status and a CDR score of 0.5 or 1. A total of 388 of the 697 participants in this group also had abnormal tau accumulation as define by CSF pTau levels greater than 27 pg/mL. In this group, the average age was 73.4 (7.2), 43.5% were female, and the education level was 15.7 (2.8) years. The average MMSE at baseline was 26.0 (2.7) and the average PACC score at baseline was -9.8 (6.2). This group had a total of 1,510 MRI observations, 1,457 plasma NfL observations, and 3,559 cognition observations.

In terms of differences between groups, the demographics were quite similar between all groups, although the mild AD group had slightly less educational attainment than the other groups. The preclinical AD group had similar baseline MMSE levels as the control group but lower baseline PACC levels.

Predicting amyloid PET status

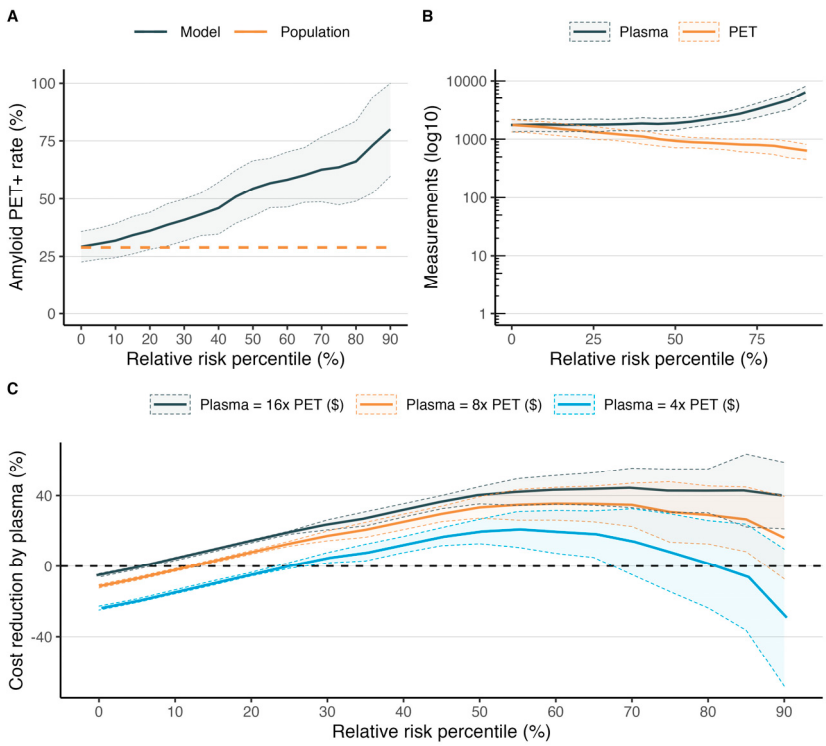
The first step for investigating how biomarkers perform for screening is to fit a statistical model to predict the relevant screening outcome. Here, a logistic regression model was fit to predict amyloid PET status (normal versus abnormal) from plasma Aβ42/Aβ40, *APOE* status, and age. This model had an AUC of 0.87 (95% CI [0.82, 0.92]) and there was a significant independent effect for plasma Aβ42/Aβ40 (OR = 5.98 [3.27, 12.32], P < 0.0001) and *APOE* status (OR = 1.85 [1.26, 2.76], P = 0.002), but not age (P = 0.62). A model which excluded age was tested and found to have a similar performance for predicting amyloid PET status as the full model.

As there was an overrepresentation of amyloid PET positivity in the SCD group compared to the CN group, a sensitivity analysis was performed in which the logistic regression analysis was restricted to CN participants. The results here showed only a small decrease in model performance (AUC = 0.84 [0.75, 0.92]). A significant effect remained for plasma Aβ42/Aβ40 (OR = 4.21 [1.92, 11.02], P = 0.001), but the effect became only a trend for *APOE* status (OR = 1.55 [0.92, 2.64], P = 0.098) and remained non-significant for age.

Trial screening

The results from testing this model in a screening scenario with different risk percentile cutoffs demonstrated that the expected amyloid PET+ rate would increase from 28.9% with no screening (i.e., the baseline rate in the entire population) to 38.7% (CI [29.8, 47.7]; P < 0.0001 versus no pre-screening) with a 25th percentile risk score cutoff (i.e., the 75% of screened individuals with highest predicted amyloid PET risk scores would be invited for PET scanning). With a 50th percentile risk score cutoff, the amyloid PET+ rate would increase to 54.6% (CI [43.0, 66.2], P < 0.0001). With a 75th percentile cutoff, the amyloid PET+ rate would increase even further to 63.7% (CI [48.8, 78.6]; P < 0.0001).

In terms of the number of plasma measurements versus PET scans which would be required to recruit 500 amyloid PET+ CU individuals, the baseline amyloid PET+ rate of 28.9% in the overall population means that 1,749 amyloid PET scans would be needed to fulfill recruitment. However, implementing pre-screening with the fitted model above (i.e., only taking amyloid PET scans in the pre-screened population) would reduce the expected number of amyloid PET scans to 1,312 at a 25th percentile risk score cutoff. At a 50th percentile risk score cutoff, 926 amyloid PET scans would be needed. And at a 75th percentile risk score cutoff, only 796 amyloid PET scans would be needed. These results are visualized in the figure below.



While 1,749 amyloid PET scans would be expected to recruit 500 amyloid PET+ individuals without pre-screening, this strategy requires no plasma measurement. In contrast, implementing pre-screening with the fitted model above would require 1,750 individuals to receive plasma measurement to screen in enough amyloid PET+ individuals with a 25th percentile risk score cutoff. With a 50th percentile risk score cutoff, 1,850 plasma measurements would be required. And with a 75th percentile risk score cutoff, 3,184 plasma measurements would be required.

As these results show, there is a tradeoff between the number of plasma measurements and the number of amyloid PET scans which would be expected under various pre-screening risk score cutoffs. A cost-benefit analysis performed under the assumption that amyloid PET scanning was only four-times (4x) as expensive as plasma biomarker measurement showed that there would be no significant cost savings by employing the fitted pre-screening model with a 25th percentile risk cutoff (Δcost = -8.7% [-12.1, 29.5], P = 0.18), or a 75th percentile risk cutoff (Δcost = -0.27% CI [-1.3, +5.6], P = 0.05). However, with a 4x PET:plasma cost ratio there was indeed a significant cost saving with a 50th percentile risk cutoff (Δcost = -20.8% CI [-26.9, -14.7], P < 0.001).

When assuming an 8x PET:plasma cost ratio, there was always a significant cost savings to be had by plasma pre-screening. The cost savings by pre-screening with a 25th percentile risk score cutoff compared to no pre-screening was 31.3% ([16.2, 46.4], P = 0.002). With a 50th percentile risk score cutoff at an 8x PET:plasma cost ratio, the cost savings was 33.9% ([28.9, 38.9], P < 0.001), and the cost savings with a 75th percentile risk score cutoff was only 12.8% ([11.2, 14.3], P < 0.0001).

Lastly for the cost-benefit analysis, assuming an 16x PET:plasma cost ratio showed that there was again always a significant cost savings to be had by plasma pre-screening regardless of the risk score cutoff to be invited further for an amyloid PET scan. The cost savings by pre-screening compared to no pre-screening increased to 42.7% with a 25th percentile risk score cutoff, increased to 30.6% with a 50th percentile risk score cutoff, and increased to 18.9% with a 75th percentile risk score cutoff.

Moreover, the cost-benefit analysis was performed in a logistic regression model which excluded age. This sensitivity analysis was performed because age was not a significant predictor in the original logistic regression model that included plasma Aβ42/Aβ40, *APOE* status, age. Here, the results showed almost no difference compared to the results found when age was included in the model. Thus, age had little impact on the amyloid PET risk scores.

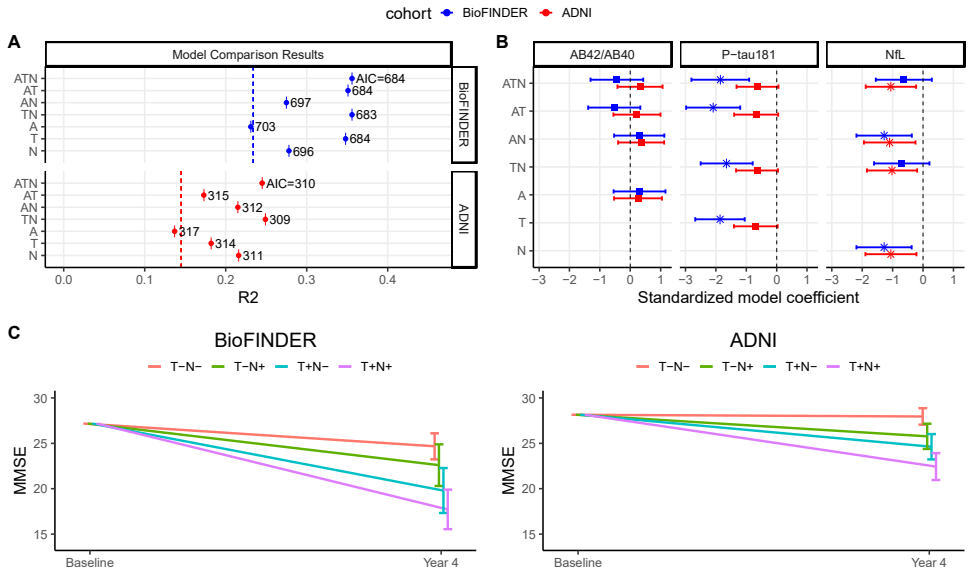
Predicting AD-related decline

Investigating how plasma biomarkers perform for enrichment involves first fitting a statistical model to predict longitudinal outcomes such as cognitive decline or risk for AD dementia from baseline biomarker levels. In the case that the best combination of biomarkers to use at baseline for predicting longitudinal outcomes is not known, there can also be a model selection step in which different combinations of biomarkers are tested and the most parsimonious (i.e., the best performing combination set of variables with the fewest number of variables) set of variables is identified.

In Paper 2, in which plasma Aβ42/Aβ40, pTau181, and NfL was analyzed in MCI participants from the BioFINDER and ADNI cohorts, a model selection procedure was performed to identify the best performing combination of plasma biomarkers to predict longitudinal cognition (four-year MMSE) and clinical conversion (four-year risk of AD dementia). All possible combinations of the three plasma biomarkers were tested, but these combinations were always tested when added to a "basic" model consisting of age, sex, education, and baseline MMSE score.

With four-year MMSE as outcome in the BioFINDER sample (n = 118), the model which included all three plasma biomarkers (R2 = 0.36, AIC = 684) fit the data significantly better than the basic model which included age, sex, education, and baseline MMSE score (R2 = 0.24, AIC = 702, P = 0.0001). The most parsimonious model, however, which achieved the lowest AIC score included only plasma P-tau181 and plasma NfL (R2 = 0.36, AIC = 683). In this most parsimonious model, there was a significant effect of plasma pTau181 (B = -1.65 points decline in MMSE per standard deviation increase in biomarker value; P < 0.0001) but not for plasma NfL (B = -0.70; P = 0.13).

This finding was successfully replicated in the ADNI model selection sample (n = 64) with four-year MMSE as outcome. Here, the performance of the full plasma biomarker model containing all three plasma biomarkers (R2 = 0.25, AIC = 310) was significantly better than that of the basic model consisting only of age, sex, education, and baseline MMSE (R2 = 0.15; P = 0.01). Again, the model which included only plasma P-tau181 and plasma NfL was the most parsimonious in terms of AIC value (R2 = 0.25, AIC = 309). However, in this model there was a significant individual effect for plasma NfL (B = -1.02, P = 0.02) but only a near-significant effect for plasma pTau181 (B = -0.64, P = 0.06). These results are visualized in the figure below.



When performing the same model selection procedure in the BioFINDER and ADNI cohorts with four-year conversion to AD dementia as the outcome, the results were again replicated showing that 1) the full plasma biomarker outperformed the basic model and 2) the model consisting of only plasma pTau181 and plasma NfL (in combination with the basic model) was the most parsimonious model in terms of AIC.

More specifically, in the BioFINDER cohort, the performance of the full plasma biomarker model for predicting four-year conversion to AD dementia (AUC = 0.88, AIC = 106) was significantly better than the basic model (AUC = 0.70, AIC = 140; $P < 0.0001$). The most parsimonious model included only plasma P-tau181 and plasma NfL (AUC = 0.88, AIC = 104). Again, this model had a significant effect of plasma pTau181 (OR = 5.87, P = 0.0001) but only a trend for plasma NfL (OR = 1.73, P = 0.10).

In the ADNI cohort, the performance of the full plasma biomarker model for predicting four-year conversion to AD dementia (AUC = 0.88, AIC = 50) was again significantly better than the basic model alone (AUC = 0.74, AIC = 57, P = 0.005). The most parsimonious model again included only plasma pTau181 and plasma NfL (AUC = 0.89, AIC = 49). In this model, there was a significant effect of plasma pTau181 (or = 4.58, P = 0.009) but not plasma NfL (OR = 2.15, P = 0.20).

A sensitivity analysis was performed in which apoE4 status was included as part of the basic model in addition to age, sex, education, and baseline MMSE, but this did not affect any of the findings – i.e., the full plasma biomarker model still

outperformed the basic model even with apoE4 status included, and the combination of plasma pTau181 and plasma NfL was generally the most parsimonious when added to the basic model.

Due to the low independent prognostic value provided by plasma Aβ42/Aβ40 in the tested models, another sensitivity analysis was performed in which a more sensitive, mass spectrometry-based assay was used to measure Aβ42/Aβ40 in plasma. However, while the results were qualitatively stronger for this assay, it still had no significant effect on the model selection procedure – i.e., plasma Aβ42/Aβ40 was never selected as part of the most parsimonious model.

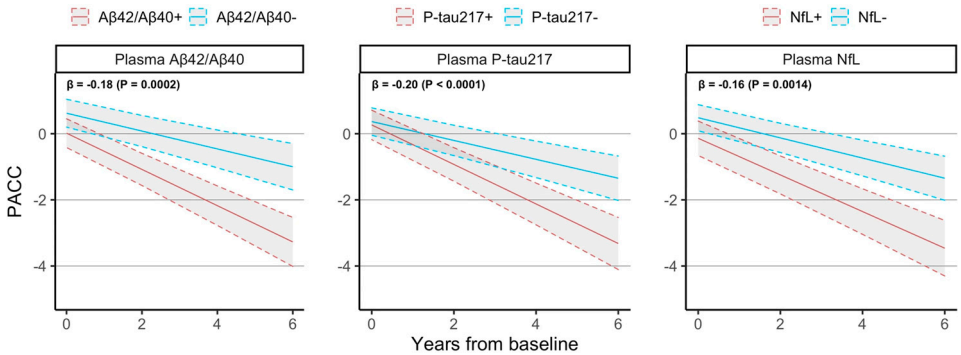
Because the analysis in Paper 2 involved data from two independent cohorts (BioFINDER and ADNI), it was possible to perform an out-of-sample validation where the statistical models described above to predict cognitive decline and conversion to AD could be fit in one of the cohorts and then evaluated in the other. However, because the plasma assays were not harmonized across studies, this analysis was performed after dichotomizing plasma biomarker levels for each participant into "normal" or "abnormal" based on pre-defined cutoffs. For this validation analysis, only the most parsimonious biomarker model (i.e., plasma pTau181 and plasma NfL) was tested along with the basic model consisting of age, sex, education, and baseline MMSE.

When externally validating the plasma biomarker models, it was found that the plasma biomarker model significantly improved participant-level, out-of-sample prediction of four-year MMSE score compared to the basic model, both when the model was fit on BioFINDER participants and tested on ADNI participants (MAE = 3.74 versus 4.08, P = 0.0006; 8.3% improvement) and when the model was fit on ADNI and tested on BioFINDER (MAE = 4.15 versus 5.19, P < 0.0001; 20.1% improvement.

The same result was found when four-year conversion to AD was used as the outcome. When fitting on BioFINDER and testing on ADNI, the out-of-sample AUC was 0.62 with the basic model while the out-of-sample AUC was 0.73 with the plasma biomarker model (P < 0.0001; 18.3% improvement). Similarly, when fitting on ADNI and testing on BioFINDER, the out-of-sample AUC was 0.61 with the basic model and the out-of-sample AUC was 0.79 with the plasma biomarker model (P < 0.0001; 29.3% improvement).

A similar procedure to that of Paper 2 for evaluating plasma biomarkers in terms of predicting AD-related decline was also performed in Paper 3. The primary difference was that Paper 3 featured CU individuals from the BioFINDER cohort rather than MCI individuals. Moreover, the analysis in Paper 3 included plasma pTau217 instead of plasma pTau181 (reflecting advancements in plasma biomarker assay technology), and the analysis in Paper 3 also included PACC instead of MMSE as the primary outcome (reflecting the need to have a cognitive outcome which reflects more subtle changes in cognition when studying CU individuals).

The first analysis tested whether plasma biomarkers were significantly associated with four-year change in PACC when added individually to a basic model consisting of age, sex, and education. The main result here showed that all three plasma biomarkers were significant associated with cognitive decline. Plasma pTau217 had the strongest standardized association with four-year PACC (B = -0.2 points/year per standard deviation increase in biomarker value, P < 0.0001), followed by plasma Aβ42/Aβ40 (B = -0.18, P = 0.0002), and then plasma NfL (B = -0.16, P = 0.001). These results are visualized in the figure below.



When combining all three biomarkers together with the basic model, all three biomarkers retained their significant association (P = 0.004 for Aβ42/Aβ40, P = 0.002 for pTau217, P = 0.01 for NfL). Interestingly, a sensitivity analysis of adding *APOE* e4 status to this model did not remove the significance of any individual plasma biomarkers, but it did decrease the effect size of plasma Aβ42/Aβ40 the most (B = -0.16, P = 0.002 with *APOE* e4 status included).

Looking at model performance for predicting four-year change in PACC, the basic model had an R2 of 0.07 (CI [0.06, 0.11]) and an AIC of 6699. These values were the baseline by which plasma biomarkers models were compared. And of the plasma biomarker models, the model which included all three plasma biomarkers had the best performance (R2 = 0.14 [0.12, 0.17], dAIC = -28 versus basic model), followed by the model which included only plasma Aβ42/Aβ40 (R2 = 0.11, dAIC = -14 versus basic model, P < 0.0001), the model which included only plasma pTau217 (R2 = 0.10, dAIC = -9 versus basic model, P = 0.0002), and finally the model which included only plasma NfL (R2 = 0.10, dAIC = -9 versus basic model, P = 0.002).

Two additional analyses were performed when investigating the association between plasma biomarkers and four-year change in PACC in this study. First, the analysis was done again using data only from SCD participants. The results here showed a significantly increased basic model fit (R2 = 0.17 for SCD-only participants compared to R2 = 0.07 for all CU participants), but no qualitative difference in the model

performance or association of plasma biomarkers with PACC. Secondly, the same analysis was performed using corresponding CSF biomarkers (Aβ42/Aβ40, pTau181, NfL) instead of plasma biomarkers. The results here showed that a model which combined all CSF biomarkers was significantly better at predicting four-year change in PACC compared to the model which combined all plasma biomarkers (dAIC = -98 for CSF model compared to basic model versus dAIC = -28 for plasma compared to basic model; dAIC = -70 for CSF compared to plasma models).

A similar analysis as above was performed with four-year conversion to AD dementia as the outcome. In general, the results were similar: the performance of the basic model for predicting four-year conversion to AD dementia (AUC = 0.64 [0.55, 0.77], AIC = 274) was significantly improved by adding all three plasma biomarkers (AUC = 0.82 [0.77, 0.91], dAIC = -25 versus basic model, $P < 0.0001$).

There were only a few differences in the results when using four-year conversion to AD dementia as outcome. For one, plasma pTau217 had the strongest association with four-year conversion to AD dementia – as opposed to plasma Aβ42/Aβ40 for four-year change in PACC – and plasma NfL on its own (with the basic model) did not have a significant association with four-year conversion to AD dementia (OR = 1.51, $P = 0.07$). However, when using four-year conversion to any form of dementia as the outcome, the association between plasma NfL and the outcome became significant. Importantly, adjusting for CN/SCD status in the models did not have an effect on any associations between plasma biomarkers and the outcomes.

Trial enrichment

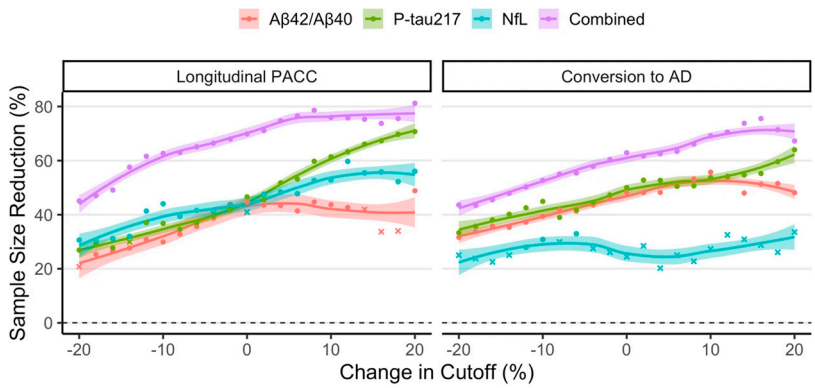
After having fit statistical models to predicting four-year PACC change in CU participants using plasma biomarkers, a natural extension to this analysis was to investigate the utility of these plasma biomarker-based models for clinical trial enrichment. This analysis involved determining whether using predicted cognitive decline from the plasma biomarker models to include participants would reduce the required sample size of a theoretical clinical trial of CU individuals aimed at slowing change in PACC by 30% over four years compared to an identical clinical trial which did not use plasma biomarker-based enrichment.

The hypothesis was that using plasma biomarkers for enrichment would result in a clinical trial with more participants who progressed during the trial (rather than remaining cognitively stable), which in turn would increase the likelihood of detecting a treatment effect with fewer participants.

Enrichment using individual plasma biomarkers was investigated first. The results here showed that enriching for plasma pTau217 – by including only individuals with abnormal plasma pTau217 levels as determined by a pre-defined cutoff – would

reduce the required sample size by 47% (CI [16, 65]; P = 0.007) compared to no enrichment. Similar results were found when simulating enrichment by other individual plasma biomarkers. Plasma Aβ42/Aβ40 enrichment would reduce sample size by 45% (CI [20, 63], P = 0.003) and plasma NfL would reduce sample size by 41% (CI [5, 63], P = 0.03).

Most importantly, combining all three plasma biomarkers for enrichment reduced the sample size by 70% (CI [54, 81], P < 0.001) compared to no plasma biomarker enrichment for this simulated clinical trial in CU participants aimed at slowing change in PACC by 30% over four years. These results were robust when simulating an identical clinical trial in CU participants using conversion to AD dementia as the primary endpoint, and these results were also robust when testing systematic bias in plasma biomarker levels between -20% and 20%. These results are visualized in the figure below.



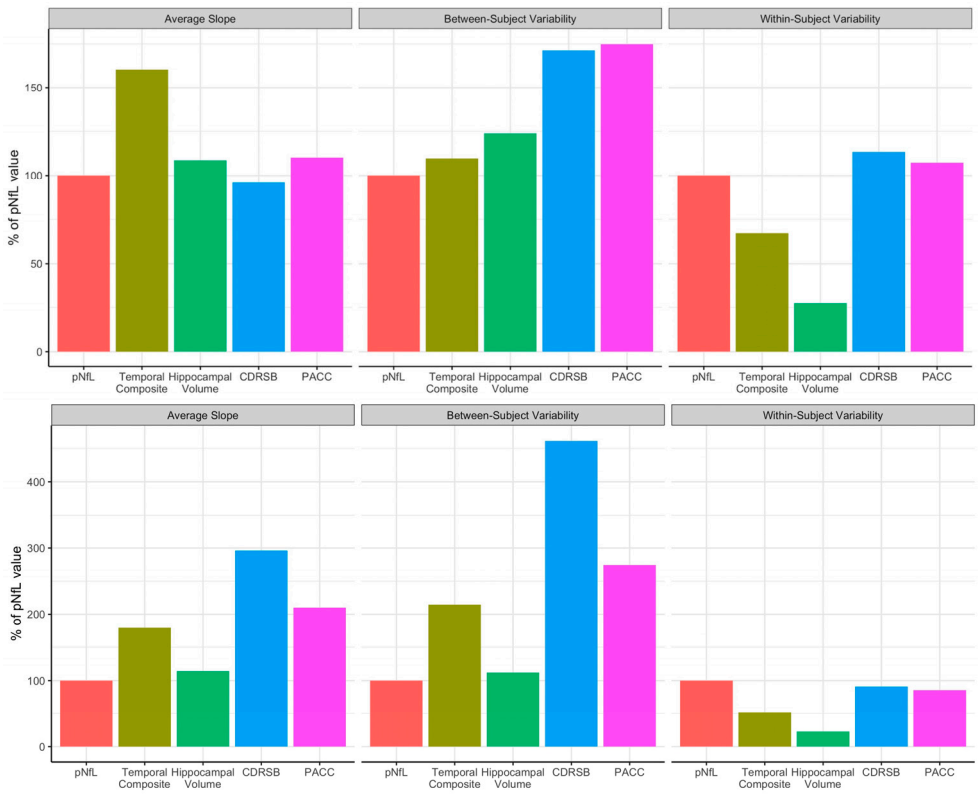
Modelling longitudinal biomarker trajectories

In Paper 4, longitudinal trajectories of plasma NfL, hippocampal volume, a cortical thickness temporal composite ("temporal composite"), CDR-SB, and PACC were estimated in participants from the ADNI study classified as controls, preclinical AD, or mild AD.

In the preclinical AD group, plasma NfL levels did not increase significantly faster over time compared to the control group (dB = 0.04 standard deviations per year, P = 0.10). Moreover, plasma NfL levels were not found to be significantly elevated at baseline in the preclinical AD group compared to controls (dB = 0.21 standard deviations, P = 0.07). All other biomarkers (hippocampal volume, temporal composite, CDR-SB, PACC) saw significantly greater worsening over time in the preclinical AD group compared to controls.

In the mild AD group, plasma NfL levels did worsening significantly faster compared to controls (dB = 0.06 standard deviations per year, P = 0.002) and were also significantly higher at baseline (dB = 0.64 standard deviations, P < 0.0001). All other biomarkers also saw significantly greater worsening over time in the mild AD group compared to the preclinical AD group.

The longitudinal trajectories of the biomarkers described above were estimated using linear mixed-effects models. There are three inherent factors of these LME models which are relevant to determine the statistical power to detect a treatment effect using a given biomarkers: average slope (how much a biomarker changes over time), between-subject variability (how much biomarker slopes vary across individuals), and within-subject variability (how much biomarker values vary for an individual around their slope). In general, the utility of a biomarker to detect treatment effects is better with a higher average slope, lower between-subject variability, and lower within-subject variability. Therefore, these three factors were quantified for the above biomarkers and compared. These results are visualized in the figure below.



The results of this analysis in preclinical AD showed that plasma NfL had the lowest average slope, with cognitive measures (CDR-SB, PACC) having the highest average slope and MRI measures (hippocampal volume, temporal composite) in a middle tier with slightly higher slopes compared to plasma NfL. These results were also qualitatively the same when looking at between-subject variability, but in this case lower between-subject variability implies better utility as a trial endpoint. Looking at within-subject variability (where lower is better) in the preclinical AD group, plasma NfL had the highest levels, followed closely by cognitive measures. MRI measures, on the other hand, had greatly lower within-subject variability levels.

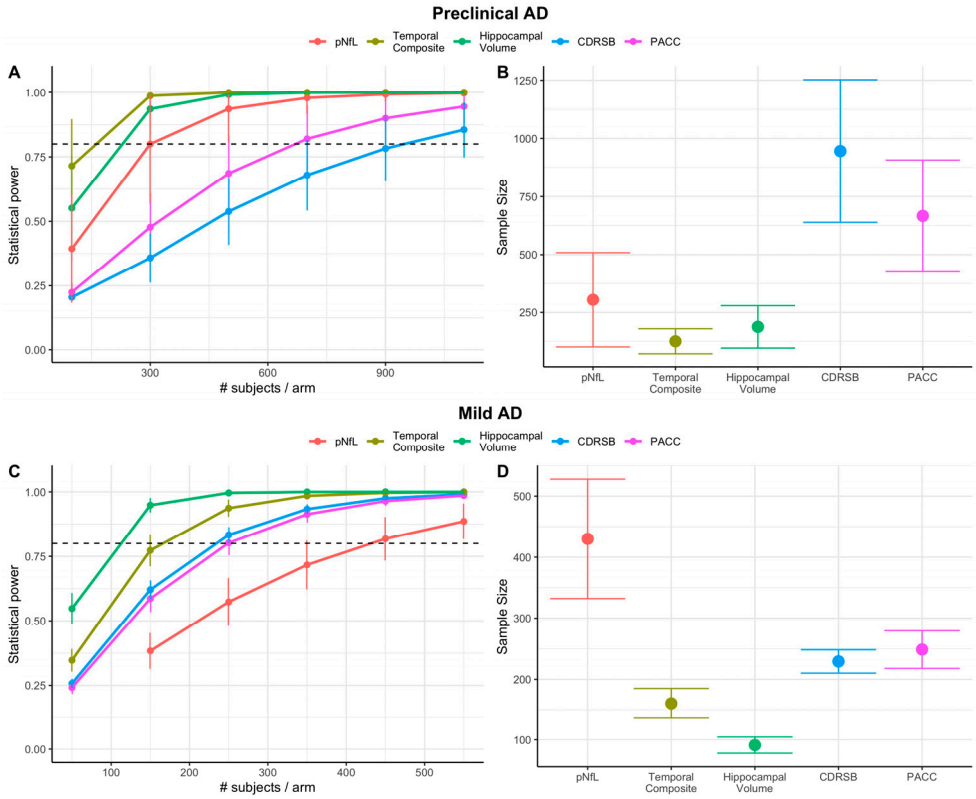
In the mild AD group, the results were similar in the fact that plasma NfL had the lowest average slope and lowest between-subject variability among the biomarkers, but also had the highest within-subject variability levels. Again, MRI measures had extremely low levels of within-subject variability in the mild AD group, but also actually had higher average slope in this group compared to cognitive measures. MRI measures had by far the highest levels of between-subject variability in the mild AD group, indicating a large spread in cognitive change over time among participants in this group.

Trial power analysis

Using the three characteristics of longitudinal trajectories which are derived from linear mixed effects models – average slope, between-subject variability, within-subject variability – a power analysis was performed in which the number of participants needed to detect a 30% reduction in biomarker levels for the preclinical AD group or the mild AD group was calculated. A power analysis also requires specification of key trial parameters, particularly the length of the trial and the frequency of visits (i.e., biomarker measurement). For the primary analysis, the trial length was defined as 30 months for a trial in preclinical AD and 18 months for a trial in mild AD, while the biomarker sampling frequency was defined as every month for plasma NfL and every three months for MRI and cognitive measures.

In a theoretical 30-month clinical trial in preclinical AD patients, using plasma NfL as a progression marker would require 289 participants (CI [76, 496]) to achieve 80% power to detect a 30% treatment effect, while using the MRI-based temporal composite would require 125 participants (CI [78, 172], $P < 0.0001$ compared to plasma NfL) and using hippocampal volume would require 184 participants (CI [91, 278], $P < 0.0001$ compared to plasma NfL). On the other hand, using CDRSB as trial endpoint in a 30-month clinical trial of preclinical AD patients would require 939 participants (CI [634, 1245], $P < 0.0001$ compared to plasma NfL) and PACC

would require 669 participants (CI [430, 909], P < 0.0001 compared to plasma NfL). These results are visualized in the figure below.



In a theoretical 18-month clinical trial in mild AD patients, using plasma NfL as a progression marker would require 432 participants (CI [334, 529]) to achieve 80% power to detect a 30% treatment effect, while using the MRI-based temporal composite would require 161 participants (CI [140, 180], P < 0.0001 compared to plasma NfL) and using hippocampal volume would require 90 participants (CI [76, 104], P < 0.0001 compared to plasma NfL). On the other hand, using CDRSB as trial endpoint in an 18-month clinical trial of mild AD patients would require 230 participants (CI [211, 249], P < 0.0001 compared to plasma NfL) and PACC would require 249 participants (CI [220, 277], P < 0.0001 compared to plasma NfL).

Sensitivity analyses were performed in which the design parameters of trial duration and biomarker measurement frequency were systematically altered across a range of values. Altering such parameters has the effect of placing more or less emphasis on various longitudinal biomarker characteristics – i.e., a trial with more frequent biomarker measurement will be more favorable to biomarkers with high within-

subject variability, and a longer trial will be more favorable to biomarkers with low average slope.

The result of this analysis showed that even when the measurement frequency of plasma NfL was fixed to one month while measurement frequency was increased for MRI and cognition endpoints from one per 3 – 9 months, the utility of plasma NfL as a trial endpoint was still low. In other words, plasma NfL was less efficient than MRI measures for preclinical AD and was less efficient than MRI and cognition measures for mild AD regardless of the less frequent measurement frequency of the other measures.

When increasing the trial duration for a theoretical trial with MRI or cognition as endpoints compared to a theoretical trial with plasma NfL endpoint, however, the results were different. Here, there was a much stronger effect – plasma NfL became as efficient than other MRI and cognition measures for theoretical trials which were between 1.5 – 2 times longer.

Together, these results suggest that increasing sampling frequency of plasma NfL would not make it as efficient as MRI or cognition measures for detecting treatment effects in preclinical AD or mild AD, while increasing trial duration may have a slightly larger effect.

Discussion

The studies presented here investigated some of the possible applications for blood-based biomarkers to improve clinical trials in the future – either indirectly via improved diagnosis and prognosis or directly through clinical trial design decision-making. Blood-based biomarkers were tested for three tasks in particular: screening, enrichment, and tracking response to treatment. These three areas are all vital to improving clinical trials of AD which have been plagued by failure over the past few decades.

The findings related to screening demonstrated that plasma Aβ42/Aβ40, in combination with *APOE* and age, could greatly reduce the number of amyloid PET scans that are needed to recruit and include enough participants for secondary prevention trials of AD in CU individuals. Reducing the number of PET scans is a major need for such trials since the prevalence of amyloid positivity is low in CU elderly individuals.

Interestingly, the cost-benefit analysis for plasma screening demonstrated that the optimal relative risk threshold to refer individuals further for amyloid PET scanning was around 50%. In other words, this finding suggests that around half of those CU elderly individuals who receive a plasma screening test should be referred for a PET scan. As the plasma screening risk threshold becomes stricter, the expected amyloid PET positivity rate will increase among those who receive an amyloid PET scan. However, the number of plasma screening tests which must be performed in this case also increases. Therefore, the tradeoff between being increasing or decreasing the risk threshold required to refer individuals for a PET scan depends greatly on the cost ratio between PET and plasma. A higher PET-to-plasma cost ratio (e.g., at 16x as was tested) suggests that fewer individuals should be referred for PET scans as there is a strong incentive to avoid negative (i.e., normal) amyloid PET scans.

A major limitation of the screening work is that it did not consider the contribution of multiple blood-based biomarkers. It is likely that including multiple plasma biomarkers such as pTau181 or pTau217 could greatly increase the screening utility. Still, the panel tested in the study is the only blood-based biomarker test which is commercially available (West et al. 2021). This fact makes it highly relevant to investigate its performance in the BioFINDER cohort.

Another potential limitation of the plasma screening analysis is that it did not take into consideration the ethical aspects of denying potentially amyloid PET positive

participants access to an amyloid PET scan. In other words, there was no consideration of the sensitivity of the plasma screening test for detecting true amyloid PET positive individuals. In practice, such a strategy may result in the exclusion from clinical trials of a large number of CU elderly individuals who would eventually develop AD (Visser et al. 2008; Insel et al. 2020). It is likely more applicable to instead consider minimum thresholds for detection sensitivity in MCI and AD populations, and then compare which blood-based biomarker(s) perform best for screening while meeting the minimum sensitivity threshold.

Nonetheless, blood-based biomarker screening has important role to play in the future of clinical trials for AD (Leuzy et al. 2022). There are already efforts to incorporate blood-based biomarkers in screening pipelines such as those developed by "trial-ready cohort" initiatives (Aisen et al. 2020). The way such trial-ready cohort initiatives have set up their screening pipeline is by first making a short cognitive screening test available online from individuals in the community, and then inviting individuals for further testing if they meet criteria based on the first cognitive and/or demographic screening tests. In this way, blood-based biomarkers are still likely to be the second line of screening after demographics and cognition. It is highly important that blood-based biomarkers are tested in combination with demographic and cognition variables, rather than as a replacement. These biomarkers should not a replacement for cognitive assessment.

The next set of studies investigated the performance of core blood-based biomarkers of AD to predict cognitive decline in an MCI population (Paper 2) and a CU population (Paper 3). These biomarker models of cognitive decline were then considered in the context of clinical trial enrichment, where inclusion into a clinical trial would involve individuals who are predicted to show cognitive decline during the trial. Enrichment is important because many participants remain stable during AD clinical trials due to the variable nature of disease progression (Freidlin and Korn 2014; Holland, McEvoy, Desikan, et al. 2012; Kerr et al. 2017).

For predicting cognition in MCI patients, the combination of plasma pTau181 and plasma NfL was consistently the most parsimonious model across the different outcomes and sensitivity analyses that were performed. Plasma pTau181 provided the strongest independent information, while plasma NfL was more relevant for non-AD specific outcomes such as predicting risk of any type of dementia rather than predicting AD dementia specifically. This result cements the role of plasma NfL as an important blood-based biomarker for AD – a result which only strengthened by its importance in other neurodegenerative diseases (Byrne et al. 2017; Gisslén et al. 2016; Lewczuk et al. 2018). Importantly, this blood-based biomarker panel was tested relative to a basic model consisting of age, sex, education, and baseline MMSE score. Any P-values or comparisons were made first relative to this basic model. Having a fair model to compare with is greatly important when investigating blood-based biomarkers for AD prognosis.

The combination of plasma pTau181 and NfL performed well in out-of-sample testing across the BioFINDER and ADNI cohorts. Interestingly, the test performance of the panel was better when fitting the model on BioFINDER and testing the model on ADNI, rather than vice-versa. This supports the idea that deriving biomarker cutoffs or model coefficients from more symptomatically diverse and representative population goes a long way to improve the generalizability of biomarker models (Barnes 2019). Conversely, more pure amnestic AD cohorts like ADNI may be more suitable for investigating biological questions related to AD but less suitable on their own for developing generalizable biomarker models for diagnosis or prognosis in more representative populations. In any case, much more work is needed to ensure that biomarker-based clinical prediction models in AD can generalize in different scenarios and populations (Steyerberg 2019).

In a CU population, the results when testing the entire combination of plasma Aβ42/Aβ40, pTau217, and NfL were quite similar. Both pTau217 and NfL were independently significant in predicting AD-related outcomes such as cognitive decline and risk of AD dementia – with the importance of plasma NfL increasing when predicting risk of all-cause dementia. One difference in this cohort was that plasma Aβ42/Aβ40 also was important for prediction of AD-related outcomes. It is likely that the increased importance of plasma Aβ42/Aβ40 has to do with a combination of two factors: 1) improved assay technology between Study 2 in MCI patients and Study 3 in CU individuals, and 2) a potential greater importance of knowing amyloid information in CU populations (Wang et al. 2017; Reitz 2012).

While plasma Aβ42/Aβ40 was important in this population, however, adding *APOE* status to the combined model greatly decreased the effect size of plasma Aβ42/Aβ40 in predicting cognitive decline and risk for AD dementia. This supports what is already known – that *APOE* e4 status primarily affects amyloid-related processes – and additionally suggests that knowing *APOE* e4 status may partially substitute on a group-level for knowing plasma amyloid status (Belloy, Napolioni, and Greicius 2019; O'Donoghue et al. 2018).

This result from Study 3 in a CU population suggests that the core plasma ATN biomarkers provided significant prognostic information. Combining the plasma biomarkers was much better than individual biomarkers. Still, the overall model performance was quite low, even for the combined model. This suggests one of three things: 1) that there is still room to improve plasma ATN biomarkers to make them more accurate, 2) that it is important to add other plasma biomarkers which reflect other biological processes such as vascular disease, inflammation, etc., and 3) that the measurement error (i.e., "noise") related to measuring cognition in CU participants puts a low ceiling on how well future outcomes can be predicted.

All three of these ideas presented above likely contribute to the low overall model performance of plasma biomarkers in CU populations. The finding in Paper 3 that

corresponding CSF biomarkers greatly outperform plasma biomarkers suggests that there may be room to improve plasma biomarkers further to make them correspond more closely with their CSF counterparts.

Testing the prognostic models for clinical trial enrichment in a CU population showed that between 40 – 50% reduction in trial sample size could be achieved by using individual plasma biomarkers. This reduction improved even further to 70%+ when combining plasma biomarkers. Looking at the number of participants required to detect treatment effects in CU populations, it is hard to see how some form of screening or enrichment can be avoided (Holland, McEvoy, Dale, et al. 2012; Huang et al. 2017). The issue of running a clinical trial in this population is simply that the overwhelming majority of participants would remain stable – their cognitive trajectories remaining completely flat – during the trial. This phenomenon would make it extremely difficult to detect a significant group-level difference in cognitive trajectories between the placebo and treatment groups.

Importantly for the enrichment results, simulating large, systematic shifts (i.e., bias) in plasma biomarker cutoffs did not significantly decrease the sample size reduction seen by using these plasma biomarkers for enrichment. Such systematic shifts may be rare but may be more likely for plasma biomarkers compared to CSF biomarkers (Mattsson et al. 2013). In other words, even in the worst-case scenario where plasma biomarker assays do not perform as well as expected, it would still be preferable to enrich with plasma biomarkers rather than do nothing.

While the first three papers were focused on design decisions which must be undertaken before the trial begins, the focus of Paper 4 was on an important design decision which affects execution of the trial itself – namely, selecting trial endpoints. Determining which biomarkers to measure during a clinical trial is important because biomarker endpoints have a large influence on whether a trial will continue to Phase III from Phase II studies, as well as whether a Phase III trial has a high likelihood of detecting a true treatment effect – i.e., power.

Evaluating biomarkers as potential endpoints requires first investigating group-level and individual-level trajectories of the biomarkers when measured longitudinally. This was explored in Paper 4 for plasma NfL, two MRI-based measures, and cognition. A primary application of this analysis can be considered the comparison of MRI versus plasma NfL as biomarkers of general neurodegeneration during clinical trials of AD, with cognition as a key reference point.

The results from this investigation showed that plasma NfL had both a high within-subject variability compared to MRI, as well as a low or similar change over time compared to MRI. In fact, plasma NfL was the only biomarker tested whose longitudinal change over time was not significantly greater in the preclinical AD group compared to controls; although this is likely due to the fact that the groups really only differed by amyloid positivity. In any case, the high within-subject variability – i.e., the finding that plasma NfL levels often jump up and down

drastically for an individual across visits – greatly hinders its ability as a biomarker endpoint.

And indeed, taking the estimated longitudinal trajectories and applying them to a power analysis showed that plasma NfL was greatly outperformed by MRI-based measures in terms of efficiency to detect a treatment effect in clinical trials of preclinical AD or mild AD. The main reason for this finding likely goes back to the within-subject variability: it is high for plasma NfL and extremely low for MRI. The finding that MRI is preferable to plasma NfL for "standard" clinical trials of AD was not changed when biomarker measurement frequency in the simulated trial was increased or when the length of the simulated trial increased. The properties of MRI-based measures are simply too strong.

Nonetheless, clinical trials are not run via simulation. The use of MRI in clinical trials of AD has been widespread, yet recent trials of AD have actually shown adverse effects on MRI-based measures (Howard and Liu 2020; Swanson et al. 2021; Mintun et al. 2021). This demonstrates how simulation and reality can be at odds, making integration of biological findings and assumptions into clinical trial design a key to success.

Conclusion

The future of clinical trials for AD involves clever integration of blood-based biomarkers at various stages. The open questions when this thesis was started were best described by the classic "who, what, where, when, and how?" In other words, the open questions were many.

Work from this thesis has shown that blood-based biomarkers are theoretically likely to perform well for pre-screening to identify individuals across the AD spectrum who are likely to be amyloid PET positive. This may even be possible for tau PET positivity, as well. It can be inferred from these results that blood-based biomarkers should be employed to reduce the number of PET scans or lumbar punctures in AD clinical trials in the future.

For enrichment, on the other hand, the results for blood-based biomarkers are less obvious. Despite the fact that blood-based biomarkers relate to cognitive decline, they do not relate so strongly when compared to corresponding proteins measured via CSF or PET. Perhaps a few smaller, Phase II trials may be run based solely on blood-based biomarker levels, but it does not appear that PET or CSF collection is likely to being taken out of service any time soon. In fact, it will probably only grow more important with the future approval of AD therapies likely to require confirmation of amyloid and/or tau pathology. And because PET or CSF will be required for inclusion and enrichment decisions occur after participants are already confirmed to meet inclusion criteria, then they will always be available for enrichment.

Finally, the use of blood-based biomarkers to measure treatment effects during AD clinical trials is one of the most exciting frontiers in this area of research. The idea of gaining a fine-grained understanding of how participants respond to various treatments seems on the horizon via blood-based biomarkers. These biomarkers allow for frequent analysis and track AD-related processes well enough to provide biological insights which can inform future trial decisions. Blood-based biomarkers have already become an integral progression biomarker for recent AD clinical trials, and their role in this area is likely to increase in the future.

Acknowledgements

Thank you to everyone from the BioFINDER team, including participants and family members of participants in the BioFINDER study. Thank you as well to those from the ADNI study.

References

Aisen, P S. 2015. “Cognitive/Clinical Endpoints for Pre-Dementia AD Trials.” *The Journal of Prevention of Alzheimer’s Disease* 2 (2): 82–84.

Aisen, P S, R A Sperling, J Cummings, M C Donohue, O Langford, G A Jimenez-Maggiora, R A Rissman, et al. 2020. “THE TRIAL-READY COHORT FOR PRECLINICAL/PRODROMAL ALZHEIMER’S DISEASE (TRC-PAD) PROJECT: AN OVERVIEW.” *The Journal of Prevention of Alzheimer’s Disease* 7 (4): 1–5. doi:10.14283/jpad.2020.45.

Anderson, Roy M, Christoforos Hadjichrysanthou, Stephanie Evans, and Mei Mei Wong. 2017. “Why Do so Many Clinical Trials of Therapies for Alzheimer’s Disease Fail?” *The Lancet* 390 (10110): 2327–29. doi:10.1016/s0140-6736(17)32399-1.

Barnes, Lisa L. 2019. “Biomarkers for Alzheimer Dementia in Diverse Racial and Ethnic Minorities—A Public Health Priority.” *JAMA Neurology* 76 (3): 251–53. doi:10.1001/jamaneurol.2018.3444.

Belloy, Michaël E., Valerio Napolioni, and Michael D. Greicius. 2019. “A Quarter Century of APOE and Alzheimer’s Disease: Progress to Date and the Path Forward.” *Neuron* 101 (5): 820–38. doi:10.1016/j.neuron.2019.01.056.

Betthauser, Tobey J, Rebecca L Koscik, Erin M Jonaitis, Samantha L Allison, Karly A Cody, Claire M Erickson, Howard A Rowley, et al. 2019. “Amyloid and Tau Imaging Biomarkers Explain Cognitive Decline from Late Middle-Age.” *Brain* 143 (1): 320–35. doi:10.1093/brain/awz378.

Brookmeyer, Ron, Nada Abdalla, Claudia H. Kawas, and María M. Corrada. 2018. “Forecasting the Prevalence of Preclinical and Clinical Alzheimer’s Disease in the United States.” *Alzheimer’s & Dementia* 14 (2): 121–29. doi:10.1016/j.jalz.2017.10.009.

Byrne, Lauren M, Filipe B Rodrigues, Kaj Blennow, Alexandra Durr, Blair R Leavitt, Raymund Roos, Rachael I Scahill, et al. 2017. “Neurofilament Light Protein in Blood as a Potential Biomarker of Neurodegeneration in Huntington’s Disease: A Retrospective Cohort Analysis.” *The Lancet Neurology* 16 (8): 601–9. doi:10.1016/s1474-4422(17)30124-2.

Carotenuto, Antonio, Beniamino Giordano, George Dervenoulas, Heather Wilson, Mattia Veronese, Zachary Chappell, Sotirios Polychronis, et al. 2020. "[18F]Florbetapir PET/MR Imaging to Assess Demyelination in Multiple Sclerosis." *European Journal of Nuclear Medicine and Molecular Imaging* 47 (2): 366–78. doi:10.1007/s00259-019-04533-y.

Chapman, Kimberly R, Hanaan Bing-Canar, Michael L Alosco, Eric G Steinberg, Brett Martin, Christine Chaisson, Neil Kowall, Yorghos Tripodis, and Robert A Stern. 2016. "Mini Mental State Examination and Logical Memory Scores for Entry into Alzheimer's Disease Trials." *Alzheimer's Research & Therapy* 8 (1): 9. doi:10.1186/s13195-016-0176-z.

Cummings, Jeffrey. 2018. "Lessons Learned from Alzheimer Disease: Clinical Trials with Negative Outcomes." *Clinical and Translational Science* 11 (2): 147–52. doi:10.1111/cts.12491.

Cummings, Jeffrey L, Travis Morstorf, and Kate Zhong. 2014. "Alzheimer's Disease Drug-Development Pipeline: Few Candidates, Frequent Failures." *Alzheimer's Research & Therapy* 6 (4): 37. doi:10.1186/alzrt269.

Donohue, M. C., and P. S. Aisen. 2012. "Mixed Model of Repeated Measures versus Slope Models in Alzheimer's Disease Clinical Trials." *The Journal of Nutrition, Health & Aging* 16 (4): 360–64. doi:10.1007/s12603-012-0047-7.

Donohue, Michael C, Reisa A Sperling, David P Salmon, Dorene M Rentz, Rema Raman, Ronald G Thomas, Michael Weiner, et al. 2014. "The Preclinical Alzheimer Cognitive Composite: Measuring Amyloid-Related Decline." *JAMA Neurology* 71 (8): 961–70. doi:10.1001/jamaneurol.2014.803.

Dubois, Bruno, Howard H Feldman, Claudia Jacova, Steven T DeKosky, Pascale Barberger-Gateau, Jeffrey Cummings, André Delacourte, et al. 2007. "Research Criteria for the Diagnosis of Alzheimer's Disease: Revising the NINCDS–ADRDA Criteria." *The Lancet Neurology* 6 (8): 734–46. doi:10.1016/s1474-4422(07)70178-3.

Fischl, Bruce. 2011. "FreeSurfer." *NeuroImage* 62 (2): 774–81. doi:10.1016/j.neuroimage.2012.01.021.

Freidlin, Boris, and Edward L. Korn. 2014. "Biomarker Enrichment Strategies: Matching Trial Design to Biomarker Credentials." *Nature Reviews Clinical Oncology* 11 (2): 81–90. doi:10.1038/nrclinonc.2013.218.

Gisslén, Magnus, Richard W. Price, Ulf Andreasson, Niklas Norgren, Staffan Nilsson, Lars Hagberg, Dietmar Fuchs, Serena Spudich, Kaj Blennow, and Henrik Zetterberg. 2016. "Plasma Concentration of the Neurofilament Light Protein (NFL) Is a Biomarker of CNS Injury in HIV Infection: A Cross-Sectional Study." *EBioMedicine* 3: 135–40. doi:10.1016/j.ebiom.2015.11.036.

Hampel, Harald, Sid E O'Bryant, José L Molinuevo, Henrik Zetterberg, Colin L Masters, Simone Lista, Steven J Kiddle, Richard Batrla, and Kaj Blennow. 2018. "Blood-Based Biomarkers for Alzheimer Disease: Mapping the Road to the Clinic." *Nature Reviews. Neurology* 14 (11): 639–52. doi:10.1038/s41582-018-0079-7.

Hansson, Oskar. 2021. "Biomarkers for Neurodegenerative Diseases." *Nature Medicine* 27 (6): 954–63. doi:10.1038/s41591-021-01382-x.

Holland, Dominic, Linda K McEvoy, Anders M Dale, and Alzheimer's Disease Neuroimaging Initiative. 2012. "Unbiased Comparison of Sample Size Estimates from Longitudinal Structural Measures in ADNI." *Human Brain Mapping* 33 (11): 2586–2602. doi:10.1002/hbm.21386.

Holland, Dominic, Linda K McEvoy, Rahul S Desikan, Anders M Dale, and Alzheimer's Disease Neuroimaging Initiative. 2012. "Enrichment and Stratification for Predementia Alzheimer Disease Clinical Trials." *PLoS ONE* 7 (10): e47739. doi:10.1371/journal.pone.0047739.

Howard, Robert, and Kathy Y. Liu. 2020. "Questions EMERGE as Biogen Claims Aducanumab Turnaround." *Nature Reviews Neurology* 16 (2): 63–64. doi:10.1038/s41582-019-0295-9.

Huang, Zhiyue, Graciela Muniz-Terrera, Brian D.M. Tom, and Alzheimer's Disease Neuroimaging Initiative. 2017. "Power Analysis to Detect Treatment Effects in Longitudinal Clinical Trials for Alzheimer's Disease." *Alzheimer's & Dementia: Translational Research & Clinical Interventions* 3 (3): 360–66. doi:10.1016/j.trci.2017.04.007.

Insel, Philip S., Michael C. Donohue, Reisa Sperling, Oskar Hansson, and Niklas Mattsson-Carlgren. 2020. "The A4 Study: B-amyloid and Cognition in 4432 Cognitively Unimpaired Adults." *Annals of Clinical and Translational Neurology* 7 (5): 776–85. doi:10.1002/acn3.51048.

Insel, Philip S, Niklas Mattsson, R Scott Mackin, John Kornak, Rachel Nosheny, Duygu Tosun-Turgut, Michael C Donohue, Paul S Aisen, Michael W Weiner, and Alzheimer's Disease Neuroimaging Initiative. 2015. "Biomarkers and Cognitive Endpoints to Optimize Trials in Alzheimer's Disease." *Annals of Clinical and Translational Neurology* 2 (5): 534–47. doi:10.1002/acn3.192.

Jack, Clifford R., Heather J. Wiste, Stephen D. Weigand, David S. Knopman, Michelle M. Mielke, Prashanthi Vemuri, Val Lowe, et al. 2015. "Different Definitions of Neurodegeneration Produce Similar Amyloid/Neurodegeneration Biomarker Group Findings." *Brain* 138 (Pt 12): 3747–59. doi:10.1093/brain/awv283.

Jagust, William J., Susan M. Landau, Robert A. Koeppe, Eric M. Reiman, Kewei Chen, Chester A. Mathis, Julie C. Price, Norman L. Foster, and Angela Y. Wang. 2015. “The Alzheimer’s Disease Neuroimaging Initiative 2 PET Core: 2015.” *Alzheimer’s & Dementia* 11 (7): 757–71. doi:10.1016/j.jalz.2015.05.001.

Janelidze, Shorena, Niklas Mattsson, Sebastian Palmqvist, Ruben Smith, Thomas G Beach, Geidy E Serrano, Xiyun Chai, et al. 2020. “Plasma P-Tau181 in Alzheimer’s Disease: Relationship to Other Biomarkers, Differential Diagnosis, Neuropathology and Longitudinal Progression to Alzheimer’s Dementia.” *Nature Medicine* 26 (3): 379–86. doi:10.1038/s41591-020-0755-1.

Janelidze, Shorena, Josef Pannee, Alvydas Mikulskis, Ping Chiao, Henrik Zetterberg, Kaj Blennow, and Oskar Hansson. 2017. “Concordance Between Different Amyloid Immunoassays and Visual Amyloid Positron Emission Tomographic Assessment.” *JAMA Neurology* 74 (12): 1492. doi:10.1001/jamaneurol.2017.2814.

Janelidze, Shorena, Charlotte E. Teunissen, Henrik Zetterberg, José Antonio Allué, Leticia Sarasa, Udo Eichenlaub, Tobias Bittner, et al. 2021. “Head-to-Head Comparison of 8 Plasma Amyloid-β 42/40 Assays in Alzheimer Disease.” *JAMA Neurology* 78 (11): 1–8. doi:10.1001/jamaneurol.2021.3180.

Jansen, Willemijn J, Rik Ossenkoppele, Dirk L Knol, Betty M Tijms, Philip Scheltens, Frans R J Verhey, Pieter Jelle Visser, et al. 2015. “Prevalence of Cerebral Amyloid Pathology in Persons Without Dementia: A Meta-Analysis.” *JAMA* 313 (19): 1924–38. doi:10.1001/jama.2015.4668.

Karikari, Thomas K, Tharick A Pascoal, Nicholas J Ashton, Shorena Janelidze, Andréa Lessa Benedet, Juan Lantero Rodriguez, Mira Chamoun, et al. 2020. “Blood Phosphorylated Tau 181 as a Biomarker for Alzheimer’s Disease: A Diagnostic Performance and Prediction Modelling Study Using Data from Four Prospective Cohorts.” *The Lancet Neurology* 19 (5): 422–33. doi:10.1016/s1474-4422(20)30071-5.

Kerr, Kathleen F, Jeremy Roth, Kehao Zhu, Heather Thiessen-Philbrook, Allison Meisner, Francis Perry Wilson, Steven Coca, and Chirag R Parikh. 2017. “Evaluating Biomarkers for Prognostic Enrichment of Clinical Trials.” *Clinical Trials* 14 (6): 629–38. doi:10.1177/1740774517723588.

Keshavan, Ashvini, Josef Pannee, Thomas K Karikari, Juan Lantero Rodriguez, Nicholas J Ashton, Jennifer M Nicholas, David M Cash, et al. 2021. “Population-Based Blood Screening for Preclinical Alzheimer’s Disease in a British Birth Cohort at Age 70.” *Brain* 144 (2): 434–49. doi:10.1093/brain/awaa403.

Leuzy, Antoine, Niklas Mattsson-Carlgren, Sebastian Palmqvist, Shorena Janelidze, Jeffrey L Dage, and Oskar Hansson. 2022. “Blood-based Biomarkers for Alzheimer’s Disease.” *EMBO Molecular Medicine* 14 (1): e14408. doi:10.15252/emmm.202114408.

Lewczuk, Piotr, Natalia Ermann, Ulf Andreasson, Christian Schultheis, Jana Podhorna, Philipp Spitzer, Juan M Maler, Johannes Kornhuber, Kaj Blennow, and Henrik Zetterberg. 2018. “Plasma Neurofilament Light as a Potential Biomarker of Neurodegeneration in Alzheimer’s Disease.” *Alzheimer’s Research & Therapy* 10 (1): 71. doi:10.1186/s13195-018-0404-9.

Mattsson, Niklas , Ulf Andreasson, Henrik Zetterberg, and Kaj Blennow. 2017. “Association of Plasma Neurofilament Light With Neurodegeneration in Patients With Alzheimer Disease.” *JAMA Neurology*. doi:10.1001/jamaneurol.2016.6117.

Mattsson, Niklas, Ulf Andreasson, Staffan Persson, Maria C. Carrillo, Steven Collins, Sonia Chalbot, Neal Cutler, et al. 2013. “CSF Biomarker Variability in the Alzheimer’s Association Quality Control Program.” *Alzheimer’s & Dementia* 9 (3): 251–61. doi:10.1016/j.jalz.2013.01.010.

Mattsson, Niklas, Nicholas C. Cullen, Ulf Andreasson, Henrik Zetterberg, and Kaj Blennow. 2019. “Association Between Longitudinal Plasma Neurofilament Light and Neurodegeneration in Patients With Alzheimer Disease.” *JAMA Neurology* 76 (7). doi:10.1001/jamaneurol.2019.0765.

McKhann, Guy M., David S. Knopman, Howard Chertkow, Bradley T. Hyman, Clifford R. Jack, Claudia H. Kawas, William E. Klunk, et al. 2011. “The Diagnosis of Dementia Due to 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): 263–69. doi:10.1016/j.jalz.2011.03.005.

Mintun, Mark A., Albert C. Lo, Cynthia Duggan Evans, Alette M. Wessels, Paul A. Ardayfio, Scott W. Andersen, Sergey Shcherbinin, et al. 2021. “Donanemab in Early Alzheimer’s Disease.” *New England Journal of Medicine* 384 (18): 1691–1704. doi:10.1056/nejmoa2100708.

O’Donoghue, M. Clare, Susannah E. Murphy, Giovanna Zamboni, Anna C. Nobre, and Clare E. Mackay. 2018. “APOE Genotype and Cognition in Healthy Individuals at Risk of Alzheimer’s Disease: A Review.” *Cortex* 104: 103–23. doi:10.1016/j.cortex.2018.03.025.

Palmqvist, Sebastian, Shorena Janelidze, Yakeel T. Quiroz, Henrik Zetterberg, Francisco Lopera, Erik Stomrud, Yi Su, et al. 2020. “Discriminative Accuracy of Plasma Phospho-Tau217 for Alzheimer Disease vs Other Neurodegenerative Disorders.” *JAMA* 324 (8): 772–81. doi:10.1001/jama.2020.12134.

Palmqvist, Sebastian, Shorena Janelidze, Erik Stomrud, Henrik Zetterberg, Johann Karl, Katharina Zink, Tobias Bittner, et al. 2019. “Performance of Fully Automated Plasma Assays as Screening Tests for Alzheimer Disease–Related β-Amyloid Status.” *JAMA Neurology* 76 (9): 1060–69. doi:10.1001/jamaneurol.2019.1632.

Palmqvist, Sebastian, Henrik Zetterberg, Kaj Blennow, Susanna Vestberg, Ulf Andreasson, David J. Brooks, Rikard Owenius, et al. 2014. “Accuracy of Brain Amyloid Detection in Clinical Practice Using Cerebrospinal Fluid β-Amyloid 42: A Cross-Validation Study Against Amyloid Positron Emission Tomography.” *JAMA Neurology* 71 (10): 1282–89. doi:10.1001/jamaneurol.2014.1358.

Palmqvist, Sebastian, Henrik Zetterberg, Niklas Mattsson, Per Johansson, Lennart Minthon, Kaj Blennow, Mattias Olsson, et al. 2015. “Detailed Comparison of Amyloid PET and CSF Biomarkers for Identifying Early Alzheimer Disease.” *Neurology* 85 (14): 1240–49. doi:10.1212/wnl.0000000000001991.

Petersen, R. C., P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, et al. 2010. “Alzheimer’s Disease Neuroimaging Initiative (ADNI) Clinical Characterization.” *Neurology* 74 (3): 201–9. doi:10.1212/wnl.0b013e3181cb3e25.

Reitz, Christiane. 2012. “Alzheimer’s Disease and the Amyloid Cascade Hypothesis: A Critical Review.” *International Journal of Alzheimer’s Disease* 2012: 369808. doi:10.1155/2012/369808.

Selkoe, Dennis J. 2019. “Alzheimer Disease and Aducanumab: Adjusting Our Approach.” *Nature Reviews Neurology* 15 (7): 365–66. doi:10.1038/s41582-019-0205-1.

Sevigny, Jeff, Joyce Suhy, Ping Chiao, Tianle Chen, Gregory Klein, Derk Purcell, Joonmi Oh, Ajay Verma, Mehul Sampat, and Jerome Barakos. 2016. “Amyloid PET Screening for Enrichment of Early-Stage Alzheimer Disease Clinical Trials.” *Alzheimer Disease & Associated Disorders* 30 (1): 1–7. doi:10.1097/wad.0000000000000144.

Steyerberg, Ewout W. 2019. “Clinical Prediction Models.” *Statistics for Biology and Health*, 59–93. doi:10.1007/978-3-030-16399-0_4.

Swanson, Chad J., Yong Zhang, Shobha Dhadda, Jinping Wang, June Kaplow, Robert Y. K. Lai, Lars Lannfelt, et al. 2021. “A Randomized, Double-Blind, Phase 2b Proof-of-Concept Clinical Trial in Early Alzheimer’s Disease with Lecanemab, an Anti-Aβ Protofibril Antibody.” *Alzheimer’s Research & Therapy* 13 (1): 80. doi:10.1186/s13195-021-00813-8.

Thambisetty, Madhav, Robert Howard, M. Maria Glymour, and Lon S. Schneider. 2021. “Alzheimer’s Drugs: Does Reducing Amyloid Work?” *Science (New York, N.Y.)* 374 (6567): 544–45. doi:10.1126/science.abl8366.

Thijssen, Elisabeth H, Renaud La Joie, Amelia Strom, Corrina Fonseca, Leonardo Iaccarino, Amy Wolf, Salvatore Spina, et al. 2020. “Plasma Phosphorylated Tau 217 and Phosphorylated Tau 181 as Biomarkers in Alzheimer’s Disease and Frontotemporal Lobar Degeneration: A Retrospective Diagnostic Performance Study.” *The Lancet. Neurology* 20 (9): 739–52. doi:10.1016/s1474-4422(21)00214-3.

Veitch, Dallas P., Michael W. Weiner, Paul S. Aisen, Laurel A. Beckett, Nigel J. Cairns, Robert C. Green, Danielle Harvey, et al. 2018. “Understanding Disease Progression and Improving Alzheimer’s Disease Clinical Trials: Recent Highlights from the Alzheimer’s Disease Neuroimaging Initiative.” *Alzheimer’s & Dementia* 15 (1): 106–52. doi:10.1016/j.jalz.2018.08.005.

Visser, P.J., F.R.J. Verhey, M. Boada, R. Bullock, P.P. De Deyn, G.B. Frisoni, L. Frölich, et al. 2008. “Development of Screening Guidelines and Clinical Criteria for Predementia Alzheimer’s Disease.” *Neuroepidemiology* 30 (4): 254–65. doi:10.1159/000135644.

Wang, Jun, Ben J Gu, Colin L Masters, and Yan-Jiang Wang. 2017. “A Systemic View of Alzheimer Disease — Insights from Amyloid-β Metabolism beyond the Brain.” *Nature Reviews Neurology* 13 (10): 612–23. doi:10.1038/nrneurol.2017.111.

West, Tim, Kristopher M. Kirmess, Matthew R. Meyer, Mary S. Holubasch, Stephanie S. Knapik, Yan Hu, John H. Contois, et al. 2021. “A Blood-Based Diagnostic Test Incorporating Plasma Aβ42/40 Ratio, ApoE Proteotype, and Age Accurately Identifies Brain Amyloid Status: Findings from a Multi Cohort Validity Analysis.” *Molecular Neurodegeneration* 16 (1): 30. doi:10.1186/s13024-021-00451-6.

Zetterberg, Henrik. 2017. “Applying Fluid Biomarkers to Alzheimer’s Disease.” *American Journal of Physiology. Cell Physiology* 313 (1): C3–10. doi:10.1152/ajpcell.00007.2017.

Zetterberg, Henrik, and Kaj Blennow. 2013. “Cerebrospinal Fluid Biomarkers for Alzheimer’s Disease: More to Come?” *Journal of Alzheimer’s Disease : JAD* 33 Suppl 1: S361-9.

Printed by Media-Tryck, Lund 2022 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences, Malmö

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2022:128
ISBN 978-91-8021-290-8
ISSN 1652-8220