PhD-FHSE-2024-016
The Faculty of Humanities, Education and Social Sciences

DISSERTATION

Defence held on 06/05/2024 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN PSYCHOLOGIE

by

Matthias KLEE
Born on 30 October 1994 in Bad Kreuznach (Germany)

INVESTIGATION OF MODIFIABLE SOCIAL AND BEHAVIOURAL RISK FACTORS OF COGNITIVE DECLINE AND DEMENTIA

Dissertation defence committee

Dr Anja Leist, dissertation supervisor
Professor, Université du Luxembourg

Dr David Rehkopf
Professor, Stanford University

Dr Anna Elena Kornadt, Chair
Professor, Université du Luxembourg

Dr Pavla Brennan Kearns
Professor, Charles University

Dr Pedro Cardoso-Leite, Vice-Chair
Professor, Université du Luxembourg

“This is the age of the mechanism.”

(Anja K. Leist, personal communication)

Dedication

To my parents, Andrea, and Armin Klee, who never expected me to do more than finding my way, who never doubted me, and whom I will never cease to be grateful to, for raising me with support and love, unconditionally.

To my fiancé, my partner, my backup, my love, Deborah. With you by my side, growing old cannot not be successful.

To my future self, in the hope it reminds me of how no single person may ever flourish without support, and that everyone may reciprocate.

.

Acknowledgements

Recapitulating the last three years, I met many people that accompanied, guided, and supported me, on the way to bringing this thesis to paper.

Without doubt this thesis would not have been possible without the dedicated and engaged supervision of Anja Leist who always supported me in finding my path through the initially overwhelming endeavour. This support not only entailed professional guidance, and connecting me to global collaborators, but extended to simultaneously broadening and focussing my scope, to start to formulate my own research in the field of cognitive aging. I would like to thank you for supporting me to grow into the role of a colleague and mentoring me along the way.

I would further like to thank and acknowledge the support I have received from all members of my supervisory committee, especially David Rehkopf who motivated me to delve further into the field of social epidemiology, Pedro Cardoso-Leite who continuously drove me to dive deeper thinking about constructs, research questions and implications, and Benedikt Wilbertz, for his support in the realm of machine learning.

I am deeply honoured and thank Pavla Brennan Kearns, to serve as an external member of the thesis defence committee.

I would further like to thank Anna Elena Kornadt to serve as chair, and for her support as a mentor.

I am further grateful for having had the opportunity to conduct my research in the Institute for Research on Socio-Economic Inequality. Being part of the interdisciplinary team not only helped me to better understand my research questions in individuals, embedded in their environments and communities, but also provided an incredibly supportive community of colleagues. I would thus like to thank Louis Chauvel, Sofie Van Herzeele, who always

supported me, but especially in times of crisis, as well as all other members of IRSEI for the great environment and research community.

I would like to moreover acknowledge and thank all the co-authors that have contributed and supported me along the way.

Lastly, I would like to thank international colleagues and collaborators that have informally or formally supported me navigating research and further identifying new targets based on this thesis. I would thus like to thank Graciela Muniz-Terrera, for her guidance and connecting me to international collaborators. Further I would like to thank Heather Allore and Ana Quiñones for hosting, mentoring, and teaching me during a research stay, but more for how engaged they were in making my time in the U.S. an experience.

This work was supported by the European Research Council under the European Union's Horizon 2020 research and innovation programme [grant number 803239, to AKL].

Authors' Contributions

Throughout the thesis I will report results and implications of individual studies interchangeably using pronouns I/my or we/our, since all involved collaborative efforts. This paragraph aims to describe my role in and contribution to the individual studies in more detail. In all studies, I generated ideas and research goals thus conceptualising the studies, which included designing the methodology and conducting the formal analysis to confirm or falsify research questions. In that I independently led the presented studies as a first author and was responsible for validating procedures and findings. Besides conventional modelling during formal analysis, involving survival analysis, prediction tasks and mediation analysis, I enriched analytic strategies by applying methods based in the fields of machine learning and causal inference.

More specifically, for study 1 presented in Chapter II, I conducted multiple imputation of missing data using random forest models prior to formal analysis with survival models. For study 2 presented in Chapter III, I ran causal mediation analysis based on counterfactual imputation as well as an inverse regression-based approach to mediation analysis for individual or collections of microbes as potential mediators. For study 3 presented in Chapter IV, I compared a previously established algorithm to a broad range of machine learning functions, including extreme gradient boosting. This comparison entailed application of machine learning based algorithms to overcome class imbalance as well as the evaluation of various specifications in a train-test pipeline involving hyperparameter tuning and k-fold cross-validation. I moreover wrote initial drafts and organised feedback iterations. I led revisions and communicated with editors and reviewers. Throughout drafting, submission, and post-publication, I created visualisations and presentations to disseminate findings at scientific meetings and further prepared open access of analytical pipelines (https://github.com/makleelux/; Appendix I Contributions).

Summary

Dementia is recognised as a public health priority, affecting not only people living with the syndrome but also formal and informal carers as well as society at large. The growing number of people living with dementia worldwide strains economies increasingly, in part due to costs for disease management and care. In this thesis, key challenges of dementia prevention research were identified and targeted, with a special emphasis on the examination of modifiable social and behavioural risk factors.

Chapter I introduces dementia as a public health priority in ageing societies across the globe, considering current population trends and increasing knowledge on the role of modifiable social and behavioural risk factors. Chapter I furthermore provides an overview of operationalisation strategies and outlines a research framework underlying this thesis, focussing on dementia and cognitive function in later life as well as risk and protective factors. In that, the underlying aetiological theories, involved interactions of endogenous and environmental factors, as well as a life course perspective on dementia prevention are acknowledged. Chapter I closes with identified key challenges and derived research questions evolving around socioeconomic disparities in dementia risk, the lack of knowledge about underlying working mechanisms and the need for identification of individuals at-risk of dementia. Chapter II (study 1) draws on data from a large observational cohort, the UK Biobank, to examine area-level socioeconomic deprivation as a risk factor for dementia, while accounting for individual-level socioeconomic deprivation and genetic predisposition. Our findings suggest higher area-level socioeconomic deprivation to be associated with higher risk of dementia irrespective of individual-level socioeconomic deprivation, lifestyle, or genetic risk. Chapter III (study 2) reports an examination of features of the gut microbiome as potential mediators of the association of years of education with the risk of cognitive impairment in old age. Using data from the Luxembourg Parkinson's Study, findings

suggested that education is associated with gut microbiome composition and risk of mild cognitive impairment. However, there was no significant mediation. Chapter IV (study 3) presents the adaptation and validation of a dementia classification algorithm in the European, multi-country, context. Analysis was based on data from the Survey of Health, Ageing and Retirement in Europe. Results provide evidence for the usefulness of a dementia classification algorithm using a minimal predictor set, to help identify 'probable dementia' and to reduce cross-country variation in underreporting of dementia. Chapter V delivers a general discussion and synthesis of findings stemming from the individual studies. Contributions to the research field are further discussed, alluding to outlooks for future research.

In sum, reported findings suggest that both individual characteristics as well as contextual features, i.e., of the environment in which people live, work, and age are associated with the risk of cognitive impairment and dementia in later life. As such results of this thesis extend on previous research reinforcing the potential to dementia prevention by targeting individual, and contextual factors. Findings further reinstate the need for utilising deeply-phenotyped data to further elucidate working mechanisms underlying associations of modifiable risk factors and dementia. Lastly, this thesis provides a transportable solution to 'probable dementia' status classification in absence of clinical diagnosis in observational studies.

Note that Chapters II to IV have been reproduced from printed versions, with slight modifications to ensure consistency in formatting and terminology.

Correspondence may be addressed to matthias.klee@uni.lu.

List of Contents

List of Tables

List of Figures

Table of Abbreviations

Aβ	Amyloid Beta
AD	Dementia due to Alzheimer's Disease
ADI	Alzheimer's Disease International
ADL	Activities of Daily Living
ADRD	Alzheimer's Disease and Related Dementias
APA	American Psychiatric Association, DSM-5 Task Force
APOE	Apolipoprotein E Gene
ASV	Amplicon Sequence Variant
ATB	Antibiotic Medication Usage
AUC	Area Under the Receiver Operating Characteristic Curve
BDI-I	Beck Depression Inventory
BM	Brain Maintenance
BMI	Body Mass Index
BR	Brain Reserve
CDE	Controlled Direct Effect
CIND	Cognitive Impairment Without Dementia
CR	Cognitive Reserve
CSF	Cerebrospinal Fluid
DAA	Differential Abundance Analysis
DSM-5^{TM}	Diagnostic and Statistical Manual of Mental Disorders
EOD	Early Onset Dementia
FTD	Frontotemporal Dementia
GLM	Logistic Regression Model
HCP	Health Care Providers
HRS	Health and Retirement Study
IADL	Instrumental Activities of Daily Living
ICD-i	International Classification of Diseases ith Revision
LOD	Late Onset Dementia
LW	Dementia Classification Algorithm of Langa, Kabeto and Weir
MCI	Mild Cognitive Impairment
ML	Machine Learning
MMSE	Mini-Mental State Examination
MoCA	Montreal Cognitive Assessment
MRI	Magnetic Resonance Imaging
NCD	Neurocognitive Disorder
NDE	Natural Direct Effect
NIA-AA	National Institute on Aging – Alzheimer's Association
NIE	Natural Indirect Effect
OECD	Organisation for Economic Co-operation and Development
PAF	Population Attributable Fraction
PE	Proportion Eliminated

PET	Positron Emission Tomography
PS	Partnership Status
RCT	Randomised Controlled Trial
RF	Random Forest Classifier
SDoH	Social Determinants of Health
SES	Socioeconomic Status
SHARE	Survey of Health, Ageing and Retirement in Europe
SMOTE	Synthetic Minority Oversampling Technique
SNP	Single Nucleotide Polymorphism
VaD	Dementia due to Cerebrovascular Disease/Vascular Dementia
WHO	World Health Organization
WMH	White Matter Hyperintensities
XGB	XGBoost Classifier

Preamble

Chapter I – General Introduction

I.1 Motivation for the Dissertation

I.1.1 A Public Health Priority

Over a decade ago, the World Health Organization (WHO) recognised dementia as a public health priority (Alzheimer's Disease International [ADI] et al., 2012). "Dementia is a syndrome, usually of a chronic or progressive nature, caused by a variety of brain illnesses that affect memory, thinking, behaviour and ability to perform everyday activities" (ADI et al., 2012, p. 2).

The number of people living with dementia is growing, which is largely attributed to population ageing and growth worldwide (ADI et al., 2012; WHO, 2021). In 2019, dementia was the seventh leading cause of death (GBD 2019 Collaborators et al., 2021; WHO, 2020). Of note, mortality due to non-communicable diseases may vary across countries. Nonetheless, on a global scale, dementia reflects the leading cause of disability or dependency in old age, to date without an available cure (WHO, 2017, 2021).

As such, preventive efforts are key to respond to and address the growing prevalence of dementia, with a special emphasis on social and behavioural risk factors, given their theoretical modifiability. Before moving to previously established risk factors and defining risk and protective factors at the core of the thesis, the following sections will discuss the impact of dementia on an individual and population level and allude to relevant current trends.

I.1.2 Burden of Dementia

Consequences of the syndrome encompass cognitive impairment and resulting limitations in the ability of people living with dementia to successfully manage activities of daily living (ADL). This loss of autarchy denotes a varying need for support of dementia patients in different stages. Disease burden is evident in an elevated number of emergency

hospital admission in participants living with dementia (Sommerlad et al., 2019). Due to the progredient nature of the syndrome, likelihood of death and of admission to nursing homes increases, with only half of people residing at home 4 years after having retrieved a diagnosis (Mjørud et al., 2020). Of note, dementia is a deadly condition, which is not appreciated by research or reflected in care planning just yet, causing additional strain for those living with dementia, or those accompanying them (Harrison et al., 2019). Albeit the consequences of the syndrome centring around people living with dementia, the WHO acknowledges the scope of the burden also encompassing partners, families as well as carers (WHO, 2021). Of note, in many cases informal carers accompany people living with dementia over longer periods prior to nursing home admission. As a result, informal carers are at risk of psychological morbidity, social isolation and not least financial hardship (Brodaty & Donkin, 2009).

I.1.3 Current Trends and Developments

Disease burden on individual, societal, or economic level is not least driven by population growth and improved longevity and hence related, a growing number of people living with dementia (WHO, 2021). Previous findings suggest a three-fold increase from 2019 (57.4 million people living with dementia [95% CI, 50.4, 65.1]) to 2050 (152.8 million people living with dementia [95% CI, 130.8, 175.9]), globally (Nichols et al., 2022). However, future projections may vary, as they are based on analytical decisions regarding e.g., the inclusion of risk factors, and assumptions about the future, involving but not limited to e.g., the development of the prevalence of risk factors, or management/treatment options. This is especially important considering region-specific variation in secular trends, or the relative importance of risk factors and consequently, their population attributable fractions (PAF) for dementia (Mukadam et al., 2019; Nichols et al., 2022). As an example, estimating dementia prevalence projections when considering e.g., a reduced age-specific dementia incidence in high income countries, suggests up to 36% lower increase in prevalence

compared with projections assuming stable incidence rates in the Netherlands (Brück et al., 2022).

Regarding dementia treatment, potential to disease modification is acknowledged with traditional approval for lecanemab by the Food and Drug Administration in the United States, but regulatory review in Europe is pending at the time of writing. Therapies, e.g., lecanemab, apply to at-risk populations, e.g., with Mild Cognitive Impairment (MCI) and biomarker evidence of dementia due to Alzheimer's disease (AD), i.e., amyloid positivity (van Dyck et al., 2023). Appropriate use recommendations have been published recently (Cummings et al., 2023). Currently approved therapeutics need to be delivered early in the disease trajectory. Despite an increasing potential to defining at-risk populations prior to clinical impairment, risk prediction comes with uncertainty given a significant subpopulation with biomarker evidence (e.g., amyloid positivity) or biologically defined AD as per positron emission tomography (PET) scans but without cognitive impairment (Jack et al., 2013, 2019; Perez-Nievas et al., 2013). As such, a discussion about the ethics of disclosing biomarkers, i.e., to research participants without cognitive impairment, as well as the diagnostic utility of biomarkers has emerged (Alzheimer's Association Workgroup, 2023; Bunnik et al., 2022; Gómez-Isla & Frosch, 2019; Grill & Karlawish, 2022; The Lancet Neurology, 2024).

Given the increasing challenges imposed by a growing number of individuals living with dementia worldwide, the following section will briefly introduce aetiological theories relating to dementia and provide a definition of constructs at the core of this thesis.

I.2 Alzheimer's Disease and Related Dementias

Dementia is a syndrome caused by different underlying diseases of the brain and comprises impairment of multiple higher order cognitive functions, such as memory, executive functions, attention, language, amongst others. Dementia is a progredient and ultimately terminal condition. Aside cognitive impairment, neurobehavioural changes are

common (WHO, 2022b). Subtypes of dementia may further be distinguished clinically regarding the underlying disease. In that, AD, dementia due to cerebrovascular disease (here referred to as Vascular dementia; VaD), or Lewy Body disease, and Frontotemporal dementia (FTD) are most frequent, aside existence of dementia in other specified diseases, such as Parkinson Disease (WHO, 2022b). AD may occur with an early onset (EOD) or late onset (LOD) of symptoms characterised by age below or above 65 (WHO, 2022b).

I.2.1 Aetiology

Dementia is inherently irreversible and progredient in nature. However, due to the variety in underlying diseases causing dementia, unification of stages of the syndrome is hampered. Since the overarching number of dementia cases is diagnosed with AD or VaD, this section will briefly introduce current theory and evidence on the pathophysiology and disease progression for these subtypes.

Alzheimer's Disease Dementia

AD is the most common form of dementia and is caused by Alzheimer's disease. The most prominent aetiological theory of AD is the amyloid cascade hypothesis (Hardy & Higgins, 1992). In that, protein accumulation, i.e., of amyloid β (Aβ) is followed by an increase in phosphorylation and secretion of the protein tau, ultimately leading to neurodegeneration and cognitive symptoms (Hardy & Higgins, 1992; Zetterberg & Bendlin, 2021). A wide array of evidence for the hypothesis suggests presence of extracellular amyloid plaques and intraneuronal neurofibrillary tangles, i.e., hyperphosphorylated and truncated tau, in people with AD.

The assumed cascade has been validated predominantly in rare familial AD cases (Zetterberg & Bendlin, 2021). Such familial cases are driven by genetic mutations in genes that for example encode Aβ turnover related proteins (Zetterberg & Bendlin, 2021). Critically, most AD cases occur sporadically in older age. Additionally, it is unclear to what

extent observed pathophysiological changes reflect ageing processes or unique features of the disease which translate into impaired cognition. As such, people may show lesions or neurodegeneration in signature regions of AD without dementia or cognitive impairment (Gómez-Isla & Frosch, 2022; Zetterberg & Bendlin, 2021; Y. Zhang et al., 2010). Moreover, common co-pathologies such as synucleinopathies, e.g., characterised by occurrence of α-synuclein in brain tissue, a major constituent of Lewy Bodies, suggest more complex, overlapping working mechanisms in neurodegenerative diseases such as AD, Lewy Body disease and PD (Zetterberg & Bendlin, 2021).

Nonetheless, recent drug trials have reinforced the established hypothesis, suggesting slowing of cognitive decline by amyloid clearance in people with incipient AD (van Dyck et al., 2023). Despite potential to disease modification, it is still unclear how the assumed cascade is started. Sporadic LOD and AD in old age, are likely due to a multi factorial process. The complex pathogenesis may involve genetic predisposition, but more critically also vascular changes, or exposure to modifiable risk factors that may contribute to resilience in presence of hallmark indicators of AD (Gómez-Isla & Frosch, 2022; Kivipelto et al., 2018; Livingston et al., 2020; Zetterberg & Bendlin, 2021).

Vascular Dementia

VaD may be characterised as dementia resulting from underlying cerebrovascular disease (WHO, 2022b). Interestingly, the first person acknowledged with AD, dating back to the beginning of the 20th century, was Auguste Deter. Her case presented vascular changes as key characteristics of a newly identified disease. Thus, long before the amyloid cascade hypothesis, neuropathological findings including microvascular injuries, endothelial proliferation, and neovascularisation suggested a vascular pathway leading to dementia (Alzheimer, 1907; Fierini, 2020). Dementia as a syndrome characterised by brain atrophy

was assumed to be caused by impaired blood supply following atherosclerosis and consequent hardening of arteries (Fierini, 2020).

In more recent approaches to define VaD, variation in underlying causes is acknowledged and as such, VaD may follow general vascular damage in the brain, i.e., also post-stroke (Bir et al., 2021). VaD thus reflects cognitive impairment as a consequence of vascular injury. Importantly, clinical representation such as affected cognitive domains and staging may vary depending on the type and localisation of underlying vascular changes (Bir et al., 2021).

I.2.2 Terminology in the Present Work

Given the wide range of diseases such as Alzheimer's disease, or cerebrovascular disease, that primarily or secondarily affect the brain and may underlie dementia, differential diagnosis is challenging for health care providers (HCP), especially in primary care settings. Critically, most cases occur in later life and older patients at risk for dementia may present themselves with multiple health conditions. Barriers further include but are not limited to time constraints in clinical practice, as well as awareness about early symptoms not reflecting a normal ageing process (Porsteinsson et al., 2021). Time and cost sensitive testing procedures need to be employed to differentiate diseases underlying a potential dementia diagnosis. These involve e.g., blood tests, genotyping, neurological examination, or imaging such as magnetic resonance imaging (MRI) or PET scans. Still, more accurate diagnosis requires neuropathological examination.

Differential diagnosis is required to appropriately adapt treatment strategies to the specific needs implicated by underlying aetiology. Critically, symptom representation (AD: e.g., impaired/semantic memory, abstract thinking; VaD: e.g., impaired executive functioning, processing speed) may differ within and between individuals with dementia and thus interfere with specificity of differential diagnoses (Fierini, 2020). Diagnostic procedures

are further complicated by high frequency of mixed pathologies in later life, e.g., comorbid diabetes or heart failure, with in part overlapping pathophysiology (Alzheimer's Association, 2023). Additionally, presence of multiple brain pathologies is common in people living with dementia, e.g., indicated by half of AD patients showing Lewy Body co-pathology (Robinson et al., 2018; Schneider et al., 2007). While the majority of dementia cases are specified as AD or VaD, with distinct underlying pathology, recent evidence suggests a continuum of neurodegeneration and vascular dysfunction, reflecting mixed dementia (Fierini, 2020; Flier & Scheltens, 2005). As such, hallmark indicators of AD, e.g., neurodegeneration, or amyloid deposition, appear intertwined with vascular disease burden and downstream damage (Gottesman et al., 2017; Robinson et al., 2018; Snyder et al., 2015; Stakos et al., 2020). Recent clinical evidence showed that people living with similar dementia severity show less AD pathology when presenting with versus without vascular damage (Fierini, 2020; Zekry et al., 2002). Potential working mechanisms of such a mixed dementia phenotype involve altered Aβ production following vascular insufficiency, and in turn, loss of vascular homeostasis (Fierini, 2020). In sum, isolated, prototypical manifestations of subtypes are likely rare, compared with mixed forms of dementia. Consequently, research on risk factors targeting prevention of dementia has commonly focused on all-cause dementia or Alzheimer's disease and related dementias (ADRD) as umbrella concept, encompassing different subtypes, to capture the total contribution to and preventive potential for dementia as a clinical syndrome.

Although most ADRD cases occur in older age, evidence on prevalence of EOD is scarce and previous findings suggest likely underestimation of EOD prevalence due to, e.g., reliance on registry data (Hendriks et al., 2021). With a likely time-lag between symptom onset and diagnosis, as well as a steep increase in EOD incidence in later life, differentiation of EOD and LOD is to some extent arbitrary and researchers have argued to instead expand

on young onset dementia, e.g., prior to age 60/65, and 45 (Prince et al., 2015; van de Veen et al., 2021). There has been no consensus reached about aetiological differences between EOD, LOD or young onset dementia. Increased risk of EOD is likely conferred by genetic predisposition (autosomal dominant AD), albeit recent findings suggest associations with modifiable risk factors in line with those identified for LOD (Hendriks et al., 2024). The clinical presentation of EOD, although similar to LOD, is characterised by more rapid progression of cognitive impairment (WHO, 2022b). Young onset dementia is likely due to rare conditions, such as developmental disorders, epilepsy, psychiatric disorders, Korsakoff's syndrome, or traumatic brain injury (van de Veen et al., 2021).

Given the exponential growth in EOD and LOD above age 60, the absence of biological reasoning for established age cutoffs, scarcity of clinically validated subtyping and synergies in underlying risk factors, the term dementia will be used interchangeably with ADRD in this thesis. ADRD thus refers to dementia as a syndrome manifesting in later life, i.e., above age 60, or 65, most likely due to AD, VaD, FTD or Lewy Body disease (Prince et al., 2015; van de Veen et al., 2021). Following this terminology, dementia and ADRD reflect commonly investigated umbrella variables, without further specification of the subtype or the disease causing the syndrome.

I.3 Operationalisation of Dementia

The three individual studies at the core of this thesis respond to research questions relating to ADRD. The following sections will introduce relevant outcomes and their operationalisation in different settings and allude to implications for case ascertainment and subsequent formal analysis. In that, outcome definitions may vary with respect to dementia progression over time.

I.3.1 Dementia in the Clinical Context

Aetiological models informing the diagnostic process regarding AD and other dementia subtypes are increasingly defined as probabilistic models of disease progression. This is in part due to insufficient knowledge to formulate deterministic models for different subtypes (Fierini, 2020; Frisoni et al., 2022). However, as an introductory example for dementia progression, the following section aims to deliver a blueprint of AD staging, including a prodromal stage, a preceding preclinical stage, MCI, or mild behavioural impairment and subsequently, major neurocognitive disorder.

Prodromal Stage

Research on AD biomarkers suggest pathophysiological alterations, i.e., identified in cerebrospinal fluid (CSF) and blood as early as midlife, decades before the onset of clinical symptoms as (Zetterberg & Bendlin, 2021). A growing body of evidence supports AD-related decreases in the 42 amino acid-long Aβ42 and subsequent increases in total and phosphorylated tau in CSF and plasma as early as midlife (Zetterberg & Bendlin, 2021). Further biomarker candidates involve (ordered by time of positive testing) biomarkers of Aβ pathology (amyloid PET), neuroinflammation (sTREM2), synaptic dysfunction (CSF neurogranin), tau pathology (tau PET) and neurodegeneration (e.g., CSF neurofilament light, or hippocampal volume), respectively (Weston et al., 2019; Zetterberg & Bendlin, 2021).

The diagnostic process currently tests presence of clinical symptoms and relating limitations in everyday activities, as well as biomarker presence, determined e.g., with amyloid PET scans. Due to high costs, limited infrastructure and invasiveness of PET scans, uptake of fluid biomarkers in clinical practice may follow, soon, provided positive evaluation of clinical trials (Zetterberg & Bendlin, 2021). However, biomarkers, such as neurofilament light may increase, e.g., indicating neuroaxonal degeneration, over a decade before onset of clinical symptoms and eventual cognitive impairment (Weston et al., 2019; Zetterberg &

Bendlin, 2021). As such, biomarker-based diagnosis would reflect a biological definition of e.g., AD in absence of clinical symptom complexes. Given complexities in the definition of cutoffs for biomarker burden, limited causal evidence, likely small effect sizes, heterogeneity among asymptomatic individuals and potential resilience of some individuals to neuropathology, application in primary care settings, e.g., for initial screening and monitoring in clinical trials may follow as an first step (Gómez-Isla & Frosch, 2019; Jack et al., 2024; Zetterberg & Bendlin, 2021).

Preclinical Stage: Mild Cognitive Impairment

MCI is characterised by early and subtle changes in memory and thinking which however do not affect a person's ability to engage in activities required to master their daily living (Alzheimer's Association, 2022; Petersen et al., 2018). MCI may be classified either as amnestic, or nonamnestic depending on symptom representation. As such, impaired memory reflects amnestic MCI whereas impaired executive function, language, or visuospatial abilities reflect nonamnestic MCI (Petersen et al., 2018). MCI may be due to different underlying conditions, e.g., sleep deprivation, medication or anxiety, and entirely unrelated to dementia (Alzheimer's Association, 2022; Petersen et al., 2018). Accelerating impairment in cognitive function, however, is a hallmark indicator of dementia, and especially amnestic MCI may indicate preclinical AD. As such, MCI prevalence increases with age. The relative risk of dementia is 3.3-fold, and of AD 3-fold higher in people living with MCI compared to age-matched controls (Petersen et al., 2018).

While up to 55.6% of people with MCI revert to unimpaired cognition, MCI prompts continued evaluation of cognitive performance, and further thorough diagnostic testing, i.e. functional impairment, when reported to HCPs (Petersen et al., 2018). Of note, efforts in predicting progression from MCI to dementia have proven challenging. Previous findings suggest limited additional value of complex modelling strategies involving, e.g., imaging

data, over established cognitive tests (Ansart et al., 2021). This may be due to testing algorithms in short-term prediction settings. Previous findings suggest, negligible performance gains of prediction models based on imaging data in terms of accuracy compared with a constant prediction, i.e., assuming stable MCI for all observations (Ansart et al., 2021). Other reasons for current limitations in predicting conversion from MCI to dementia may be increased selection of stable subjects into observational cohorts, or failure of included biomarkers to account for resilience to neurodegeneration (Oosterhuis et al., 2023). Moreover, diversity in aetiological processes underlying dementia (e.g., in MCI due to AD, vascular MCI, or MCI related to Lewy Body pathology) may not be reflected in MCI diagnoses that are used to train algorithms and hence limit the validation. However, the evolving research landscape around biomarkers may offer opportunities to further improve subtyping or prediction of conversion from MCI to dementia (Franciotti et al., 2023).

Major Neurocognitive Disorder

Both International Classification of Diseases 11th Revision (ICD-11) and Diagnostic and Statistical Manual of Mental Disorders (DSM-5™) categorise dementia as neurocognitive disorder (NCD) or subtype thereof (American Psychiatric Association, DSM-5 Task Force [APA], 2013; WHO, 2022b).

Following ICD-11, a clinical diagnosis of dementia requires impairment in at least two cognitive domains, reflecting a lower than expected level of functioning given the respective individual's age and baseline performance, i.e., diagnosis requires a marked deterioration (WHO, 2022b). Following DSM-5™, a clinical diagnosis of dementia, i.e., major NCD, requires significant decline in only one cognitive domain to capture severity of functional impairment (APA, 2013; First et al., 2021). Thus, severe memory impairment may be classified major NCD in DSM-5™ but reflect amnestic disorder (not dementia) in ICD-11 (First et al., 2021).

Of note, a prodromal or preclinical stage (i.e., MCI) would be categorised minor NCD following DSM-5™ or mild NCD following ICD-11 (APA, 2013; WHO, 2022b). For a dementia or major NCD diagnosis, cognitive impairment needs to interfere with the person's ability to master ADLs and must not be attributable to alternative explanations e.g., ageing, substance intoxication or withdrawal (WHO, 2022b).

I.3.2 Dementia in Research

Diagnostic procedures to determine the clinical status of an individual have direct and indirect implications for the patient's journey, e.g., the provision of targeted treatment. In research, a diagnosis of dementia and more specifically, of a subtype due to a specific underlying disease may however not be of equal importance, depending on the research question. Research questions may be concerned with, e.g., identification of risk factors, quantification of the strength of associations with dementia, identification of at-risk populations, clustering of risk factors, classification of disease status, or evolution of cognitive function over time, to name a few. The following sections will discuss two operationalisation strategies for dependent variables i.e., outcomes, or primary endpoints, relevant for research questions in this thesis, namely cognitive function and (incident) dementia.

Cognitive Function as Outcome

Measurement of cognitive function as dependent variable, relates to research questions concerned with e.g., classification of MCI or dementia status when no clinical diagnosis is available, or investigating subtle changes and longitudinal trajectories of cognitive function, i.e., cognitive decline (Gianattasio et al., 2019; Hunt et al., 2021).

Given heterogeneity of symptom representation across individuals and underlying diseases, screening tools or measures of cognitive function may vary between studies. As an example, in research about AD risk factors, memory may be operationalised with recall

performance as a potential measure (e.g., immediate/delayed word list recall). Comparability of cognitive function outcomes across studies focussing on different domains may thus be limited.

More comprehensively, cognitive function may be operationalised with validated cognitive assessments in the form of screening instruments or test batteries. Such instruments combine performance across multiple domains of cognitive function to derive an overall estimate used by HCPs to determine the presence or level of impairment. Prominent screening tools are e.g., the Montreal Cognitive Assessment (MoCA), and the Mini-Mental State Examination (MMSE), which may be used to test presence of MCI or AD respectively (Folstein et al., 1975; Nasreddine et al., 2005). Classification of disease status is conducted by comparison of scores to cutoffs, which are based on normative samples. However, such cutoffs may lead to differential misclassification when applied to populations that are underrepresented in the normative samples (Nasreddine et al., 2012; Rossetti et al., 2011).

Of note, major NCD or ADRD are clinically characterised by apparent limitations in everyday functioning as well as cognitive decline over time, which implicates longitudinal assessment of cognitive function (APA, 2013; WHO, 2022b). Short tests, such as word list recall may reflect a more feasible approach to assessing cognitive function longitudinally. This is in part due to a shorter duration of testing compared to larger instruments or batteries and hence easier implementation in longitudinal observational studies.

Besides, research targeting causal exposure-outcome relationships requires longitudinal data to ascertain temporality. Critically, risk factors may be associated with cognitive impairment inconsistently over time, and associations may indicate reverse causality (Brenowitz, 2021). The long prodromal or preclinical stage of ADRD, thus reinforces the need to measure domains of cognitive function and observe trajectories before clinical symptoms emerge (Kivimäki et al., 2018; Kivipelto et al., 2018). Instruments such as

the MMSE or MoCA are primarily developed to screen for clinical impairment in cognitive function (Folstein et al., 1975; Nasreddine et al., 2005). Critically, discrimination deteriorates at the upper end of the scales and resulting ceiling effects may hamper detection of subtle changes in cognitive function in a prodromal stage (Franco-Marina et al., 2010; Hoops et al., 2009).

From a clinical point of view, thresholds are necessary to classify a person's level of cognitive function and to reach a binary decision regarding subsequent access to health care. From a statistical point of view however, application of a cutoff by design leads to reduction of information compared to analysis of e.g., continuous subscales. This reduction has implications for subsequent strategies of data analysis and interpretation. As an example, a binary classification may be used longitudinally, e.g., in multi-state models, but does not allow to model cognitive function trajectories over time.

Although functional impairment is likely a consequence of impaired cognition, measuring cognitive function alone precludes generalisation about clinically meaningful cutoffs and consequences of underscoring those for an individual's everyday life. Critically, when interested in risk factors for cognitive impairment and dementia, operationalising cognitive function with continuous measures of e.g., recall performance, may reflect a more patient-centred target considering its direct relevance for everyday functioning and quality of life. In addition to that, measurement of different domains of cognition such as executive function, processing speed, or memory, may operationalise distinctive underlying aetiological processes more accurately and fine-grained than a binary diagnosis.

Dementia as Outcome

Dementia is a common outcome in research examining e.g., risk and protective factors over the life course (Geraets & Leist, 2023; Licher, Darweesh, et al., 2019). As such, dementia as an outcome may be operationalised as diagnosis of AD, or VaD, or subsume

subtypes under the umbrella category ADRD. Observational studies need to ascertain dementia e.g., through linkage of health records, consensus diagnosis, or by asking participants or accompanying proxy respondents directly if they were diagnosed by an HCP.

In case of linkage to health or death records, comparability across studies may be limited due to varying assessment protocols across sources of diagnostic records. Most dementia cases reflected in health records are likely diagnosed by general practitioners. As such, access to health care, knowledge about dementia, ageism, or differences in welfare regimes and infrastructure may affect the likelihood and timing of receiving a clinical diagnosis given presence of dementia pathology (Bond et al., 2005). Some studies, such as the UK Biobank study further include post-mortem case ascertainment (Sudlow et al., 2015). Of note, post-mortem diagnosis could also reflect informant interviews instead of neuropathological examination. Despite outlined sources of biases evolving around linkage to health records, findings suggest high validity for case ascertainment based on health and death records, e.g., in the UK Biobank (Wilkinson et al., 2019).

In some cases, legal obligations imposed by country-specific General Data Protection Regulations hamper uptake of linkage to health records. Researchers may then opt to conduct expert consensus diagnosis, or include common diagnostic instruments such as MoCA or MMSE (Folstein et al., 1975; Nasreddine et al., 2005). Of note, consensus diagnosis is time consuming (also increasing the barrier for participation) and cost intensive leading to limited uptake in large observational studies. Additionally, dementia diagnosis within a study protocol imposes ethical challenges. Testing may reveal prevalent, formerly unknown early signs or symptoms that may be associated with cognitive impairment and dementia. This would prompt disclosure and often further follow-up by HCPs, which is likely exceeding resources of researchers.

A third option to ascertain dementia status is to ask participants or related proxies directly, if they were ever given a dementia diagnosis by an HCP, as is for instance done in SHARE (Börsch-Supan et al., 2013). However, selective attrition and cognitive impairment may limit reliability of such strategies. Moreover, previous findings suggest limited consistency of self-reported health outcomes (Cigolle et al., 2018). Consequently, absence of clinically validated cognitive assessments and barriers to linkage and monitoring of health care records has led researchers to develop classification algorithms to identify participants with 'likely' or 'probable' dementia (Alzheimer's Association, 2010; Crimmins et al., 2011; Herzog & Wallace, 1997; Hurd et al., 2013; Q. Wu et al., 2013). Such classification algorithms may also exploit further information collected in representative cohort surveys, including modifiable risk factors, to improve accuracy of prediction. In that, machine learning (ML) offers a fruitful extension to traditional modelling approaches allowing amongst other things faster processing of large amounts of (unstructured) data as well as modelling of nonlinear associations and higher-order interactions (Leist et al., 2022).

In sum, using binary dementia status as outcome may help to improve our understanding of risk factors of (subtypes of) dementia, to model likelihood of progression to a clinically meaningful state or yield relevant information for HCPs supporting patients. Conversely, continuous measures of cognitive function may facilitate detection of early changes in cognitive function and more generally, observe inter- and intraindividual variability of cognitive function over time. Analytical decisions are thus directly linked to the intended focus of the research question at hand.

Given the wealth of operationalisation strategies relating to dementia, the following sections provide an introduction of research frameworks relevant for dementia prevention. This involves a brief overview of the nomenclature of prevention, a structural framework to characterise risk factors and their interactions, implications of a life course approach to

dementia prevention, and related, constructs underlying potential working mechanisms of modifiable risk factors.

I.4 Dementia Prevention

Although dementia prevalence increases dramatically with age, dementia in later life is not considered an inevitable consequence of ageing (ADI et al., 2012; Licher, Darweesh, et al., 2019). LOD, i.e., ADRD, likely results from complex gene-environment interactions, with multiple domains contributing to or protecting against neurodegeneration and vascular changes which then cause the onset of clinical symptoms and disease progression (Finch & Kulminski, 2019). In absence of a cure and given the current trends outlined in I.1.3, including but not limited to population ageing and growth, targeting modifiable risk and protective factors to foster prevention is paramount to reduce the burden of disease globally.

I.4.1 Conceptualisation of Prevention in ADRD Research

Traditionally, prevention can be classified into primary, secondary, and tertiary prevention. Primary prevention is delivered prior to any pathological manifestation, such as cortical thinning, or reduced hippocampal volume, thus aiming to prevent dementia altogether. Interventions for primary prevention may target lifestyle-related risk factors such as a sedentary lifestyle or increased blood sugar levels, in cognitively healthy at-risk individuals with e.g., high polygenic risk for dementia (Barbera et al., 2023; Kivipelto et al., 2013). Secondary prevention is concerned with stopping or slowing progression of pathological changes associated with dementia, e.g., with experimental antibodies against amyloid, so that the time to functional impairments interfering with daily activities is maximised (van Dyck et al., 2023). Target groups for secondary prevention interventions are living with asymptomatic ADRD in a prodromal or preclinical stage, e.g., Aβ positive individuals without pathological evidence, or with MCI due to AD (J. Lee et al., 2022). Tertiary prevention includes treatment and delay of disease progression after clinical

manifestation, such that partial independence of people living with dementia is maintained, and quality of life is optimised (J. Lee et al., 2022). Consequently, target groups for tertiary prevention may fulfil criteria for minor or major NCD, but the distinction to secondary prevention is less clear (J. Lee et al., 2022).

Early tertiary prevention trials suggested modest improvements of clinical symptoms but no modification of the underlying progression of pathology, e.g., for acetylcholinesterase inhibitors (J. Lee et al., 2022). On the contrary, a new generation of medication, monoclonal antibodies, aims at clearing β amyloid plaques, specifically leaning on the amyloid cascade hypothesis of AD. Recent trials suggest slower disease progression in terms of cognitive decline after 18 months by reduction of amyloid burden in the brain (Ackley et al., 2021; Avgerinos et al., 2021; Cummings, 2023; Pang et al., 2023; van Dyck et al., 2023). However, lacking longitudinal follow-up or validation in severe AD, and sporadically occurring adverse events (e.g., amyloid related imaging abnormalities), as well as the magnitude of estimated costs associated with screening and delivery of this new treatment group challenge their benefit for society at large at this point in time (Bradshaw et al., 2024; Jönsson et al., 2023; Villain et al., 2022; Wimo et al., 2023). Still, future trials may reveal characteristics of better responders, or observe more pronounced benefits of combination therapies with longer follow-up to meet criteria for clinical relevance levelling risk (Villain et al., 2022).

Furthermore, randomised controlled trials (RCT) investigating potential for prevention so far mainly focussed on single drugs targeting a specific step in the amyloid cascade, e.g., with monoclonal antibodies, or vascular risk factors, e.g., with metformin (J. Lee et al., 2022). Critically, alternative pathways (e.g., vascular) are likely contributing to cognitive impairment and risk of subsequent dementia. This is backed by observational research, which further suggests modifiable risk and protective factors to be involved in the pathogenesis of dementia (Ajnakina et al., 2020; Livingston et al., 2020; N. Zhang et al.,

2021). As an example, cardiometabolic multimorbidity at midlife is associated with increased dementia risk in later life, suggesting interventions to lower the risk of cardiometabolic diseases may reduce dementia risk (Dove & Xu, 2023; Jin et al., 2023). Critically, face validity of an association of lifestyle factors e.g., physical activity or nutrition with cardiometabolic conditions and those of cardiometabolic conditions with dementia is greater, and supported by aetiological theories, opposed to an association of lifestyle factors with dementia, for which causal evidence is frequently challenged (Ciria et al., 2023; Desai et al., 2023; Kuźma et al., 2018; Seblova et al., 2021). However, lowering the burden of individual cardiometabolic conditions or their accumulation by intervening on e.g., lifestyle, even in absence of direct causal links of lifestyle factors with the pathology underlying dementia, would likely improve the quality of life for people living with dementia, and may extend to prolonging the time to dementia onset or prevent cases altogether (Livingston et al., 2020). Against this background, an extension of existing theories of cognitive ageing including environmental, social, and behavioural factors is imperative to improve dementia prevention frameworks.

I.4.2 Alzheimer's Disease Exposome

Conceptual frameworks such as the AD Exposome allow the highly granular and at the same time holistic examination of cognitive ageing as a process involving genetic and environmental (i.e., complementary to genome, non-genetic) factors (Finch & Kulminski, 2019; Wild, 2012). The AD Exposome postulates exogenous macrolevel (e.g., air pollution), as well as exogenous individual (e.g., diet) and endogenous factors (e.g., gut microbiome) relevant to AD pathogenesis (Finch & Kulminski, 2019).

Evidence about risk and protective factors of dementia, especially exogeneous macrolevel and individual factors mainly stems from large-scale observational studies. This is in part due to ethical limitations evolving around intervention on modifiable risk factors

and/or social determinants of health (SDoH) as exposures of interest. SDoH reflect structural and societal components shaping opportunities to healthy aging over the life course, e.g., formal and informal sources of support reflected in access to healthcare (Marmot & Wilkinson, 2005). Since they are generally specified as superordinate, higher level factors, they resonate with external macrolevel factors, as per the AD Exposome (Finch & Kulminski, 2019).

SDoH-related health inequalities may be based on occupation, education, or income (Marmot & Wilkinson, 2005). As such, previous research suggests determinants of cognitive functioning such as education to be conferred by compulsory schooling, denoting education as an in part external macrolevel factor (Glymour et al., 2008; Leist et al., 2021; Schneeweis et al., 2014). As a result, alteration of e.g., education as SDoH with consequences for later life cognitive function may require policy intervention. Of note, despite the definition of SDoH as structural components, factors such as education may require ascertainment on an individual level, e.g., when investigating individual-level dementia risk.

The report of Livingston et al. (2020) suggested twelve potentially modifiable social and behavioural risk factors to account for up to 40% of dementia cases. Livingston et al. (2020) acknowledge variation in the degree of modifiability and underlying risk propagation mechanisms by categorising risk factors as early, mid, or late life risk factors. Identified risk factors include sociodemographic (less education), health (obesity, depression, diabetes, hypertension, hearing impairment, traumatic brain injury), lifestyle-related (smoking, physical inactivity, low social contact, alcohol consumption) as well as environmental (air pollution) indicators (Livingston et al., 2020).

Despite limited knowledge about the causality of identified factors, a large body of evidence suggests significant reduction of the burden of dementia by targeting e.g., lifestyle-related risk factors (Kivipelto et al., 2018; Licher, Ahmad, et al., 2019). Frequently discussed

working mechanisms linking risk factors to cognitive impairment or dementia, involve but are not limited to neurodegeneration and vascular damage, inflammation, or oxidative stress (Grande et al., 2021; Kivipelto et al., 2018).

Following the AD Exposome, the different domains are further assumed to interact in a synergistic manner (Finch & Kulminski, 2019). This is supported by observational studies, reporting interindividual differences in cognitive decline, even in high genetic risk groups, i.e., suggesting potential amplification of a genetic predisposition to dementia (e.g., in carriers of the ε4 allele of the apolipoprotein E gene [*APOE*; OMIM:107741]) by behavioural or environmental factors (Ajnakina et al., 2020; Lourida et al., 2019; Solomon et al., 2018; N. Zhang et al., 2021). Further, exogenous individual risk factors (i.e., sleep deprivation) may interact with microbial communities in the gastrointestinal tract conferring pathophysiological alterations in mouse models. As an example, reduced butyrate secretion may mediate downregulation of inflammatory response and neuronal apoptosis and consequent memory impairment via a gut-brain-axis (Finch & Kulminski, 2019; C. Wang et al., 2022; X. Wang et al., 2023). These examples reinforce the potential of interventions, not only by directly intervening on theoretically modifiable risk factors, but also on the biological, endogenous mechanisms conferring cognitive impairment, e.g., by providing butyrate (C. Wang et al., 2022).

Of note, estimation of the combined contribution of the twelve risk factors reported by Livingston et al. (2020) to the global number of people living with dementia may not necessarily reflect real world preventive potential, considering outstanding causal evidence and likely bounded efficacy of interventions targeting theoretically modifiable risk factors. However, even modest delays regarding onset of dementia would result in significant decreases in incidence and thus prevalence across the globe (Brück et al., 2022; Kivipelto et al., 2018; Licher, Darweesh, et al., 2019; Livingston et al., 2020).

While the AD Exposome categorises (notably not rigid but fluid) non-genetic factors regarding exogeneity (and further macrolevel versus individual) or endogeneity, it inherently acknowledges an exposure history across the life course, and as such, time-varying effects (Finch & Kulminski, 2019; Wild, 2012). Of note, VaD and other subtypes of dementia are not formally constituting the AD Exposome (Finch & Kulminski, 2019). However, synergies between subtypes are acknowledged as are shared risk factors, and the notion of an exposure history (Livingston et al., 2020). As an example, for time-dependent effects, carrying an *APOE* ε4 allele may reflect a genetic predisposition to and hence elevated risk of dementia, determined at birth. In the same group of individuals, vascular damage may occur due to lifestyle-related risk factors in midlife, such as limited physical activity. In turn, this could explain a potentially reinforced cognitive decline in later life compared to other individuals with a similar genetic predisposition. Consequently, adopting a life course risk model of dementia is imperative to deepen our understanding of the disease progression and fostering identification of individuals at risk to deliver tailored prevention interventions in-time.

I.4.3 Life Course Model of Risk and Protective Factors of Dementia

Adopting a life course perspective of dementia risk, mechanistic questions arise about how risk propagates over time (Kuh et al., 2003). Livingston et al. (2020) acknowledged time dependency of risk factors, distinguishing early life (e.g., less education), from midlife (e.g., obesity) and later life risk factors (e.g., air pollution). Of note, a thorough description of risk propagation with regard to accumulation, critical, or sensitive periods (e.g., in childhood) is beyond the scope of this thesis (e.g., Kuh et al., 2003). However, grouping in Livingston et al. (2020) resonates with critical, (i.e., air pollution would affect brain health only in later life) or sensitive periods (i.e., air pollution would affect brain health more profoundly in later life). As an example of a sensitive period in later life, air pollution may reinforce impaired cognition given pre-existing neurodegeneration, by increasing downstream vascular damage.

Alternatively, risk may accumulate over the life course (i.e., air pollution and resulting vascular damage may affect brain health and cognitive decline uniformly over time), challenging sensitive or critical period formulation. Presupposing accumulating exposure to air pollution over the life course, impaired cognition may manifest and be observed in later life only after a critical level of damage has emerged (Abolhasani et al., 2023; Livingston et al., 2020; Peters, Ee, et al., 2019).

While life course epidemiology may offer opportunities to formulate and examine how later life health depends on risk propagation from birth to later life, real-world limitations may interfere with testing life course models of risk aiming to inform public health efforts and prevention design. As an example, follow-up required to observe exposure to risk factors over a lifespan is excessive and likely not feasible in large-scale representative cohort studies. Consequently, investigators frequently need to rely on retrospective self-reports, studies with extensive, yet limited, follow-up periods (usually less than 20 years), or combine studies conducted across different periods in life, e.g., by mapping participants' health records to surveys conducted during their respective adolescence (Mullin et al., 2023; Zuber et al., 2023). While previous findings suggest validity of assessing e.g., retrospective anthropometry, through self-report, interindividual variability in reporting subjective measures raises concerns about the validity of such approaches to examine risk factor outcome relationships across the lifespan, especially for less salient exposures (De Rubeis et al., 2019; Gorber et al., 2007).

In order to investigate life course models of risk in spite of these limitations, theoretical or evidence-based assumptions about the flow of causality over time need to be formulated, at best prior to decisions about the strategy of data analysis, e.g., by application of directed acyclic graphs (Tennant et al., 2020). Previous findings suggest early physiological changes, indicated by biomarkers associated with amyloid deposition and

neurodegeneration, up to two decades prior to dementia diagnosis (Zetterberg & Bendlin, 2021). Similarly, accumulation of cardiometabolic multimorbidity is faster in preclinical dementia, already two decades prior to impaired cognition (Guo et al., 2024). Such findings deliver crucial insights for interpreting models examining risk propagation. As an example, higher Body Mass Index (BMI) in midlife is associated with higher dementia risk in later life (J. Li et al., 2021). Conversely, lower BMI in later life is associated with higher dementia risk in later life (J. Li et al., 2021). This goes to show that the same exposure may reflect a risk factor in midlife but capture reverse causation in later life. As such, later life BMI would be affected by early pathological alterations associated with dementia without yet clinical symptoms and thus diagnosis of dementia (Brenowitz, 2021; Brenowitz et al., 2021; Kivimäki et al., 2018; Kivipelto et al., 2018; J. Li et al., 2021). Similarly, cardiometabolic conditions relate to dementia risk inconsistently over time, with earlier onset denoting higher excess risk (Dove et al., 2023; van Gennip et al., 2024). Hence, interventions may only yield efficient reduction of excess dementia risk if delivered in midlife (targeting higher BMI, reduction of cardiometabolic burden) rather than later life (targeting lower BMI).

Recapitulating, synthesis of models describing normative human development as well as adaptation to exogenous risk and protective factors with an ageing perspective is imperative to improve prevention efforts across the lifespan (Kuh et al., 2003). In case of dementia, a growing body of evidence increasingly emphasises lifelong biological, but also psychological and social processes that may not only propagate risk, but also alleviate presence of co-occurring risk factors or increase resilience towards existing pathology (Kuh et al., 2003; Livingston et al., 2020). In line with assumed accumulation of risk, protective factors such as education may manifest in early life, and proceed to alter resilience or lifestyle choices over the life course (Lövdén et al., 2020).

In sum previous observations signal the need for early intervention, but also denote the methodological challenges arising from required follow-up periods, or interrelations of exposures over the life course. Given discussed findings, effective prevention models for dementia may require targeting individuals at risk prior to clinical manifestations associated with the syndrome (i.e., primary prevention), or prior to the development of risk factors (i.e., primordial prevention, Gillman, 2015). More generally, working mechanisms of risk propagation require further research.

I.4.4 Cognitive Reserve, Brain Reserve and Brain Maintenance

The biological underpinnings of dementia have been extensively researched. With hypothesised aetiologies such as the amyloid cascade underlying e.g., AD, or vascular pathways leading to cognitive impairment and dementia, prevention research often focussed on reducing neurodegeneration or preventing vascular damage (J. Lee et al., 2022).

As such, many of the potentially modifiable social and behavioural risk factors consolidated by Livingston et al. (2020) are assumed to alter risk of dementia by causing or relating to neurodegeneration and damage, e.g., amyloid/tau-mediated, vascular, or inflammatory (e.g., air pollution, obesity, diabetes, hypertension, smoking, traumatic brain injury). Such risk factors may be targeted for dementia prevention. However, excess dementia risk may result from accumulated exposure over longer periods of time before measurement, e.g., with obesity, or smoking, and thus windows of opportunity for prevention are of question. Still, a consequent approach to prevention may target the reduction or restoration of a disease-free state in individuals that are identified at-risk, given risk factors.

Other factors such as engaging in social contact or physical activity could increase or maintain cognitive function by protecting against cognitive decline. Such protective factors may thus denote potential for resilience and reflect targets for secondary prevention in mid- or later life (Livingston et al., 2020; Stern et al., 2020). Of note, the concept of resilience

likely reflects multiple underlying reserve-related processes (Stern et al., 2020). While working mechanisms are still under investigation, protective effects may be conceptualized with cognitive reserve (CR), brain reserve (BR) and brain maintenance (BM). In that, CR, BR, and BM allow inclusion of SDoH and individual-level modifiable social and behavioural risk factors into life course risk and resilience models of dementia, within the AD Exposome (Finch & Kulminski, 2019).

The concept of CR may be defined as adaptability of cognitive processes leading to maintained functioning and thus increased protection against the consequences of brain ageing, pathology, or insult (Stern et al., 2020). In addition to CR, the concept of BR has emerged characterising structural brain differences to explain interindividual variability in responses to brain ageing and neuropathology (e.g., number of neurons) at a specified point in time (Stern et al., 2020). Contrasting BR from CR, BR does not imply active adaptation of cognitive processes but is assumed to reflect a buffering quantity individuals may lose prior to experiencing impairment (Stern et al., 2020). Contrary to BR, BM is conceptualised longitudinally, in terms of a reduced development of brain changes and pathology (Stern et al., 2020).

Functional or cognitive brain processes constituting CR, are theoretically defined constructs opposed to BR reflecting neurobiological capital (Stern et al., 2020). As such, CR-related processes and interindividual differences in CR may be targeted with functional MRI (e.g., via mapping of resting state or task-related functional activation brain networks). As an example, given a cognitive task one could measure activation in functional brain networks to determine efficiency (minimum level of activation necessary for task completion) or capacity (maximum level of activation possible given increasing demands) to operationalise the neural implementation of CR (Stern et al., 2020). In contrast to CR, BR may be operationalised with structural MRI measures. Critically, measures of brain structures, e.g., cortical thickness, may

reflect a combination of (concurrent) BR and (longitudinal) BM (Stern et al., 2020). As such, longitudinal assessments are necessary to differentiate e.g., those with preferable premorbid cortical thickness (i.e., BR) from those experiencing less volume loss (i.e., BM). Further candidates for measurement of BR involve grey matter volume, or cortical surface area (Stern et al., 2020).

Alternatively, proxy measures such as education, or engagement in social activities may be used to operationalise CR. However, candidate CR-proxies may be associated with cognitive test performance but not reflect CR. Residual methods, i.e., regressing cognitive function on brain pathology associated with dementia such as amyloid burden or white matter hyperintensities (WMH) and using the residuals as measure of CR, have been proposed to overcome this issue (Elman et al., 2022). Similarly, BM may be operationalised with a residual approach regressing brain status on age (Stern et al., 2020). However, residual methods are under debate regarding limited validity. More precisely, given collinearity of residuals with the outcomes of interest, a more valid approach to operationalising CR or BM may be the investigation of effect modification by proxies, i.e., of the association of brain pathology and cognitive impairment (Elman et al., 2022).

CR-proxies may impact BM over time. As an example, education may influence lifestyle choices and subsequently the likelihood of experiencing stroke and subsequent vascular damage. In such a scenario, education would relate to BM but not necessarily BR (Stern et al., 2020). Factors such as hearing impairment could be interpreted as CR-proxy if you would consider treatment of hearing loss as means to maintain functioning despite an apparent damage (Livingston et al., 2020). Other factors such as depression, physical activity, or alcohol consumption may reflect CR, BR, or BM. As an example, by intervening you could limit damage or target CR, but current evidence about working mechanisms does not yield a clear classification (Livingston et al., 2020).

CR is not fixed but assumed to be modifiable by lifetime exposures (Stern et al., 2020). In an alternative framework, the concept of resilience is extended by incorporating factors imposing a detriment to CR. As such, a more holistic inclusion of risk factors such as depression is possible assuming factors contribute to neural resource enrichment or to its depletion (revised Scaffolding Theory of Aging and Cognition, Oosterhuis et al., 2023). Critically, CR-proxies are assumed to covary or contribute to CR, without necessarily reflecting a functional mechanism (Stern et al., 2020). In addition to that, reverse causation may drive associations of CR-proxies such as engagement in social activities or physical activity with dementia risk. Sophisticated methodologies involving Mendelian Randomisation have been applied to uncover the potential causal contribution of typical CR-proxies to dementia risk by introducing genetic instruments for exposures of interest. However, independence assumptions underlying Mendelian Randomisation may not hold when applied to factors constituting in part individual-level exposures but also SDoH such as education, and e.g., survivor bias may obscure findings (Desai et al., 2023; European Alzheimer's & Dementia Biobank Mendelian Randomization Collaboration et al., 2023; McMartin & Conley, 2020).

Despite advantages in measurement and examination of CR, BR, and BM, discussed challenges impose limitations for prevention research. Critically, linkage of CR and BR or BM on a biological level (i.e., level of molecules, cells, and systems) needs to be established to fully exploit prevention potential (Stern et al., 2020).

I.5 Key Challenges for Dementia Prevention

To conclude the theoretical background, the following sections will recapitulate the relevance of modifiable social and behavioural risk factors for dementia prevention and allude to three key challenges for dementia prevention research.

I.5.1 Updated Relevance for Dementia Prevention

Firstly, the lifetime risk of dementia is estimated to be 1 in 3 for women, and 1 in 5 for men (Licher, Darweesh, et al., 2019). While projections about future dementia prevalence by design reflect a reduction of real-world complexities, it is highly uncertain if past improvements (e.g., increasing levels of education) will continue to decrease dementia incidence, if a plateau will be reached, or if widespread worsening of health, e.g., as indicated by increasing AD risk in hypertension, or increases in prevalence of hypertension and type 2 diabetes, may counteract or even overturn potential reductions in dementia incidence (Adesuyan et al., 2023; Brück et al., 2022; Mills et al., 2020; Ong et al., 2023). Researchers have argued that improvements in e.g., lifestyle and healthcare may contribute to decreasing age-specific dementia incidence (Livingston et al., 2020). However, current trends, e.g., an increasing prevalence of early and midlife risk factors of dementia are not well represented in models projecting future dementia prevalence, and bias may emerge from period, or cohort effects (WHO, 2021). High costs for healthy nutrition and widespread availability of ultra-processed food may contribute to socioeconomically patterned increases in the prevalence of obesity and type 2 diabetes, and related dementia (Delpino et al., 2022; Henney et al., 2024; Marmot, 2020). Furthermore, the increasing transition of social interactions to online spaces, or a modern lifestyle including sedentary behaviour, e.g., in the workplace, may affect further risk factors such as loneliness or physical activity, with likely detrimental contributions to dementia risk (Guthold et al., 2018; Lei et al., 2024). Without clear evidence of declining age-specific rates of dementia, the number of people living with dementia will most likely continue to increase due to population ageing and growth (Bussel et al., 2017; Prince et al., 2016; Y.-T. Wu et al., 2016). Moreover, dementia remains a severe condition conferring an outrageous challenge for those affected by the syndrome and societies at large. Loss of autarchy and impaired cognition mark major limitations to participation of individuals in

everyday life, increasingly so given the complexity and pace of modern-day developments. As such, every individual living with dementia counts, and each year without impairment is precious. Detriments are multiplicated by unmet needs and inequalities in care, as well as psychosocial, financial and health consequences for the surrounding of those living with the syndrome, including informal carers, friends, and families (Bond et al., 2005; Lindeza et al., 2020).

Secondly, synergies among modifiable risk factors may be most efficiently addressed with personalised multi-domain interventions, which is also acknowledged in WHO guidelines for evidence-based dementia risk reduction (Barbera et al., 2023; Stephen et al., 2021; WHO, 2019). Of note, current evidence on modifiable risk factors for dementia is still based largely on cohort surveys with limited follow-up, and variation in operationalisations of risk factors or analytical strategies may contribute to reduced comparability across studies (Peters, Booth, et al., 2019). Further, causality of the twelve most established modifiable social and behavioural risk factors of dementia is under debate (e.g., Desai et al., 2023; Kuźma et al., 2018). However, combined evidence from previous studies suggest a dose-response relationship of modifiable risk factors with higher late life dementia risk and consequently potentials to prevention, aiming at a reduction of exposure (Peters, Booth, et al., 2019). Of note, targeting of prevention efforts may be conducted on a continuum of proximity towards the individual. As one example, the Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability suggested improvements in cognitive function and secondary outcomes such as lower risk of decline, particularly in carriers of an *APOE* ε4 allele, which denotes elevated risk of AD (Solomon et al., 2018). More generally, beneficial effects of multi-domain interventions for global cognition and cognitive domains relevant for functional abilities have been shown (e.g., executive functioning, processing speed), especially when focussing on at-risk populations (Andrieu et al., 2017; Espeland et al., 2017;

Hafdi et al., 2021; Kivipelto et al., 2018; Kulmala et al., 2019; Ngandu et al., 2015; Yaffe et al., 2024).

Thirdly, the emergence of disease modifying therapies is accompanied by discussion relating to lacking infrastructure necessary for screening of eligible patient subgroups and related cost, also for the provision of therapeutics (Jönsson et al., 2023; Wimo et al., 2023). As an example, amyloid positivity determined with PET scans reflects the most established biomarker for AD. Due to the short half-life of amyloid PET tracers and the scarcity of cyclotron-equipped production facilities, screening is hampered even in countries with widespread PET scanner infrastructure (Zetterberg & Bendlin, 2021). Moreover, despite the aim to ensure equitable access, costs of treatment will supposedly be high and may reinforce socioeconomic disparities in health care within countries but also globally (Bradshaw et al., 2024; The Lancet Neurology, 2024; Wimo et al., 2023). While recent developments in treatment may denote a new era, related challenges regarding equitable access, feasibility constraints and costs may exceed health care systems' budgets (Jönsson et al., 2023; Wimo et al., 2023).

In sum, more research on potentially modifiable risk factors is necessary to capture and unleash the full potential of prevention for society at large. Even in a best-case scenario, assuming the introduction of feasible screening of individuals at-risk, e.g., based on fluid biomarkers, improvements in overall population health and exposure to risk factors, the number of people living with dementia may increase due to secular trends (Brück et al., 2022). Given well-established importance of theoretically modifiable risk factors for dementia, related risk factors and overall health and constraints to disease modification by treatment, targeting modifiable risk factors reflects a fruitful entry point for dementia prevention and related research (Stephen et al., 2021; WHO, 2019).

I.5.2 Socioeconomic Disparities in Health

Economically patterned inequalities in dementia risk denote a sensitive gap in the research landscape so far and likely limit individuals to flourish to their full potential. Previous findings suggest stronger associations with dementia risk for SDoH, or modifiable risk factors such as socioeconomic status (SES), air pollution or education compared with e.g., genetic predisposition (Weiss et al., 2020). Moreover, risk factors captured with the LIfestyle for BRAin health score (LIBRA; considers health status [coronary heart disease, type-2 diabetes, hypercholesterolemia, hypertension, depression, obesity, renal disease]; lifestyle-related factors [smoking, physical inactivity, alcohol use, high cognitive activity, diet]) have been shown to partially mediate the association of SES with higher dementia risk (Deckers et al., 2019; Livingston et al., 2020). Given the notoriously complex interrelations of theoretically modifiable risk factors of dementia, individual contributions of candidate risk and protective factors to dementia incidence are not easily disentangled, which is however necessary to investigate prevention potential by targeting said factors. Critically, exposure histories and the potential accumulation of detrimental effects of e.g., a less favourable SES, to cognition cannot be randomly allocated or manipulated in RCTs and require further attention in statistical analyses. In a life course model of risk and resilience for dementia, the importance of baseline genetic predisposition to dementia, as well as its potential interaction with exogenous risk factors further complicates observational research (Solomon et al., 2018). Against this background, it is crucial to consider exposures holistically and to identify at-risk populations to efficiently reduce apparent socioeconomic disparities in health.

In a first step, levers for prevention need to be identified. As alluded to, SES may exert its effect on dementia risk through differential exposure to cardiovascular and related lifestyle factors (Deckers et al., 2019; Geraets & Leist, 2023). As a multidomain intervention, involving diet, physical activity, cognitive training, and vascular risk monitoring, previous

findings reinforce the potential to reduce socioeconomic disparities in dementia risk by targeting modifiable risk and protective factors that promote cardiovascular health (Kivipelto et al., 2018; Livingston et al., 2020; Ngandu et al., 2015; Qiu & Fratiglioni, 2015).

SES may be linked to selection into areas or neighbourhoods with fewer socioeconomic resources as well. Importantly, higher area-level deprivation has been linked to AD neuropathology, previously (Powell et al., 2020). To identify potentials to intervention in well-defined target groups, individual-level characteristics thus need to be examined simultaneously with area-level indicators, thereby dissecting contributions of environmental exposures over individual-level exposures. This imposes challenges for researchers largely due to the availability of longitudinal measures of exogenous macrolevel and individual modifiable social and behavioural risk factors, accompanied by genetic assessments and clinical follow-ups.

I.5.3 Examining Biological Working Mechanisms

Education, as a constituent of SES is mainly acquired in early life. In later life, lower education reflects a risk factor of cognitive decline and dementia (Livingston et al., 2020). Previous findings suggest higher education to exert its protective effect primarily through increasing levels of cognitive abilities in early life, rather than affecting cognitive decline (Lövdén et al., 2020). Of note, formal education and related educational attainment is frequently assessed as e.g., highest degree obtained, or total years of schooling. Suggested resilience mechanisms may not generalise to different operationalisations of education, e.g., capturing socioeconomic position (Chapko et al., 2018; Lövdén et al., 2020; UNESCO Institute for Statistics, 2012). While I will proceed to examine education in this thesis, I adhere to a definition of education as formal education, i.e., years of schooling accumulated in early life. As such, education is assessed on an individual level and participants vary in

educational attainment, despite partial dependence on e.g., policy. I will thus further discuss education as exogenous individual risk factor.

While a comprehensive overview of the current landscape of resilience research and related methodological challenges is beyond the scope of this thesis, physiological working mechanisms or neurological correlates, relating e.g., education to BR continue to be investigated (Bocancea et al., 2021). Higher educated people living with dementia are on average older at the time of diagnosis and show a higher pathological burden linked to dementia, suggesting higher BR (Chapko et al., 2018; Lövdén et al., 2020). As such, a recent study identified brain connectivity (i.e., node degree operationalised as average number of connections of the structural connectome) as potential neurological correlate of resilience in the brain, supporting the notion of BR (DeJong et al., 2023). With implied critical or sensitive periods, it is unclear whether intervention on education in later life would alter risk of cognitive impairment. Previous research further suggests that education may relate to dementia risk indirectly. Mediation could follow through occupational status or complexity, lifestyle-related exposures, or reflect education as indicator of prolonged periods of extensive cognitive stimulation (Lövdén et al., 2020; Oosterhuis et al., 2023). Alternative pathways have been suggested by previously identified associations of SES and related, education, with gut microbiome alterations, that are in turn linked to later life health (Bowyer et al., 2019; Ding & Schloss, 2014; G. E. Miller et al., 2016). More specifically, the gut microbiome, i.e., the collection of microbes residing in the gut may relate to cognitive impairment via the gut-brain-axis (Morais et al., 2021; Saji, Niida, et al., 2019). Importantly, such pathways point to potentially modifiable intervention targets in later life.

I.5.4 Limited Access to Diagnosis and Risk Prediction

Risk prediction is crucial to identify and target individuals benefiting most from interventions and to tackle socioeconomic inequalities translating into dementia risk (WHO,

2022a). Moreover, risk profiling may allow improved recruitment strategies for cohort studies and drug trials. Despite advancement in large-scale data collection, predictive algorithms, identification of biomarkers, and harmonisation of large-scale survey designs, individual risk prediction remains challenging. This is in part due to inherent feasibility constraints of representative cohort surveys.

In general, clinically valid dementia ascertainment in longitudinal studies would require resource intensive testing procedures. One potential solution to this caveat is the application of dementia classification algorithms, determining 'probable dementia' based on readily available cognition or function measures. However, cross validation of such algorithms is scarce, and different study setups may lead differential performance of classification algorithms. To advance the availability, reliability, and stability of dementia ascertainment across data silos, research on the generalisability of dementia classification algorithms is needed, given restrictions to data collection.

I.6 Aims of the Thesis

This cumulative thesis responds to key challenges in the field of dementia prevention regarding social and behavioural risk factors, potential mechanisms conferring cognitive impairment, and risk prediction to classify 'probable dementia' status. This section will formulate the research questions evolving around identified challenges of the field.

To recapitulate, the number of people living with dementia will likely proceed to grow despite potential overestimation in previous prevalence projections (Brück et al., 2022; Nichols et al., 2022). Moreover, recent findings suggest availability of biomarker-guided, early diagnosis and disease modifying treatments in a foreseeable future, albeit current discussion regarding feasibility constraints and levelling of a risk/benefit ratio for the treated (Ackley et al., 2021; Avgerinos et al., 2021; Bradshaw et al., 2024; Cummings, 2023; Pang et al., 2023; van Dyck et al., 2023; Zetterberg & Bendlin, 2021). Besides, contributions of

modifiable social and behavioural risk factors to dementia risk may be underestimated due to not sufficiently accounting for synergies, communalities and clustering (Deckers et al., 2019; Livingston et al., 2020; Welberry et al., 2023). Against this background, prevention remains the most important aim in the race against dementia, and a focus on modifiable risk factors is paramount to lower the burden of dementia globally.

Firstly, the AD Exposome postulates interactions of exogenous and endogenous factors contributing to dementia risk (Finch & Kulminski, 2019). As such, exogenous macrolevel factors, such as neighbourhood deprivation opposed to exogenous individual factors such as wealth may interact with endogenous predisposition to develop dementia (Ajnakina et al., 2020). Identification of such interactions, contrasting exogenous macrolevel and individual domains, would allow to assess potential for prevention and in turn more accurate risk prediction and subsequent provision of targeted interventions. Additionally, explorative observations may point to mechanisms leading to increased dementia risk. This is investigated in paper 1. *How are socioeconomic indicators of the environment in which people reside associated with individual dementia risk given potential interaction with endogenous risk factors?*

Secondly, prevention efforts depend in part on knowledge about working mechanisms through which identified risk factors are associated with dementia risk. As an example, education is assumed to relate to BR, and thus resilience to neuropathology, associated with cognitive impairment (Lövdén et al., 2020). However, previous research mainly focussed on formal education accumulated in early life. Examining potentially modifiable biological mediators of a likely protective effect of education regarding cognitive health may offer new directions for the development of effective prevention strategies, extending to later life. One such avenue is reflected in previous findings concerning gut microbiome alterations relating to SES and later life health (Bowyer et al., 2019; Ding & Schloss, 2014; G. E. Miller et al.,

2016). This is investigated in paper 2. *What are potential working mechanisms linking education with the risk of cognitive impairment in later life?*

Thirdly, dementia case ascertainment is hampered in large-scale observational studies, in turn limiting statistical power and analysis of dementia as an outcome. Previous findings suggest validity of classifying dementia status with algorithms using minimal predictor sets based on available cognitive test records (Gianattasio et al., 2019). Given previously established classification algorithms, the question arises if applicability to cross-national contexts is given (Crimmins et al., 2011). Critically, dementia is a global burden and generalisability of findings is crucial. This is investigated in paper 3. *How can information collected in representative cohort surveys be used to classify 'probable dementia' status given cross-national differences and absence of clinically validated cognitive assessments?*

The following chapters, cover three individual studies (Chapters II to IV) responding to the three research questions. They are followed by a synthesis of findings and a discussion of implications for the field as well as future directions for research.

Publications

Chapter II – Identification of Risk Factors: Socioeconomic Deprivation, Genetic Risk, and Incident Dementia

Klee, M., Leist, A. K., Veldsman, M., Ranson, J. M., & Llewellyn, D. J. (2023). Socioeconomic Deprivation, Genetic Risk, and Incident Dementia. *American Journal of Preventive Medicine*, *64*(5), 621–630. https://doi.org/10.1016/j.amepre.2023.01.012

Abstract

Socioeconomic factors and genetic predisposition are established risk factors for dementia. It remains unclear whether associations of socioeconomic deprivation with dementia incidence are modified by genetic risk.

Participants in the UK Biobank age 60 years and older, of European ancestry without dementia at baseline (2006-2010) were eligible to the analysis, with main exposures area-level deprivation based on the Townsend Deprivation Index, individual-level socioeconomic deprivation based on car and home ownership, housing type and income, and polygenic risk of dementia. Dementia was ascertained in hospital and death records. Analysis was conducted in 2021.

In this cohort study, *N*=196 368 participants (*M* [*SD*] age=64.1 [2.9] years, 52.7% female) were followed-up for 1 545 316 person-years (median [interquartile range] follow-up=8.0 [7.4 to 8.6] years). In high genetic risk and high area-level deprivation 1.71% (95% CI, 1.44%, 2.01%) developed dementia compared to 0.56% (95% CI, 0.48%, 0.65%) in low genetic risk and low-to-moderate area-level deprivation (hazard ratio=2.31 [95% CI, 1.84, 2.91]). In high genetic risk and high individual-level deprivation 1.78% (95% CI, 1.50%, 2.09%) developed dementia compared to 0.31% (95% CI, 0.20%, 0.45%) in low genetic risk and low individual-level deprivation (hazard ratio=4.06 [95% CI, 2.63, 6.26]). There was no significant interaction between genetic risk and area-level (p=.77) or individual-level (p=.07) deprivation. An imaging substudy including 11 083 participants found greater burden of white matter hyperintensities associated with higher socioeconomic deprivation.

Individual-level and area-level socioeconomic deprivation were associated with increased dementia risk. Dementia prevention interventions may be particularly effective if targeted to households and areas with fewer socioeconomic resources, regardless of genetic vulnerability.

II.1 Introduction

The risk of AD and other subtypes of dementia is determined by multiple pathways including genetic, environmental and lifestyle factors (Livingston et al., 2020). Most cases occur in older adults and risk is linked to multiple common genetic variants, with PAF for single nucleotide polymorphisms (SNPs) of up to 8% or 27.3% for *APOE* ε4 allele (Lambert et al., 2013). Many studies have therefore employed polygenic risk scores (PRS) to quantify genetic risk of dementia suggesting almost two times higher incidence in high versus low polygenic risk (Ajnakina et al., 2020; Jansen et al., 2019; Lambert et al., 2013; Licher, Ahmad, et al., 2019; Lourida et al., 2019).

Moreover, individuals with fewer socioeconomic resources are at higher risk of dementia (Ajnakina et al., 2020; Cadar et al., 2018; Powell et al., 2020). Socioeconomic deprivation has been measured before, using both individual-level indicators such as income or wealth, and area-level indices like the Townsend Deprivation Index that captures unemployment rates, car and home ownership, and household overcrowding (Ajnakina et al., 2020; Cadar et al., 2018; Lourida et al., 2019; Townsend, 1987). Despite lower PAFs of risk factors related to socioeconomic deprivation (air pollution, 2.3%; education, 7.1%) recent findings suggest higher importance of wealth-related compared to genetic risk factors (Livingston et al., 2020; Weiss et al., 2020). While low area-level socioeconomic deprivation has been linked to CR and lower rates of cognitive decline in some studies, others found area-level deprivation no longer significant after adjustment for individual-level wealth (Cadar et al., 2018; Clarke et al., 2012, 2015). This suggests previous studies captured potentially distinct drivers of associations such as access to green space or air pollution, which are yet to be fully understood (de Keijzer et al., 2020; Peters, Ee, et al., 2019).

Cross-sectional findings link higher area-level deprivation to AD neuropathology (Powell et al., 2020). Further, a recent study found higher socioeconomic deprivation,

amongst others, associated with higher brain age (de Lange et al., 2021). Longitudinally, cognitive decline and accelerated degeneration in signature regions of AD including the medial temporal lobe were associated with higher area-level socioeconomic deprivation (Hunt et al., 2021). Additionally, links of WMH to a more rapid cognitive decline in MCI patients have been established before (Tosto et al., 2014).

No study has yet investigated the interplay of said factors jointly. Consequently, net associations of area-level above individual-level socioeconomic deprivation and their potential mechanisms have not been fully elucidated. While polygenic scores quantify a diathesis for dementia, it is yet to be examined if genetic predisposition may exacerbate associations of area-level and individual-level socioeconomic deprivation with incident dementia. Earlier research found interactions of polygenic risk with wealth and educational attainment and of *APOE* genotype with smoking (Ajnakina et al., 2020; N. Zhang et al., 2021). Findings show improved resilience to AD-related neurodegeneration, but there is also evidence suggesting more complex interactive pathways involving inflammation, which are not well understood yet (Ajnakina et al., 2020; Mole et al., 2020; Pan et al., 2020; Y. Zhang et al., 2021). Identifying potential interaction effects is crucial since they may point to risk factors and population groups that are most effective to target in dementia risk reduction interventions.

The purpose of this study was to use data from a large population-based cohort to investigate the hypothesis that associations between individual- and area-level socioeconomic deprivation and dementia may be modified by genetic risk. Complementary to previous research, the UK Biobank study offers unique opportunities. With over 500 000 participants analyses are well-powered to detect potentially small interactions. Additionally, information on genetics, imaging, area-level and individual-level socioeconomic deprivation is provided.

Lastly, linkage to health records and death registries allows extensive follow-up and dementia ascertainment (Wilkinson et al., 2019).

II.2 Methods

II.2.1 Study Sample

Data was provided by the UK Biobank, a population-based cohort study in the UK (Sudlow et al., 2015). Participants completed baseline assessments between 2006 and 2010 hosted in 22 centres (Sudlow et al., 2015). Of 502 536 participants, *N*=196 368 were eligible to analysis excluding participants below 60 (*n*=285 037), of other than European ancestry or without genetic data (*n*=20 969), with dementia at baseline (*n*=147) or discontinued consent before time of analysis (*n*=15). Follow-up continued until the date of first diagnosis, death, dropout, or last hospital admission. Participants without technical exclusion criteria (e.g., metal implants, discontinued consent, high movement) were reinvited for imaging between 2014 and 2020 (Littlejohns et al., 2020; S. Smith M. et al., 2020; Sudlow et al., 2015). A neuroimaging substudy included 11 083 eligible participants with imaging data.

II.2.2 Measures

Area-level socioeconomic deprivation was assessed with the Townsend Deprivation Index including information on employment, home ownership, car ownership and household overcrowding, based on baseline assessments and the preceding national census output areas (Townsend, 1987). Area-level socioeconomic deprivation categories distinguish low-to-moderate (quintiles 1-4) and high (quintile 5) deprivation, as exploratory analyses suggested no significant differences in the associations of quintiles 1-4 with dementia risk (Appendix II Figure S1).

Individual-level socioeconomic deprivation was based on a weighted composite score including home (own home without mortgage; other/not disclosed) and car ownership (one or more; none/not disclosed), housing type (house/flat; other/not disclosed) and annual

household income before tax (>£31 000; £18 000 to 30 999; <£18 000; not disclosed). For comparison, the median equivalised net household income in the UK in 2010/11 (end of baseline) was ~ 22 000£. The coefficients of a Cox proportional-hazards regression, with time to incident dementia as outcome, were used to compute individual-level socioeconomic deprivation (Appendix II Table S1). The score sums the product of indicators and their regression coefficient and divides it by the total sum of coefficients. Categories distinguish low (quintile 1), intermediate (quintiles 2 to 4) and high (quintile 5) individual-level socioeconomic deprivation. Previous research suggests systematic differences between participants disclosing socioeconomic indicators like income and those that do not (Kim et al., 2007). Therefore, and due to group sizes, not disclosed information was merged with less favourable categories except for income where it was kept as a separate category. A sensitivity analysis excluded participants who did not disclose socioeconomic information yielding similar results (Appendix II Table S1).

The PRS quantifies AD and dementia risk (Lourida et al., 2019). Polygenic risk was operationalised as the z-standardised weighted sum of the number of prevalent alleles at each AD-related SNP, including *APOE* genotype. Weights are based on their association with AD determined in a meta-analysis of genome-wide association studies of individuals of European ancestry (Lambert et al., 2013). Therefore, analyses were restricted to participants of self-identified European ancestry (British, Irish, other white). In total 249 273 SNPs met the *p* value threshold for inclusion, i.e., $p<.50$ (Lourida et al., 2019). Polygenic risk groups distinguish low (quintile 1), intermediate (quintiles 2 to 4) and high (quintile 5) risk.

Participants' all-cause dementia status was derived from hospital inpatient data (England: Hospital Episode Statistics, Scotland: Scottish Morbidity Record, Wales: Patient Episode Database) and death records (England & Wales: National Health Service Digital, Scotland: Information and Statistics Division), coding ICD-9/10 denoted primary/secondary

dementia diagnosis or dementia-related cause of death (WHO, 1992). ICD codes are presented in the supplementary material of a previous publication (Lourida et al., 2019). Previous research suggests high validity of this protocol, balancing a positive predictive value of 84.5% with reasonable case ascertainment (Wilkinson et al., 2019).

The six imaging-derived phenotypes (WMH, whole brain, grey matter, white matter, left and right hippocampal volume), were generated by an image-processing pipeline developed and run on behalf of UK Biobank (Alfaro-Almagro et al., 2018; Brugulat-Serrat et al., 2020; Cox et al., 2019; Lyall et al., 2020; K. L. Miller et al., 2016; Nobis et al., 2019; S. Smith M. et al., 2020; Wardlaw et al., 2015).

All models were adjusted for baseline characteristics including age in years, education (high: College/university degree; medium: higher secondary; low: lower secondary; other: degrees not covered in response options/non-response), sex, marital status (living with husband/wife/partner; joint category other/not disclosed), ancestry (20 first principal components) and in-sample third-degree relatedness (UNESCO Institute for Statistics, 2012). Models including PRS were additionally adjusted for the number of alleles included during computation. Potential mediators presence of depressive symptoms in the last 2 weeks and a healthy-lifestyle score (favourable; intermediate; unfavourable) were included in the main analysis (Lourida et al., 2019; VanderWeele, 2011).

II.2.3 Statistical Analysis

Missing data were assumed missing at random and addressed using multiple imputation by chained equations with five imputations (Donders et al., 2006). The imputation procedure employed recursive partitioning which is beneficial in the presence of nonlinear relations (Doove et al., 2014). Dementia incidence, survival times, variables relating to genetic risk or imaging, age, sex, and housing type were complete in eligible participants (Appendix II Figure S2).

Cox proportional-hazards regressions were applied to investigate the relationship of individual-level and area-level socioeconomic deprivation with time to incident all-cause dementia. Time at risk of dementia was modelled from baseline until diagnosis, loss to follow-up, death, or end of hospital admissions (England: 31/03/2017, Wales: 29/02/2016, Scotland: 31/10/2016). Main exposures were introduced stepwise to confirm main associations. Interaction terms between socioeconomic deprivation and polygenic risk were tested to investigate moderation. The assumption of proportional hazards was confirmed using Schoenfeld residuals, i.e., p=.71 in first imputed data set (Schoenfeld, 1982).

For the main analysis, socioeconomic deprivation categories were combined with polygenic risk groups, with low genetic risk and lower socioeconomic deprivation as reference categories, to investigate variation in the associations of socioeconomic deprivation with dementia incidence for different levels of genetic risk. Absolute risk was calculated as percentage of cases based on the first imputed data set. Incidence rates per 1,000 person-years were calculated accordingly.

For the exploratory imaging substudy, potential imaging-related confounders were entered in the multivariable linear regressions as predictors for imaging-derived phenotypes, including site-specific derivatives capturing (squared) age, sex, age-sex interactions, head size, (squared) days since scanner start-up and two dummy variables coding site. In a second step, scaled residuals were used as dependent variables in multivariable linear regressions including main exposures, covariates and inverse probability weights based on logistic regression models with selection into the imaging subsample as dependent variable (Alfaro-Almagro et al., 2021; Bradley & Nichols, 2022; Cole & Hernán, 2008). WMH burden was log-transformed.

Sensitivity analyses comprised replication in complete-case data and subsamples stratified by polygenic risk and sex. For the imaging substudy, a less conservative set of

potential imaging-related confounders including age, sex, age-sex interactions, head size and site was applied (Alfaro-Almagro et al., 2021).

Results were pooled across five imputed data sets according to Rubin's rules (van Buuren & Groothuis-Oudshoorn, 2011). Significance was assessed two-sided with $p<.05$. Analyses were performed in R version 4.0.3 (R Core Team, 2022; Therneau, 2021; van Buuren & Groothuis-Oudshoorn, 2011). Analysis code is available on the GitHub page of the first author (https://github.com/makleelux).

The UK Biobank study received approval from the North West Multi-centre Research Ethics Committee (MREC), the National Information Governance Board for Health & Social Care (NIGB) and the Community Health Index Advisory Group (CHIAG). All participants signed informed consent at baseline.

II.3 Results

In total, *N*=196 368 (*M* [*SD*] age=64.1 [2.9] years) participants (52.7% female) were followed up for 1 545 316 person-years (median [interquartile range] follow-up=8.0 [7.4 to 8.6] years). During follow-up, *n*=1,769 participants developed dementia (Table 1). In complete-case data, median age at dementia diagnosis was 72.0 for low-to-moderate and 71.7 for high area-level socioeconomic deprivation, and 71.6 for low, 72.0 for intermediate and 71.7 for high individual-level socioeconomic deprivation.

Table 1 *Baseline Characteristics (Study 1)*

Characteristic	Total No. (%)	
	Incident Dementia (*n*=1,769)	No Dementia (*n*=194 599)
Age, Years, *M (SD)*	65.8 (2.7)	64.1 (2.8)
Sex		
Female	790 (44.7)	102 644 (52.8)
Male	979 (55.3)	91 955 (47.2)
Education[a,b]		
High	317 (17.9)	49 493 (25.4)
Medium	472 (26.7)	59 160 (30.4)
Low	255 (14.4)	30 939 (15.9)
Other[c]	725 (41.0)	55 007 (28.3)
Married or in a Relationship[b]	1,586 (89.7)	179 256 (92.1)
Depressive Symptoms in Last 2 Weeks[b]	411 (23.2)	32 942 (16.9)
Healthy Lifestyle[b,d]		
5 (Favourable)	251 (14.2)	39 022 (20.1)
2 to 4 (Intermediate)	1,049 (59.3)	116 772 (60.0)
1 (Unfavourable)	469 (26.5)	38 805 (19.9)
Individual-Level Socioeconomic Deprivation[b,d,e]		
1 (Low)	174 (9.8)	39 100 (20.1)
2 to 4 (Intermediate)	1,037 (58.6)	116 784 (60.0)
5 (High)	558 (31.6)	38 715 (19.9)
Area-Level Socioeconomic Deprivation[b,d,f]		
1 to 4 (Low-to-Moderate)	1,266 (71.6)	155 829 (80.1)
5 (High)	503 (28.4)	38 770 (19.9)
Genetic Risk Group[d,g]		
1 (Low)	247 (14.0)	39 027 (20.1)
2 to 4 (Intermediate)	1,038 (58.7)	116 783 (60.0)
5 (High Genetic)	484 (27.4)	38 789 (19.9)

Note. Percentages may not sum to 100 because of rounding. [a]Education was grouped based on the UNESCO ISCED 2011(UNESCO Institute for Statistics, 2012) classification system. [b]Missing values have been imputed. Reported values are averaged across 5 imputed datasets. [c]The response level other summarised options prefer not to answer and none of the above. [d]Categories based on continuous scores. Numbers indicate quintiles from lowest (one) to highest (five). [e]Individual-level socioeconomic deprivation summarises information on home and car ownership, housing type and income. [f]Area-level socioeconomic deprivation based on the Townsend deprivation index (Townsend, 1987). [g]Genetic Risk based on a polygenic risk score for dementia (Lourida et al., 2019).

Dementia risk was higher in participants living in areas with fewer socioeconomic resources. Of participants in high area-level socioeconomic deprivation, 1.28% developed dementia (95% CI, 1.17%, 1.40%; Appendix II Table S2) versus 0.81% (95% CI, 0.76%, 0.85%) in low-to-moderate area-level socioeconomic deprivation (adjusted hazard ratio=1.47 [95% CI, 1.32, 1.63]; Table 2). Inclusion of genetic risk resulted in an adjusted hazard ratio of 1.47 (95% CI, 1.32, 1.64), indicating that area-level socioeconomic deprivation is independent of genetic risk. Additional inclusion of individual-level socioeconomic deprivation resulted in an adjusted hazard ratio of 1.28 (95% CI, 1.14, 1.43), suggesting that the association between area-level socioeconomic deprivation and dementia risk is partially accounted for by individual-level socioeconomic deprivation.

Table 2 *Risk of Incident Dementia According to Area-Level Socioeconomic Deprivation*

	Model 1		Model 2[a]		Model 3[b]	
Area-Level Socioeconomic Deprivation[c]	Low-to-Moderate (*n*=157 095)	High (*n*=39 273)	Low-to-Moderate (*n*=157 095)	High (*n*=39 273)	Low-to-Moderate (*n*=157 095)	High (*n*=39 273)
No. of Dementia Cases / Person-Years[c]	1,266 / 1 240 516	503 / 304 799	1,266 / 1 240 516	503 / 304 799	1,266 / 1 240 516	503 / 304 799
HR (95% CI)	1 [Reference]	1.47 (1.32, 1.63)	1 [Reference]	1.47 (1.32, 1.64)	1 [Reference]	1.28 (1.14, 1.43)
p		<.001		<.001		<.001

Note. All Cox proportional-hazards regressions were adjusted for the 20 first principal components, third degree relatedness, age, sex, education, and marital status. HR=Hazard ratio. [a]Model 2 additionally included polygenic risk and the number of alleles used to compute the polygenic risk score. [b]Model 3 included adjustments of model 2 and individual-level socioeconomic deprivation. [c]Reported results are based on the first imputed data set.

Dementia risk also increased monotonically across individual-level socioeconomic deprivation categories. Of participants with high individual-level socioeconomic deprivation, 1.41% developed dementia (95% CI, 1.29%, 1.53%; Appendix II Table S3) versus 0.44% (95% CI, 0.38, 0.51%) with low individual-level socioeconomic deprivation (adjusted hazard ratio=2.57 [95% CI, 2.14, 3.08]; Table 3). Inclusion of genetic risk resulted in an adjusted hazard ratio of 2.57 (95% CI, 2.14, 3.09) for high individual-level socioeconomic deprivation, indicating that individual-level socioeconomic deprivation is independent of genetic risk. Additional inclusion of area-level socioeconomic deprivation resulted in an adjusted hazard ratio of 2.38 (95% CI, 1.98, 2.87) for high individual-level socioeconomic deprivation, suggesting that the association between individual-level socioeconomic deprivation and dementia risk is independent of area-level socioeconomic deprivation.

Table 3 *Risk of Incident Dementia According to Individual-Level Socioeconomic Deprivation*

Individual-Level Socioeconomic Deprivation[c]	Model 1			Model 2[a]			Model 3[b]		
	Low (*n*=39 274)	Intermediate (*n*=117 821)	High (*n*=39 273)	Low (*n*=39 274)	Intermediate (*n*=117 821)	High (*n*=39 273)	Low (*n*=39 274)	Intermediate (*n*=117 821)	High (*n*=39 273)
No. of Dementia Cases / Person-Years[c]	174 / 309 221	1,042 / 929 551	553 / 306 541	174 / 309 221	1,042 / 929 551	553 / 306 541	174 / 309 221	1,042 / 929 551	553 / 306 541
HR (95% CI)	1 [Reference]	1.63 (1.38, 1.93)	2.57 (2.14, 3.08)	1 [Reference]	1.63 (1.38, 1.93)	2.57 (2.14, 3.09)	1 [Reference]	1.62 (1.37, 1.92)	2.38 (1.98, 2.87)
p		<.001	<.001		<.001	<.001		<.001	<.001
p of Trend[d]			<.001			<.001			<.001

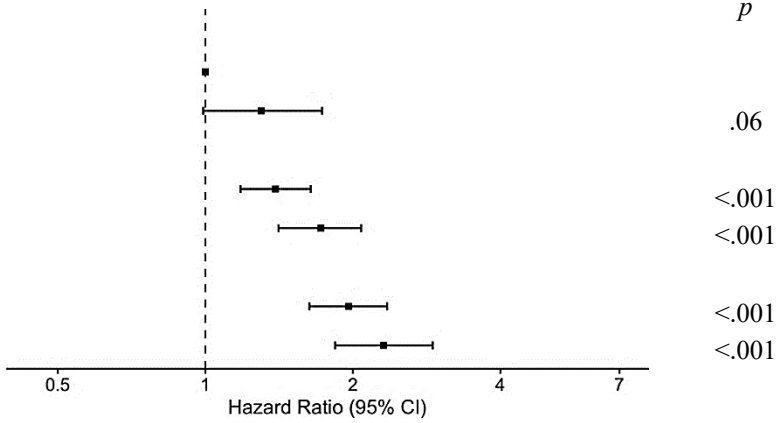
Note. All Cox proportional-hazards regressions were adjusted for the 20 first principal components, third degree relatedness, age, sex, education, and marital status. [a]Model 2 additionally included polygenic risk and the number of alleles used to compute the polygenic risk score. HR=Hazard ratio. [b]Model 3 included adjustments of model 2 and area-level socioeconomic deprivation. [c]Reported results are based on the first imputed data set. [d]*p* for trend was assessed using the continuous score of individual-level socioeconomic deprivation.

In models adjusted for socioeconomic deprivation, intermediate (adjusted hazard ratio=1.37 [95% CI, 1.19, 1.58]) and high (adjusted hazard ratio=1.91 [95% CI, 1.63, 2.23]) genetic risk were significantly associated with dementia risk. When genetic risk and socioeconomic deprivation categories were combined there was a consistent pattern of increasing dementia risk (Figure 1). Of participants with high genetic risk and high area-level socioeconomic deprivation, 1.71% (95% CI, 1.44%, 2.01%; Appendix II Table S4) developed dementia versus 0.56% (95% CI, 0.48%, 0.65%) with low genetic risk and low-to-moderate area-level socioeconomic deprivation (adjusted hazard ratio=2.31 [95% CI, 1.84, 2.91]). There was no significant interaction between area-level socioeconomic deprivation and genetic risk (p=.77; Appendix II Figure S3), indicating that the association with area-level socioeconomic deprivation did not vary substantially based on genetic risk. Of participants with high genetic risk and high individual-level socioeconomic deprivation, 1.78% (95% CI, 1.50%, 2.09%; Appendix II Table S5) developed dementia versus 0.31% (95% CI, 0.20%, 0.45%) with low genetic risk and low individual-level socioeconomic deprivation (adjusted hazard ratio=4.06 [95% CI, 2.63, 6.26]). There was no significant interaction between individual-level socioeconomic deprivation and genetic risk (p=.07; Appendix II Figure S4), indicating that the association with individual-level socioeconomic deprivation did not vary substantially based on genetic risk.

Figure 1 *Risk of Incident Dementia for A Area-Level and B Individual-Level Socioeconomic Deprivation With Genetic Risk.*

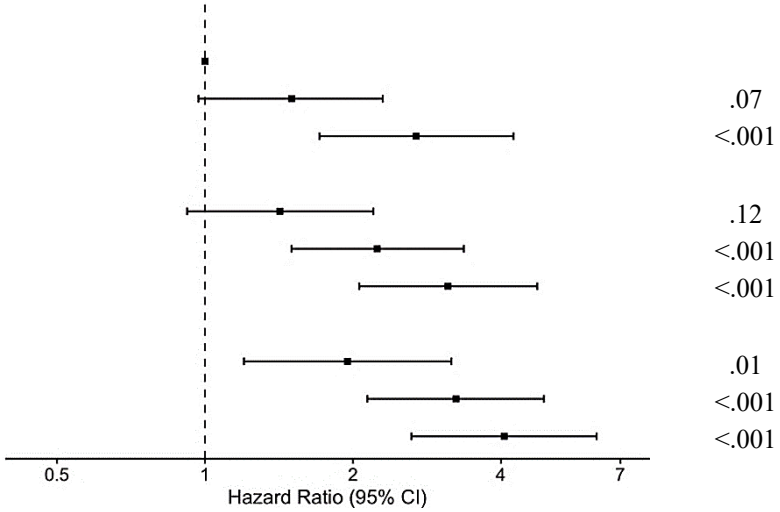
A *Area-level Socioeconomic Deprivation*

Deprivation	Total No.[a]	No. Dementia Cases / Person-Years[a]	*HR* (95% CI)	*p*
		Low Genetic Risk		
Low-to-Moderate	31 648	177 / 249 647	1 [Reference]	
High	7,626	70 / 59 124	1.30 (0.99, 1.73)	.06
		Intermediate Genetic Risk		
Low-to-Moderate	94 316	744 / 744 724	1.39 (1.18, 1.64)	<.001
High	23 505	294 / 182 389	1.72 (1.41, 2.08)	<.001
		High Genetic Risk		
Low-to-Moderate	31 131	345 / 246 144	1.96 (1.63, 2.35)	<.001
High	8,142	139 / 63 285	2.31 (1.84, 2.91)	<.001



B *Individual-level Socioeconomic Deprivation*

Deprivation	Total No.[a]	No. Dementia Cases / Person-Years[a]	*HR* (95% CI)	*p*
		Low Genetic Risk		
Low	8,110	25 / 63 790	1 [Reference]	
Intermediate	23 624	134 / 186 093	1.50 (0.97, 2.30)	.07
High	7,540	88 / 58 887	2.69 (1.71, 4.24)	<.001
		Intermediate Genetic Risk		
Low	23 417	103 / 184 307	1.42 (0.92, 2.20)	.12
Intermediate	70 774	614 / 558 529	2.24 (1.50, 3.36)	<.001
High	23 630	321 / 184 276	3.12 (2.06, 4.74)	<.001
		High Genetic Risk		
Low	7,747	46 / 61 124	1.95 (1.20, 3.17)	.01
Intermediate	23 423	294 / 184 928	3.24 (2.14, 4.89)	<.001
High	8,103	144 / 63 377	4.06 (2.63, 6.26)	<.001



Note. Coefficients correspond to combined groups of socioeconomic deprivation and genetic risk. All Cox proportional-hazards regressions were adjusted for the 20 first principal components, third-degree relatedness, number of alleles used to compute polygenic risk score, age, sex, education, marital status, healthy lifestyle, and depressive symptoms in the last 2 weeks. In addition, adjustments for (A) individual-level and (B) area-level socioeconomic deprivation were included. The number of dementia cases and dementia cases per person-years are based on the first imputed data set. HR=Hazard Ratio.

The imaging substudy comprised 11 083 participants (*M* [*SD*] age at imaging assessment=72.0 [3.2] years; 46.4% female) with available neuroimaging data. Total burden of WMH was higher in participants with high, compared with low-to-moderate, area-level socioeconomic deprivation (standardised coefficient=0.08 [95% CI, 0.01, 0.15]). Total burden of WMH was also higher in participants with high (standardised coefficient=0.10 [95% CI, 0.01, 0.19]) or intermediate (standardised coefficient=0.05 [95% CI, 0.00, 0.10]), compared with low, individual-level socioeconomic deprivation. In participants with high area-level socioeconomic deprivation, grey matter volume was lower (standardised coefficient=-0.11 [95% CI, -0.18, -0.04]). There were no significant associations with hippocampal, white matter or whole brain volumes (Appendix II Table S6 to Appendix II Table S11).

In complete-case data, participants with high genetic risk and high individual-level or area-level socioeconomic deprivation were at higher risk of dementia (Appendix II Table S12, Appendix II Table S13). Analysis in subsamples stratified by polygenic risk indicated that participants with high area-level or intermediate individual-level socioeconomic deprivation had higher risk of dementia in intermediate and high but not in low genetic risk (Appendix II Table S14, Appendix II Table S15). Participants with high individual-level socioeconomic deprivation had higher risk of dementia in all genetic risk groups. Analysis in subsamples stratified by sex yielded a similar pattern of results (Appendix II Table S16, Appendix II Table S17). Stroke may be on the causal path between socioeconomic deprivation and dementia and was therefore not included in the analyses; however, including history of stroke led to practically identical result patterns.

For the imaging substudy, result patterns were replicated, using a less conservative set of potential imaging-related confounders, except for a non-significant association of intermediate individual-level socioeconomic deprivation with WMH and a significant

association with grey matter volume. In complete-case data, the association of high individual-level and high area-level socioeconomic deprivation with WMH burden, and the association of high area-level socioeconomic deprivation with grey matter volume were not significant (Appendix II Table S6 to Appendix II Table S11).

II.4 Discussion

Individual-level and area-level socioeconomic deprivation were associated with risk of incident all-cause dementia, regardless of genetic risk. Participants with high genetic risk and area-level socioeconomic deprivation had a significantly higher risk of incident dementia compared with low genetic risk and low-to-moderate area-level socioeconomic deprivation, respectively. Similarly, participants with high genetic risk and individual-level socioeconomic deprivation had a significantly higher risk of incident dementia compared with low genetic risk and individual-level socioeconomic deprivation.

Previous studies had established that both area-level and individual-level socioeconomic deprivation were associated with an increased risk of dementia (Ajnakina et al., 2020; Cadar et al., 2018; Powell et al., 2020). Likewise, a prior meta-analysis of genome-wide association studies had established that a large proportion of the risk of developing late-onset AD is genetically determined (Lambert et al., 2013). Risk was highest in high socioeconomic deprivation and genetic risk. This finding is in line with a previous study, which however additionally found a significant interaction of lower wealth with polygenic risk of dementia accelerating the time to diagnosis, possibly owing to differences in genetic risk assessment and strategy of data analysis (Ajnakina et al., 2020; Leonenko et al., 2021). This study therefore extends prior findings by confirming pre-established associations and establishing that socioeconomic deprivation does not interact with genetic risk. In comparison the present study is considerably larger, incorporates a more comprehensive measure of

genetic risk, potential mediators, and tests moderation of socioeconomic deprivation more comprehensively.

Individual-level socioeconomic deprivation was more robustly associated with dementia risk in comparison with area-level socioeconomic deprivation. Components of individual-level socioeconomic deprivation such as low income may increase dementia risk through reduced access to healthcare, poor-quality nutrition and reduced cognitive stimulation that cannot be as effectively accounted for by area-level measures. Although there was a monotonic trend for individual-level socioeconomic deprivation, no such trend was found for area-level socioeconomic deprivation. This is in line with previous findings suggesting detrimental associations of neighbourhood socioeconomic deprivation with health outcomes at the highest levels (Hunt et al., 2021).

Contrary to earlier findings, area-level associations remained associated with increased dementia risk after adjusting for individual-level socioeconomic deprivation (Cadar et al., 2018). Area-level socioeconomic deprivation may capture dementia risk factors that are not fully explained by individual-level socioeconomic deprivation. Indeed, recent research suggests potential causal paths through cognitive stimulation at large, access to residential green space or air pollution (Clarke et al., 2012; de Keijzer et al., 2020; Peters, Ee, et al., 2019). As such, area-level socioeconomic deprivation may reflect environments with limited opportunities for cognitive stimulation, healthy nutrition, or physical exercise.

The imaging substudy explored measures of brain health that might underlie increased dementia risk associated with socioeconomic deprivation. Higher area-level and individual-level socioeconomic deprivation were associated with greater WMH burden. WMHs are a well-established indicator of cerebral small vessel disease, double the risk of dementia and are associated with more aggressive cognitive decline in MCI patients (Debette et al., 2019; Tosto et al., 2014). These results suggest a vascular pathway to dementia that might include

both individual-level vascular risk factors, such as blood pressure, and area-level risk factors such as air pollution. Importantly, these risk factors are modifiable (Livingston et al., 2020). High area-level socioeconomic deprivation was further associated with lower grey matter volume, suggesting additional, potentially neurodegenerative pathways. Although the results for hippocampal volume are inconclusive, lacking associations with other markers of brain health typically associated with dementia risk suggest global effects of area-level socioeconomic deprivation that might represent accelerated brain aging.

II.4.1 Limitations

Some limitations should be considered. First, individual-level and area-level socioeconomic deprivation are correlated (r=.33). Second, residual confounding may exist despite careful confounder adjustment. Third, reverse causation cannot be ruled out, despite a median follow-up of 8 years. Fourth, one in 71 participants above age 65 was ascertained with dementia compared to one in 14 in the general population, suggesting a healthy volunteer bias (Prince et al., 2014). Representativeness is further limited due to a low response rate to the invitation to participate in the UK Biobank study. Nonetheless, previous findings suggest health hazards correspond to findings in representative samples (Batty et al., 2020; Stamatakis et al., 2021; Sudlow et al., 2015). Fifth, without case finding, sensitivity cannot be tested, and dementia may have not been detected in all cases. Additionally, ascertainment in hospital and death records may select more severe cases, potentially biasing estimates (Wilkinson et al., 2019). Sixth, estimates may be biased since the competing risk of death can precede dementia diagnosis (Rojas-Saunero et al., 2021; Weuve et al., 2015). Seventh, analyses were restricted to age 60 years and older, limiting cases, and to European ancestry, limiting generalisability. Finally, some associations were not replicated in complete-case data, likely due to disproportionally missing data in higher socioeconomic deprivation (Appendix II Table S18).

II.4.2 Conclusions

In older adults without dementia, area-level and individual-level socioeconomic deprivation and genetic risk were significantly and independently associated with a higher risk of dementia. Dementia prevention interventions may be particularly effective if targeted to people living in households and areas with fewer socioeconomic resources, regardless of genetic vulnerability.

Chapter III – Examination of Mechanisms: Education and Mild Cognitive Impairment: The Role of the Gut Microbiome

Klee, M., Aho, V. T. E., May, P., Heintz-Buschart, A., Landoulsi, Z., Jónsdóttir, S. R., Pauly, C., Pavelka, L., Delacour, L., Kaysen, A., Krüger, R., Wilmes, P., Leist, A. K., NCER-PD Consortium. (2024). Education as Risk Factor of Mild Cognitive Impairment: The Link to the Gut Microbiome. *The Journal of Prevention of Alzheimer's Disease, 11, 759-768*. https://doi.org/10.14283/jpad.2024.19

Abstract

With differences apparent in the gut microbiome in MCI and dementia, and risk factors of dementia linked to alterations of the gut microbiome, the question remains if gut microbiome characteristics may mediate associations of education with MCI. We sought to examine potential mediation of the association of education and MCI by gut microbiome diversity or composition.

We analysed data from control participants of the Luxembourg Parkinson's Study, a cross-sectional study in Luxembourg and the Greater Region (surrounding areas in Belgium, France, Germany). Main measures were gut microbiome composition, ascertained with 16S rRNA gene amplicon sequencing, differential abundance, assessed across education groups (0-10, 11-16, 16+ years of education), and alpha diversity (Chao1, Shannon and inverse Simpson indices). Mediation analysis with effect decomposition was conducted with education as exposure, MCI as outcome and gut microbiome metrics as mediators.

After exclusion of participants below 50, or with missing data, *N*=258 participants (*n*=58 MCI) were included (*M* [*SD*] Age=64.6 [8.3] years). Higher education (16+ years) was associated with MCI (Odds ratio natural direct effect=0.35 [95% CI, 0.15, 0.81]. *Streptococcus* and *Lachnospiraceae*-UCG-001 genera were more abundant in higher education.

Education is associated with gut microbiome composition and MCI risk without clear evidence for mediation. However, our results suggest signatures of the gut microbiome that have been identified previously in AD and MCI to be reflected in lower education and suggest education as important covariate in microbiome studies.

III.1 Introduction

Modifiable social and behavioural risk factors of AD and related dementias convey potential to delay or prevent a substantial rate of cases, if targeted effectively (Livingston et al., 2020). This entails interventions to be delivered early in the disease trajectory, informed by knowledge on working mechanisms. With respect to timeliness, research on biomarkers suggests AD-related pre-clinical pathophysiological changes occurring as early as midlife (Zetterberg & Bendlin, 2021). At a later stage, MCI reflects early, subtle changes in thinking and memory (Alzheimer's Association, 2022). While potentially due to a variety of underlying diseases or disorders, MCI is a markedly strong risk factor for AD. Furthermore, synergies in risk factors of MCI and AD exist, e.g., related to education and consequentially lifestyle (Alzheimer's Association, 2022).

Education itself reflects a well-established early-life risk factor for AD. As such, higher education is associated with lower dementia risk in later life (Livingston et al., 2020). Lower dementia risk may result from education increasing cognitive abilities in early adulthood and consequent build-up of CR, BR or BM, protecting against neurodegeneration (Livingston et al., 2020; Lövdén et al., 2020). Moreover, risk factors such as obesity or smoking frequently cooccur, and vary in prevalence according to SES (Alzheimer's Association, 2022; Deckers et al., 2019). Higher exposure to lifestyle-related risk factors according to education, an indicator of SES, may further contribute to a vascular pathway linking education to dementia risk. To date, there is no consensus about working mechanisms. However, recent studies suggest education, which influences life histories and in part constitutes SES, to be associated with differences in microbial community types across multiple body sites, which may be in turn associated with MCI risk (Bowyer et al., 2019; Ding & Schloss, 2014; G. E. Miller et al., 2016; Saji, Murotani, et al., 2019).

The gut microbiome refers to a collection of microbes within the gastrointestinal tract. The gastrointestinal tract, reflecting the largest ecosystem of the human body, is composed of bacteria, archaea, eukaryotes, and other microbes. Composition of the gut microbiome, for instance differential abundance of specific taxa, is subject to interindividual variation, e.g., across the life-course or geographical locations (Greenhalgh et al., 2016). Factors affecting the microbiome over a lifetime are for instance linked to SES and early childhood conditions (e.g., mode of delivery, breast feeding) resulting in variation in consequent gut colonisation and microbiome maturation, which may in turn continue to affect microbiome composition in later life (Greenhalgh et al., 2016; Heck et al., 2006; S. A. Lee et al., 2015; Milcent & Zbiri, 2018; Wampach et al., 2018).

Gut microbiome alterations have been observed, e.g., associated with ageing, or health. As such, ageing-related changes may result from the ageing processes (changing hormonal levels), changing health conditions (associated use of medication) or age-related behavioural changes, e.g., dietary deficiency (Greenhalgh et al., 2016). Moreover, recent findings suggest a link of the gut microbiome to MCI or AD, and lifestyle-related risk factors such as diet or physical activity (Alkasir et al., 2017; Cabrera et al., 2021; Saji, Murotani, et al., 2019; Saji, Niida, et al., 2019). Potential working mechanisms along the gut-brain-axis likely involve complex pathways, e.g., triggering low-grade systemic inflammation by altering gut permeability or by synthesis of metabolites with neuroendocrine functions (Morais et al., 2021). Due to the likely involvement of specific molecules, the low resolution of marker gene-based microbiome analyses precludes further specification of molecular pathways.

The association of education to gut microbiome alterations and MCI risk motivate the investigation of the role of the gut microbiome in the relationship of education and MCI.

Thus, we sought to examine potential mediation of the association of education and MCI by the gut microbiome in the present study.

III.2 Methods

III.2.1 Study Sample

We analysed data of participants, specifically the control subjects, from the Luxembourg Parkinson's Study of the National Centre of Excellence in Research on Parkinson's disease, which received approval from the National Ethics Board (CNER Ref: 201407/13) and Data Protection Committee (CNPD Ref: 446/2017) and was conducted according to the Declaration of Helsinki (Hipp et al., 2018). Eligibility criteria for analysis were age above 50, absence of Parkinson's disease, celiac disease, and chronic inflammatory bowel disease, availability of stool samples and non-missing data. All participants provided written informed consent.

Participants collected stool samples at home and sent them to the Integrated Biobank of Luxembourg (Baldini et al., 2020). Sampling, processing, and sequencing of stool samples were done as previously described (Baldini et al., 2020; Wilmes et al., 2022). The 16S rRNA gene amplicon sequencing data was processed using the dadasnake workflow, a Snakemake pipeline to process amplicon sequencing data, based on DADA2 (Callahan et al., 2016; Mölder et al., 2021; Weißbecker et al., 2020). Amplification primers were removed using cutadapt, allowing 20% mismatches and no indels (Martin, 2011). Quality filtering, amplicon sequence variant (ASV) generation and chimera removal were performed in DADA2. Reads were truncated at positions with less than 10 Phred score quality, or at 240 bp. The quality filtering kept only sequences with a maximum expected error of 2 and 240 bp length. Downsampling was performed to 25 000 reads using seqtk (https://github.com/lh3/seqtk: RRID:SCR_018927) and samples with smaller library sizes were removed from the downstream analysis. ASVs were generated in pooled mode for the whole study using

DADA2 default parameters. For merging forward and reverse ASVs, a minimum overlap of 12 bp was required. Chimeric sequences were removed based on the consensus algorithm. Taxonomic classification was performed against SILVA v138 using the naïve Bayesian classifier implemented in mothur (Quast et al., 2013; Schloss et al., 2009). Clinical and 16S rRNA gene amplicon sequencing data are available on request from https://www.parkinson.lu/research-participation.

III.2.2 Measures

Clinical assessments were conducted by neurologists, neuropsychologists, or trained study nurses. MCI classification was based on the MoCA, a brief measure for assessing cognitive function (Nasreddine et al., 2005). MoCA scores below 26 led to MCI classification (Nasreddine et al., 2005).

Education was assessed in years. For analysis, years of education were grouped (0-10 [reference], 10-16, 16+ years of education) based on the ISCED classification scheme, group sizes, and differences in compulsory schooling duration in Luxembourg for participants of different age (Honig, M. S. & Bock, 2017; UNESCO Institute for Statistics, 2012).

Alpha diversity captures the diversity of the microbiome within individuals. Alpha diversity will be greater in individuals with a greater number of different taxa (=richness) and/or similar abundances of prevalent taxa (=evenness). Alpha diversity is subject to variation over the life-course and higher alpha diversity has been related to better health in older age (Cabrera et al., 2021; Greenhalgh et al., 2016). Three measures for alpha diversity were computed after rarefication: Chao1, Shannon and inverse Simpson (Appendix III Diversity Measures). Beta diversity reflects differences of the microbiome between individuals. In that, dissimilarity indices reflect pairwise distances between individuals based on taxa abundance. In a sample-by-sample distance matrix, a greater value in a given cell indicates a larger dissimilarity between two individuals. This information can be used to

compare similarity of variance and composition of the gut microbiome between groups of individuals. Two measures for beta diversity were computed: Bray-Curtis dissimilarity and Jaccard distance (Appendix III Diversity Measures).

Additional measures included sociodemographic indicators age, sex/gender, first language (French/Luxembourgish/German versus other), partnership status (PS; married/domestic partnership versus widowed/never married/divorced/separated), BMI, mild depressive symptoms based on the Beck Depression Inventory I (BDI-I; >9), use of antibiotic medication in the last 6 months (ATB; yes versus no), and *APOE* ε4 (at least one versus no ε4 allele) status (Beck et al., 1961).

III.2.3 Statistical Analysis

All analyses were performed in R version 4.2.0 (R Core Team, 2022). Analysis code is available online at https://github.com/makleelux/edu_biome_mci (Appendix III R Session Info Excerpt). Differences of descriptive characteristics in presence or absence of MCI were tested with Fisher's Exact Test for categorical and Student's t-Test for continuous characteristics. Differences in beta diversity were tested across education groups with *betadisper* [*vegan*] and *adonis2* [*vegan*] with 999 permutations. In short, *betadisper* compares average distances, i.e., the dispersion or homogeneity, across groups, while *adonis2* tests multivariate differences in microbiome compositions (Oksanen J. et al., 2022).

Differential abundance analysis (DAA) was conducted across education groups, adjusting for age, sex/gender, BMI, and ATB. DAA was repeated additionally adjusting for first language, PS, BDI-I, and *APOE*, as robustness check. Two commonly used functions (*ancombc* [*ANCOMBC*]*; DESeq* [*DESeq2*: RRID:SCR_000154]) were employed (H. Lin & Peddada, 2020; Love et al., 2014). Both methods identify differentially abundant taxa with estimates of statistical significance adjusted for false discovery rates (Appendix III

Description of Differential Abundance Analysis). For DAA, taxa with nonzero counts in less than 25% of samples were not tested.

Mediation analysis was specified with MCI as outcome and groups of education as exposure, adjusting for age, sex/gender, first language, PS, BDI-I, *APOE* and ATB. For alpha diversity as mediator, a regression-based, counterfactual approach to mediation was employed for which continuous mediator models (i.e., alpha diversity as outcome) and a logistic outcome model (i.e., MCI as outcome), were specified (Appendix III R Code Mediation Analysis With CMAverse; *cmest* [*CMAverse*]), including interaction terms for education and alpha diversity (Shi et al., 2021). Total effects of education on MCI were decomposed into a controlled direct effect (CDE) for alpha diversity fixed at the sample mean, a natural direct (NDE, Appendix III Direct and Indirect Effects), and a natural indirect (NIE, Appendix III Direct and Indirect Effects) effect (Valeri & VanderWeele, 2013; VanderWeele, 2014; VanderWeele & Vansteelandt, 2014). Proportion eliminated (PE) was calculated, indicating the proportion of the effect due to either mediation, interaction, or both, that would be eliminated by fixing the mediator to a specific level, i.e., the sample mean of the z-standardised alpha diversity measures (VanderWeele, 2014). As a sensitivity check, mediation analysis was repeated without interaction terms in the outcome model.

For beta diversity as mediator, a previously described inverse-regression-based approach to mediation was employed at genus level (Yue & Hu, 2022a, 2022b). In short, this approach specifies regressions for potentially mediating taxa at genus level on education, and MCI adjusted for education, in turn utilising resulting *p* values to test mediation. Two functions were used, allowing to estimate mediation by abundance of specific taxa or by the overall composition of the microbiome (Appendix III R Code Mediation Analysis With Ldm and PermanovaFL, *ldm* [*LDM*]; *permanovaFL* [*LDM*]), while controlling for false discovery rates (Hu & Satten, 2020; Yue & Hu, 2022a, 2022b). *Ldm* suggests mediation if education

affects the microbiome and consequentially the outcome. This can be tested globally (community contains any mediating taxa) and locally (mediation by specific taxa). *PermanovaFL* is a distance-based procedure, and suggests mediation if education affects some part of the community and some potentially different part of the community proceeds to affect MCI, thus being less conservative. For *ldm* an omnibus test was conducted combining analysis at three scales, i.e., relative abundance, arcsin-root transformed relative-abundance, presence-absence (Zhu et al., 2022). For *permanovaFL* individual and omnibus tests were conducted combining analysis at two scales, i.e., relative abundance, presence-absence (Zhu et al., 2022).

III.3 Results

From 524 participants without Parkinson's disease or Parkinsonism diagnosis, *N*=258 participants were eligible for analysis (*M* [*SD*] Age=64.6 [8.3] years) after exclusion of participants below age 50 (*n*=93), with celiac disease (*n*=6) or chronic inflammatory bowel disease (*n*=5), missing data (*n*=11) or without stool samples and microbiome data (*n*=149, and *n*=2 after pruning of samples with library size <10 000). Participants with MCI (*n*=58) were older, more likely male, had fewer years of education and a higher BMI (Table 4). A total of 1,150 taxa at seven taxonomic ranks were identified after trimming of ASVs occurring in <10% of samples and pruning of samples with library size <10 000.

Table 4 *Baseline Characteristics (Study 2)*

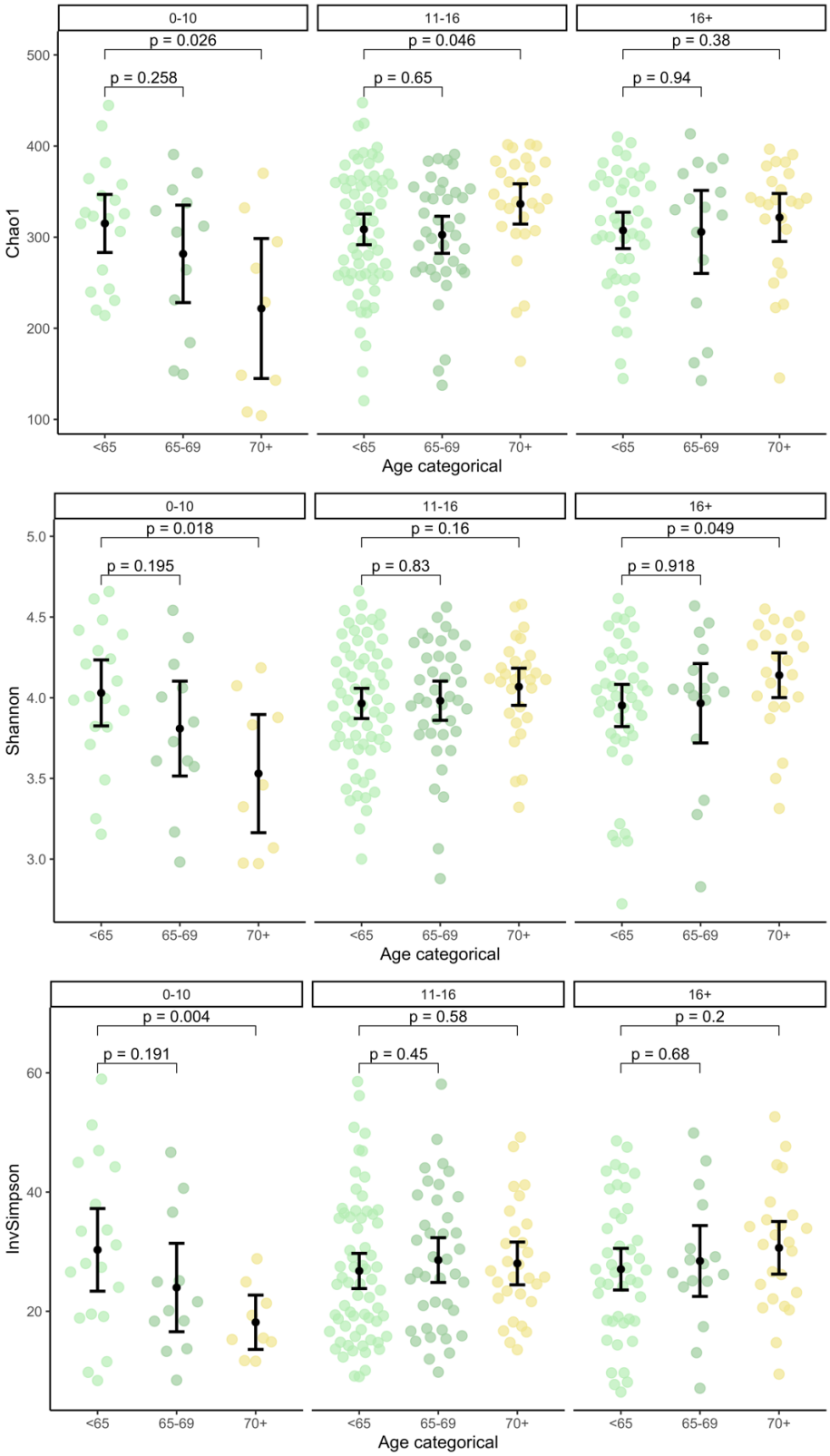
Characteristic	NC (*n*=200)	MCI (*n*=58)	*p*	Test
Age, Years	63.76 ± 7.84	67.6 ± 9	.005	*t*
Sex/Gender				
Female	87	15	.022	*Fisher*
Male	113	43		
Years of Education				
0-10	24	16		
11-16	110	24	.018	*Fisher*
16+	66	18		
First Language				
FR / LU / DE	182	52	.798	*Fisher*
Other	18	6		
Living With Partner				
No	61	17	1	*Fisher*
Yes	139	41		
BDI-I (>9)				
Yes	23	9		
No	177	49	.497	*Fisher*
APOE				
At least one ε4	54	19	.410	*Fisher*
No ε4	146	39		
Antibiotics (Last 6 Months)				
No	179	49	.351	*Fisher*
Yes	21	9		
BMI	27.04 ± 4.37	29.74 ± 5.91	.002	*t*
Alpha Diversity				
Chao1	311.56 ± 64.96	295.05 ± 86.37	.181	*t*
Shannon	3.99 ± 0.38	3.93 ± 0.46	.415	*t*
Inverse Simpson	27.75 ± 11.46	26.79 ± 11.66	.580	*t*

Note. Numbers refer to means ± standard deviations for continuous, n for categorical characteristics.

NC=Normal Cognition; MCI=Mild cognitive impairment; T=Student's t-Test; Fisher=Fisher's Exact Test; BMI=body mass index; BDI-I=Beck Depression Inventory I; *APOE*=Apolipoprotein E ε4 status; FR=French; LU=Luxembourgish; DE=German.

Alpha diversity as per Chao1 was lower in but not significantly associated with MCI. Education groups did not differ significantly in beta dispersion, tested with *anova* (p=.17), thus meeting the assumption of homogeneity of variances for *adonis2*. Education groups did not differ significantly regarding multivariate analyses with *adonis2* (p=.20 adjusting for sex/gender, age, ATB, BDI-I, first language, PS, and *APOE*), suggesting similar composition of the microbiome. However, alpha diversity was lower in lower education (Appendix III Figure S1) and was significantly lower in older age but only in lower education (Figure 2).

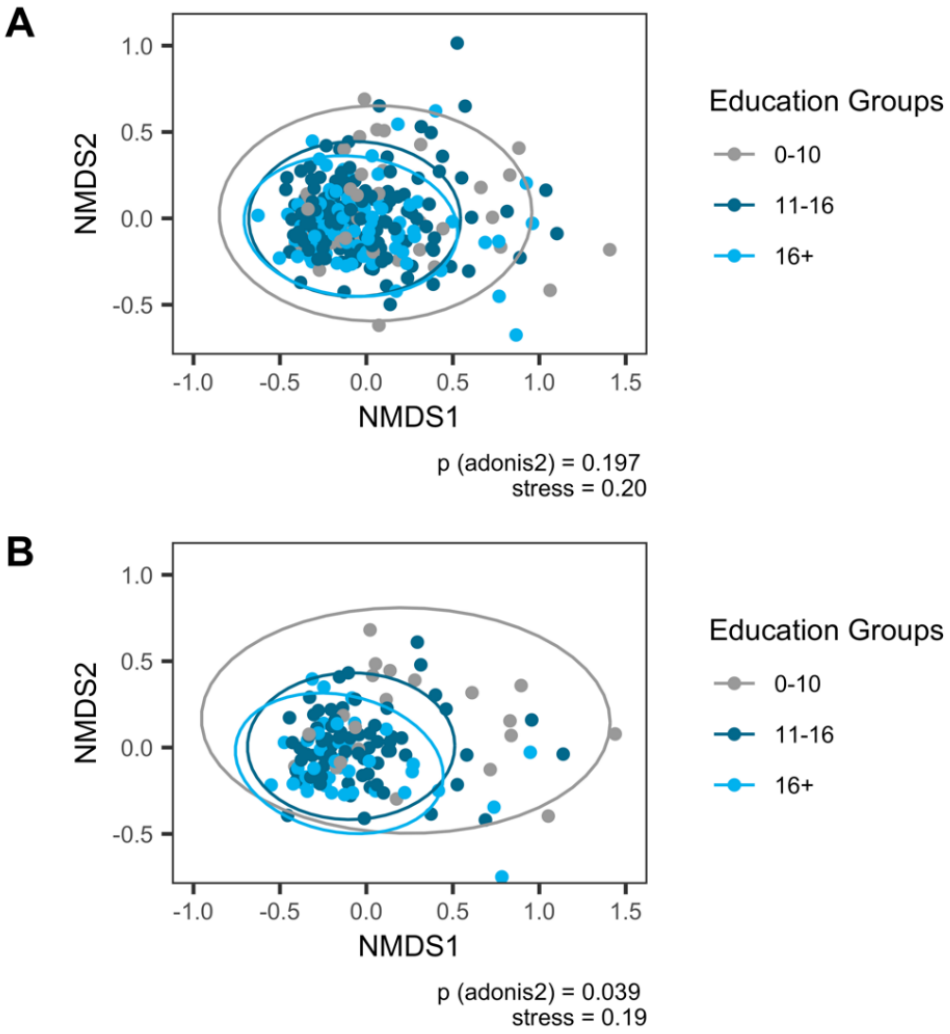
Figure 2 *Alpha Diversity Across Age and Education Groups*



Note. Chao1, Shannon and Inverse Simpson indices denoting alpha diversity stratified by age and education groups with 0-10, 11-16 and 16+ years of education. Reported *p* values result from Student's t-Tests with 0-10 years of education as reference group. InvSimpson=Inverse Simpson.

Beta diversity differed significantly across education groups (*betadisper*: p=.048; *adonis2*: p=.04; Figure 3), when restricting to age 65 and older.

Figure 3 *Ordination Plots for Education Groups*



Note. Ordination using NMDS based on Bray-Curtis dissimilarity for A the full sample and B a subset age 65 and older. Analysis with adonis2 was adjusted for sex/gender, age, use of antibiotic medication in the last 6 months, mild depressive symptoms based on the Beck Depression Inventory I, first language, partnership status, and apolipoprotein ε4 status. NMDS=Non-metric Multidimensional Scaling. Authors MK and VTEA.

There were no significant differences in beta diversity between MCI or age groups (Appendix III Figure S2). As Chao1 likely reflects an underestimate of richness with ASVs

based on DADA2, analyses were repeated with observed richness as measure of alpha diversity. These analyses yielded analogous findings (results not shown).

DAA suggested higher relative abundance of Bacilli (class), Actinobacteria (class), Lactobacillales (order), Streptococcaceae (family), *Streptococcus* (genus), with *DESeq2* and *Lachnospiraceae* UCG 001 (genus) and two ASVs with *ancombc* in higher compared to lower (0-10 years) education, adjusting for age, sex/gender, BMI, and ATB (Table 5).

Table 5 *Taxonomic Analysis Across Groups of Education*

Level	Taxon	DESeq2		ANCOMBC	
		11-16	16+	11-16	16+
Class	Bacilli[a,b,c]		**		
	Actinobacteria[a]		*		
Order	Lactobacillales[a,b,c]	**			
Family	Streptococcaceae[a,b,c]	*			
Genus	Streptococcus[a]	*			
	Lachnospiraceae UCG 001[a,b,c]				**
ASV	ASV 000053[a]				*
	ASV 000508[a]				*

Note. Significant differences in abundance across education groups 11 to 16 and 16+ years of education. Grey fill indicates higher relative abundance in higher education compared with lower education (0-10 years). [a]Adjusted for age, sex/gender, BMI, and use of antibiotic medication in the last 6 months. [b]Adjusted for age, sex/gender, BMI, use of antibiotic medication in the last 6 months, mild depressive symptoms based on the Beck Depression Inventory I, first language, and partnership status. [c]Adjusted for age, sex/gender, BMI, use of antibiotic medication in the last 6 months, mild depressive symptoms based on the Beck Depression Inventory I, first language, partnership status and Apolipoprotein E ε4 status. ASV=amplicon sequence variant; BMI=body mass index. *p<.05. **p<.01. ***p<.001 referring to lowest identified p values across individual tests.

Sensitivity analysis with additional adjustments for BDI-I, first language, PS, and/or *APOE* replicated findings, except for Actinobacteria and *Streptococcus* with *DESeq2*, and the two ASVs with *ancombc*. There was no overlap between *DESeq2* and *ancombc* (of note,

p_{adj}=.07 for *Lachnospiraceae* UCG 001 with *DESeq2*). Visual inspection of relative abundance plots suggests dose-response relationships of increasing years of education and increasing relative abundance *Lachnospiraceae* UCG 001 (Appendix III Figure S3).

With 0-10 years of education as reference, higher education was associated with higher Chao1 (11-16 years=0.42 [95% CI, 0.07, 0.77]; 16+ years=0.38 [95% CI, 0.00, 0.76]; Appendix III Table S1). With an interaction term for education and Chao1 in the outcome model, higher education was associated with lower likelihood of MCI (11-16 years=-1.24 [95% CI, -2.12, -0.35]; 16+ years=-1.26 [95% CI, -2.22, -0.30]) whereas greater Chao1 was not significantly associated with MCI (coefficient=-0.14, [95% CI, -0.76, 0.44]). Interaction terms were not significant (11-16 years of education:Chao1=-0.02 [95% CI, -0.80, 0.79]; 16+ years of education:Chao1=-0.20 [95% CI, -1.00, 0.62]). With Chao1 as mediator, NDE (16+ years of education) was 0.35 (95% CI, 0.15, 0.81, *p*=.02, Table 6) and NIE (16+ years of education) was 0.89 (95% CI, 0.68, 1.14, *p*=.33) suggesting an association of education to lower MCI risk, not mediated by Chao1 (total effect=0.31 [95% CI, 0.14, 0.72], *p*=.008; CDE=0.33 [95% CI, 0.14, 0.78], *p*=.02). PE=0.22 (95% CI, 0.00, 0.59, *p*=.049), suggests most of the association of education on MCI risk being due to a direct effect of education but also a significant amount due to interaction, mediation, or both.

Table 6 *Mediation Analysis With Chao1 as Mediator*

Estimand	Comparing 0-10 to 11-16 Years of Education				Comparing 0-10 to 16+ Years of Education			
	With Interaction		Without Interaction		With Interaction		Without Interaction	
	Estimate (95% CI)	*p*	Estimate (95% CI)	*p*	Estimate (95% CI)	*p*	Estimate (95% CI)	*p*
RCDE	0.33 (0.15, 0.76)	.012 *	0.34 (0.16, 0.74)	.010 *	0.33 (0.14, 0.78)	.018 *	0.33 (0.15, 0.76)	.015 *
RPNDE	0.33 (0.15, 0.73)	.012 *	0.34 (0.16, 0.74)	.010 *	0.35 (0.15, 0.81)	.016 *	0.34 (0.15, 0.76)	.015 *
RTNDE	0.33 (0.15, 0.77)	.011 *	0.34 (0.16, 0.74)	.010 *	0.33 (0.15, 0.79)	.016 *	0.34 (0.15, 0.76)	.015 *
RPNIE	0.95 (0.70, 1.19)	.647	0.92 (0.77, 1.07)	.284	0.95 (0.71, 1.17)	.662	0.93 (0.78, 1.06)	.316
RTNIE	0.94 (0.71, 1.33)	.693	0.92 (0.76, 1.07)	.284	0.89 (0.68, 1.14)	.334	0.93 (0.77, 1.07)	.316
RTE	0.31 (0.15, 0.68)	.006 **	0.32 (0.15, 0.68)	.006 **	0.31 (0.14, 0.72)	.008 **	0.31 (0.14, 0.71)	.008 **
ERCDE	-0.53 (-0.74, -0.17)	.012 *	-	-	-0.53 (-0.76, -0.14)	.018 *	-	-
ERINTREF	-0.14 (-0.32, 0.05)	.114	-	-	-0.11 (-0.28, 0.10)	.218	-	-
ERINTMED	0.03 (-0.23, 0.31)	.792	-	-	0.01 (-0.25, 0.25)	.966	-	-
ERPNIE	-0.05 (-0.30, 0.19)	.647	-	-	-0.05 (-0.29, 0.17)	.662	-	-
ERCDE(P)	0.77 (0.42, 0.99)	.008 **	-	-	0.78 (0.41, 1.00)	.010 *	-	-
ERINTREF(P)	0.20 (-0.09, 0.56)	.115	-	-	0.17 (-0.20, 0.49)	.216	-	-
ERINTMED(P)	-0.04 (-0.52, 0.41)	.796	-	-	-0.01 (-0.41, 0.45)	.964	-	-
ERPNIE(P)	0.07 (-0.32, 0.53)	.651	-	-	0.07 (-0.29, 0.52)	.668	-	-
PM	0.03 (-0.15, 0.28)	.694	0.04 (-0.04, 0.25)	.288	0.06 (-0.06, 0.38)	.340	0.03 (-0.04, 0.24)	.321
INT	0.16 (-0.05, 0.40)	.104	-	-	0.16 (-0.02, 0.38)	.082	-	-
PE	0.23 (0.01, 0.58)	.043 *	-	-	0.22 (0.00, 0.59)	.049 *	-	-

Note. Results of mediation analysis with or without interaction terms of education and Chao1 in the outcome model. Standard errors were estimated with 5,000 bootstraps. RCDE=controlled direct effect odds ratio (referring to CDE); RPNDE=pure natural direct effect odds ratio (referring to NDE); RTNDE=total natural direct effect odds ratio; RPNIE=pure natural indirect effect odds ratio; RTNIE=total natural indirect effect odds ratio (referring to NIE); RTE=total effect odds ratio; ERCDE=excess relative risk due to controlled direct effect; ERINTREF=excess relative risk due to reference interaction; ERINTMED=excess relative risk due to mediated interaction; ERPNIE=excess

relative risk due to pure natural indirect effect; ERCDE(P)=proportion ERCDE; ERINTREF(P)=proportion ERINTREF; ERINTMED(P)=proportion ERINTMED; ERPNIE(P)=proportion ERPNIE; PM=overall proportion mediated; INT=overall proportion attributable to interaction; PE=overall proportion eliminated. Cells with – indicate n/a. *p<.05. **p<.01. ***p<.001.

Given the moderately rare outcome (~22.5% MCI), CDE, NDE and NIE reported on the odds-ratio scale may be overestimated. As a sensitivity analysis, estimation was repeated on the risk-ratio scale using a multinomial log-linear link for the outcome model, resulting in a similar pattern of findings but without significant PE, results not shown (Valeri & VanderWeele, 2013).

Analyses without interaction terms led to similar result patterns in regression models (Appendix III Table S1). Comparison of 0-10 to 11-16 years of education led to similar result patterns in effect decomposition (Table 6). BMI was hypothesised as a potential mediator of education and MCI, or microbiome diversity and MCI, and thus not included in the main analyses but considered for robustness checks. Inclusion of BMI led to attenuated associations of education with Chao1 in the mediator, and of Chao1 with MCI in the outcome model. This in turn led to attenuated NIE and a similar, but no longer significant estimate of PE (results not shown).

Analyses with Shannon or inverse Simpson as mediator suggested similar findings but no significant PE. Analyses with inverse Simpson as mediator suggested similar findings except for no significant association of education with alpha diversity in the mediator model and a significant proportion of the total effect of education due to (additive) interaction, when comparing 0-10 to 11-16 years of education (regression models: Appendix III Table S2, Appendix III Table S3, effect decomposition: Appendix III Table S4, Appendix III Table S5).

Ldm suggested no significant mediation by individual taxa or by the composition of the microbiome (p=.99 for N=48 000 completed permutations with *ldm*.omni3). Likewise, *permanovaFL* suggested no significant mediation by the composition of the microbiome, on the relative abundance (Bray-Curtis dissimilarity, p=.70), or presence-absence scale (Jaccard dissimilarity, p=.35), or overall (p=.54 for N=600 completed permutations with

permanovaFL.omni). Robustness checks (with BMI) yielded a similar pattern of findings (results not shown).

III.4 Discussion

Higher education was associated with a lower risk of MCI, with most of this association not being due to mediation by the gut microbiome. Despite differences in taxonomic signatures and gut microbiome composition between education groups, our findings suggest no significant mediation of the association of education with MCI by measures of alpha diversity or individual taxa. However, effect decomposition indicated potential additive interaction between education and alpha diversity.

In this study, MCI risk was highest in the group with 0-10 years of education. Higher education groups did not differ in their association with MCI. This reflects earlier findings suggesting that education is related to reserve capacity, and thus lower MCI risk, by in particular increasing levels of cognitive skills in early life which then persist until old age (Lövdén et al., 2020).

Critically, more than 16 years of education likely reflect education beyond the end of adolescence, with positive effects levelling off and thus, no linear association of education with MCI.

Further analyses suggested a dominating direct effect of education. While education was associated with microbial diversity, no indicator of diversity was significantly associated with MCI, although less clear so for Chao1, reflecting richness, in models without interaction terms. Nonetheless, one fifth of the association of education on MCI could be removed (i.e., PE) by intervening to fix Chao1 at the sample mean. Four-way decomposition suggests this to be most likely attributable to an additive interaction of education and Chao1, such that their association with lower MCI risk increases with increments in education (VanderWeele, 2014). Of note, this finding reinforces most of the association of education with MCI to be

flowing through a direct causal path, which is also supported by sensitivity analysis on the risk-ratio scale.

A potential explanation for the absence of statistically significant mediation would be that lower education may proxy higher MCI risk due to factors which are not associated with the gut microbiome, such as cognitive stimulation. In that case, the observed variation in gut microbiome diversity and composition across education groups would not be causally related to MCI risk.

However, our findings highlight education-related gut microbiome diversity and composition reflecting those found in MCI and AD. Given MCI as a strong risk factor and AD as the most common cause of dementia, similarities in the gut microbiota of individuals with low education – who are at higher risk of dementia – and of people living with AD may indicate further mechanisms contributing to the disease. These may involve nutritional choices and chronic low-grade inflammation or the synthesis of metabolites leading to modulation of nerve signalling via the enteric nervous system. A previous study found reduced richness as well as a distinct composition of the gut microbiome in terms of beta diversity in participants with AD compared to healthy controls (Vogt et al., 2017). In line with a hypothesised neurodegenerative pathway involving education and the gut-brain-axis, our findings suggest that lower education is associated with reduced richness and a distinct gut microbiome composition. Conversely, another study found increasing richness with AD progression, which may be explained by an apparent gradient of education from lowest, in unimpaired cognition, to highest, in moderate AD (L. Chen et al., 2022). Considering our findings lower education may not only have altered the likelihood of belonging to patient or control groups but may also have resulted in different taxonomic signatures.

Previous findings suggest similar alterations with respect to reduced alpha diversity in lower income and area-level SES settings (Bowyer et al., 2019; G. E. Miller et al., 2016).

Education is related to income and wealth, and consequently with selection into areas with fewer socioeconomic resources. As such, education may capture community-level or spatial exposures affecting gut microbiome composition (Bowyer et al., 2019).

Critically, we found Chao1 to be lower in older age, but only in participants with lower education. Additionally, compositional differences by education were only significant in older age. This extends on earlier reports of interindividual variability and reduced biodiversity in later life by suggesting education as a key modifier (Biagi et al., 2013). Moreover, our finding of lower alpha diversity in lower education, suggests a putative association with a dysbiotic state. While lower alpha diversity has been discussed previously as a possible indicator of AD, to date, no concrete link has been established between education and dysbiosis and consequently AD or MCI (Cabrera et al., 2021; Z. Li et al., 2022). This may be due to the relatively limited depth and linked resolution of sequencing or the breadth of education measures (Cabrera et al., 2021).

Extending on mediation results with alpha diversity metrics, *ldm* and *permanovaFL* did not identify mediating taxa or compositional changes translating into decreased MCI risk (Hu & Satten, 2020; Yue & Hu, 2022a, 2022b). However, DAA results suggest differential abundance in line with an MCI or AD phenotype and consequentially a potential communality of lower education and AD pathology. Bacilli (class), Actinobacteria (class), Lactobacillales (order), Streptococcaceae (family), *Streptococcus* (genus), *Lachnospiraceae* UCG 001 (genus) and two ASVs were depleted in lower education.

Contrary to our findings given a hypothesised link of education to MCI via the gut, previous studies showed an increased ratio of Firmicutes to Bacteroidetes, and an increased relative abundance of Lactobacillales in AD, and of Firmicutes in MCI (Nagpal et al., 2019; Saji, Niida, et al., 2019). However, earlier findings were likely driven by depleted Bacteroides; increases in Firmicutes were not statistically significant (Nagpal et al., 2019;

Saji, Niida, et al., 2019). Other studies found increased Bacteroidetes in MCI without AD and, in line with our findings, depletion of Firmicutes in AD and amnestic MCI (Liu et al., 2019; Saji, Murotani, et al., 2019; Vogt et al., 2017). Increased Bacteroides may relate to impaired cognition potentially through cerebral small vessel disease and resulting WMH and, in line with our findings, education may alter BR to such damage via increased node degree (DeJong et al., 2023; Saji, Murotani, et al., 2019).

Lachnospiraceae UCG 001 were earlier found to be depleted in participants with more severe depressive symptoms, implying impaired synthesis of short-chain fatty-acids, such as butyrate and other depression-related neurotransmitters (Radjabzadeh et al., 2022). Recent findings further suggest lower cognitive performance due to decreased levels of butyrate following stool transplantation of sleep deprived to control mice (X. Wang et al., 2023). Depletion of *Lachnospiraceae* UCG 001 in lower education may be associated with lower cognitive performance and MCI classification in line with a phenotype related to depressive symptom severity. Since we adjusted for BDI-I, education and depressive symptom severity may share a common neuroendocrinal pathway to impaired cognition, e.g., via nutritional choices, involving *Lachnospiraceae* UCG 001 and metabolites synthesised by gut microbiota.

Moreover, two differentially abundant ASV were identified, one classified as *Lachnospiraceae* UCG 001, the other as *NK4A214* group, albeit classification comes with some uncertainty (Species unidentified, Appendix III Table S6).

One study found Actinobacteria and *Streptococcus* enriched in mild and moderate AD, potentially explained by higher educational attainment in AD-groups compared to controls, given our findings of higher abundance of Actinobacteria and *Streptococcus* in higher education (L. Chen et al., 2022). Consequently, both depleted Actinobacteria and *Streptococcus* in lower education may reflect an AD phenotype but the alternative

explanation that their alteration reflects educational differences cannot be ruled out. Of note, Actinobacteria was earlier found depleted in AD compared to healthy controls suggesting e.g., detriments to intestinal barrier integrity in both AD and lower education (Vogt et al., 2017).

III.4.1 Limitations

In this study, we extensively triangulated potential mediation in a large cohort and point to potential interaction of education and gut microbiome diversity regarding MCI risk. Despite careful adjustment, residual confounding may bias results. Effect decomposition assumes no unmeasured (i) exposure-outcome, (ii) mediator-outcome, or (iii) exposure-mediator confounding, and (iv) that (ii)-confounders are not affected by the exposure. However, PE and CDE do not require (iii) or (iv), and NDE is robust to (iv), assuming monotone associations (Pearl, 2001; Tchetgen & VanderWeele, 2014). MCI classification was based on a screening instrument. Differences in causes underlying MCI classification may bias DAA, which we could not formally assess (Alzheimer's Association, 2022). Moreover, SES was not formally addressed in the present analyses and may reflect a common cause of or indirect causal path variable of educational differences and MCI risk. However, DAA was carefully adjusted for different confounder sets, including lifestyle-related variables and risk factors of impaired cognition, such as BMI, depressive symptoms, or PS. Grouping of education may bias estimates, although the associations of years of education with MCI is likely non-linear. Further, compulsory schooling years vary across birthyears (Honig, M. S. & Bock, 2017). Grouping by less than 10 years of education selects older participants or those that immigrated. However, analysis was adjusted for age and first language as proxy for immigration. Limited diversity and sample size in this cohort prevented subgroup analysis and hampers generalisability (Appendix III Table S7). Further analysis

regarding functional categories and diversity are necessary to fully elucidate implications of distinct taxonomic signatures (Heintz-Buschart & Wilmes, 2018).

III.4.2 Conclusions

Our results suggest signatures of the gut microbiome that have been identified previously in AD and MCI to be reflected in lower education. We show that most of the association of education with MCI is of a direct nature and stress the importance of considering SDoH, specifically education, as key modifiers in microbiome studies. Our findings underline the potential of the gut microbiome as a biomarker and intervention target regarding MCI, which is promising, considering its modifiability until later life. Future research with longitudinal survey designs is required to further investigate potential interaction of education and the gut microbiome and their implication for neurodegenerative diseases.

Chapter IV – Towards Risk Prediction: 'Probable Dementia' Classification With Country-Level Variation in Prevalence in Europe

Klee, M., Langa, K. M., & Leist, A. K. (2024). Performance of probable dementia classification in a European multi-country survey. *Scientific Reports*, *14*(1), 6657. https://doi.org/10.1038/s41598-024-56734-7

Abstract

Feasibility constraints limit availability of validated cognitive assessments in observational studies. Algorithm-based identification of 'probable dementia' is thus needed, but no algorithm developed so far has been applied in the European context. The present study sought to explore the usefulness of the Langa-Weir (LW) algorithm to detect 'probable dementia' while accounting for country-level variation in prevalence and potential underreporting of dementia.

Data from 56 622 respondents of the Survey of Health, Ageing and Retirement in Europe (SHARE, 2017) aged 60 years and older with non-missing data were analysed. Performance of LW was compared to a logistic regression, random forest and XGBoost classifier. Population-level 'probable dementia' prevalence was compared to estimates based on data from the Organisation for Economic Co-operation and Development.

As such, application of the prevalence-specific LW algorithm, based on recall and limitations in instrumental activities of daily living (IADL), reduced underreporting from 61.0% (95% CI, 53.3%, 68.7%) to 30.4% (95% CI, 19.3%, 41.4%), outperforming tested machine learning algorithms. Performance in other domains of health and cognitive function was similar for participants classified 'probable dementia' and those self-reporting physician-diagnosis of dementia.

Dementia classification algorithms can be adapted to cross-national cohort surveys such as SHARE and help reduce underreporting of dementia with a minimal predictor set.

IV.1 Introduction

The WHO considers dementia, a condition characterised by memory and other cognitive impairments severe enough to cause the loss of independent function, to be a public health priority as the syndrome represents one of the main causes of death and dependency among older people (ADI et al., 2012). Dementia causes significant economic, health and social care burden for those living with dementia and their informal caregivers. The number of people affected by dementia is expected to increase in the coming decades (Nichols et al., 2022). Due to resource intensity of systematic dementia ascertainment in representative cohort studies, algorithmic classifications of dementia are needed to inform research and reduce potential underreporting.

Dementia classification algorithms determine participants' dementia status based on cognitive tests or sociodemographic variables that are readily accessible in cohort surveys such as the Health and Retirement Study (HRS) in the U.S. (Alzheimer's Association, 2010; Crimmins et al., 2011; Herzog & Wallace, 1997; Hurd et al., 2013; Manly et al., 2022; Q. Wu et al., 2013). Existing algorithms frequently rely on (regression-based) prediction models or composite scores with an a priori cutoff for classification. In general, score cutoff based approaches facilitate interpretation, primarily due to a lower number of indicators and straight forward computation compared with regression-based classification. Langa, Kabeto and Weir developed a widely applied and previously validated score cutoff based algorithm (LW) to infer 'probable dementia' (Alzheimer's Association, 2010; Crimmins et al., 2011; Gianattasio et al., 2019). However, established dementia classification algorithms have not been systematically tested in the European, cross-national context, yet (Alzheimer's Association, 2010; Crimmins et al., 2011).

The SHARE is a sister study to HRS. Nevertheless, direct application of well-established dementia classification algorithms is hindered due to differences in assessment

protocols. Furthermore, cutoffs are not directly transportable since sample demographics, cognitive performance, indicator-outcome relationships, or reporting styles may vary across countries (Bond et al., 2005; d'Uva et al., 2011; Formánek et al., 2019).

We sought to examine the potential of the LW classification to detect 'probable dementia' using a minimal predictor set, with the aim of compensating for underreporting of dementia in cohort studies in the European context. Thus, we investigate the performance of a range of algorithms to detect 'probable dementia' and to adjust for country-level variation in underreporting of dementia in SHARE (Börsch-Supan et al., 2013). For this purpose, we adapted the LW classification to available indicators in SHARE, defining country-specific cutoffs. Performance was compared to a set of benchmark ML algorithms to test for possible improvements with larger predictor sets and higher model complexity, specifically, a weighted logistic regression model (GLM), a random forest (RF) and an XGBoost (XGB) classifier (T. Chen & Guestrin, 2016). Validity of classifications was assessed (a) on the population level by comparing country-specific ('probable') dementia prevalence before and after application of the algorithms to projections based on data from the Organisation for Economic Co-operation and Development (OECD) and a population representative study in Israel, and (b) on the individual level by assessing performance of those classified 'probable dementia' in further domains of health and cognitive function (Kodesh, 2019; OECD, 2018).

IV.2 Methods

IV.2.1 Study Sample

SHARE is a representative, multi-country cohort study with over 140 000 participants aged 50 years and biennial follow-up from 2004 to 2021 (Bergmann, Kneip, et al., 2019; Bergmann, Scherpenzeel, et al., 2019; Börsch-Supan, 2022; Börsch-Supan et al., 2013). Activities of the SHARE-European Research Infrastructure Consortium related to human subjects research are guided by international research ethics principles such as the Respect

Code of Practice for Socio-Economic Research (professional and ethical guidelines for the conduct of socio-economic research) and the 'Declaration of Helsinki' (a set of ethical principles regarding human experimentation developed for the medical community by the World Medical Association, last revised at the 64th WMA Meeting held in Fortalezza/Brazil in October 2013). SHARE waves 4 and following were reviewed and approved by the Ethics Council of the Max Planck Society (https://share-eric.eu/data/faqs-support). The SHARE data collection procedures are subject to continuous ethics review.

We used data from SHARE wave seven (2017 to 2019) due to its large sample size across 26 European countries and Israel (Börsch-Supan, 2022). Data from countries with at least five participants self-reporting physician-diagnosis of dementia were eligible for analyses. Although we further address class imbalance, a cutoff of five was enforced to avoid biases in performance comparisons emerging from countries with outlying, extremely low dementia case numbers. Participants aged 60 years and older, with non-missing data on relevant sociodemographic, health or cognitive items were included in our analytic data set. All participants provided informed consent.

IV.2.2 Measures

Participants self-report physician-diagnosis of dementia with a single item question; 'Has a doctor ever told you that you had/Do you currently have Alzheimer's disease, dementia, […]' (CentERdata, Institute for data collection and research & SHARE Central, Munich Center for the Economics of Aging, 2024).

The LW algorithm classifies participants based on their performance in cognitive tests or based on items characterising participants' cognitive status that are provided by proxy respondents (Alzheimer's Association, 2010; Crimmins et al., 2011). Proxy respondents answer on behalf of the main respondent in case of physical or cognitive limitations. The LW algorithm classifies participants with three groups: 'normal', 'probable dementia', i.e.,

‘demented’ in original LW, and ‘cognitive impairment without dementia’ (CIND), i.e., ‘cognitive impairment, not demented’ in original LW (Alzheimer’s Association, 2010; Crimmins et al., 2011; Langa et al., 2005).

LW classifications for self-respondents are based on immediate (0 to 10) and delayed (0 to 10) recall, Serial 7’s (0 to 5) and backwards counting (0 to 2) tasks (Alzheimer’s Association, 2010; Crimmins et al., 2011). LW classifications for proxy respondents are based on proxy-rated memory (0 to 4), interviewer-perceived quality of cognition (0 to 2) and five IADL (0 to 5) limitations (Alzheimer’s Association, 2010; Crimmins et al., 2011).

Cutoffs were defined so that the prevalence of categories resulting from classification of the full HRS sample match the population-level prevalence of CIND and dementia identified in the population-representative Aging, Demographics, and Memory Study, with equipercentile equating (Alzheimer’s Association, 2010; Crimmins et al., 2011; Langa et al., 2005).

Distinctive features of SHARE and HRS hamper direct application of LW to SHARE, despite similar assessment protocols. First, there are no proxy-reported cognitive function measures available in SHARE wave seven. Second, nine limitations with IADLs are available in SHARE wave seven, but self-reported. Third, the backwards counting task is not available in SHARE wave seven. Fourth, the Serial 7’s is only available for a subset of 12 countries and was thus only used to assess validity of classifications in SHARE wave seven.

To examine the impact of including self-reported IADLs in classification, two LW adaptations were derived. LW (Recall) is based on immediate (0 to 10) and delayed (0 to 10) recall. LW (Recall & IADL) is based on LW (Recall) and nine IADLs (0 to 9) limitations (Appendix IV Table S1).

With a smaller number of items, sum score ranges are narrower, and hence pre-established cutoffs prone to misclassification. This motivated updating cutoff definitions for SHARE.

First, since there is no measure for CIND, we defined cutoffs for classifying 'probable dementia', but not CIND. However, definition of cutoffs with the equipercentile equating approach is hampered in absence of neuropsychological assessment informing about representative prevalence of dementia in SHARE. We thus introduced externally validated prevalence estimates from a national representative study in Israel and projections published by the OECD, which are based on data from the World Alzheimer Report 2015 and population structure estimates from the United Nations (Kodesh, 2019; OECD, 2018; United Nations, Department of Economic and Social Affairs, Population Division, 2017).

Second, comparison suggested varying degrees of underreporting across countries, defined as discrepancy in prevalence estimates based on OECD data and self-report physician-diagnosis of dementia in SHARE (Kodesh, 2019; OECD, 2018; Prince et al., 2015; United Nations, Department of Economic and Social Affairs, Population Division, 2017). In example, some countries with similar dementia prevalence in SHARE vary in prevalence according to OECD data (Kodesh, 2019; OECD, 2018). Moreover, cross-national differences in mean recall performance and the number of reported IADLs indicate that cutoffs need to be defined within countries (Kodesh, 2019; OECD, 2018; Prince et al., 2015).

Consequently, two sets of cutoffs for LW (Recall) were defined, based on percentiles reflecting prevalence estimates reported by the OECD (i.e., equipercentile equating approach) or based on the 2.5th percentile. With equipercentile equating, the cutoff reflects external information on country-level dementia prevalence (Kodesh, 2019; OECD, 2018). With the 2.5th percentile, the cutoff is in line with the average population weighted dementia prevalence across countries in SHARE (M=2.2%) and reflects an outlier definition, two

standard deviations below the mean (for a normally distributed variable). Scores below either cutoff led to LW (Recall) classification ‘probable dementia’.

For LW (Recall & IADL) a naïve IADL cutoff was defined to reflect outliers, one and a half interquartile ranges above Q3. In countries with Q3 equal to zero the cutoff was set to 1. Scores above this cutoff led to LW (Recall & IADL) classification ‘probable dementia’ if LW (Recall) was classified ‘probable dementia’, too.

Consequently, two LW algorithms were specified, i.e., based on Recall (LW [Recall]), or Recall and IADLs (LW [Recall & IADL]) with two alternative cutoffs for Recall (LW [Recall]^{P} – prevalence-based; or LW [Recall] – outlier based). For LW (Recall & IADL) and LW (Recall & IADL)^{P} the same naïve IADL cutoff was used, irrespective of the cutoff used for Recall.

To examine performance of different specifications we compared the four LW algorithms with different sets of indicators (based on cognitive tests and IALDs) and cutoffs (based on prevalence or outlier definitions). Additionally, we compared the four LW algorithms to three functions commonly classified as ML algorithms, GLM, RF, and XGB, both latter relaxing parametrical assumptions and allowing for non-linear higher-order interactions (T. Chen & Guestrin, 2016; Leist et al., 2022). RF classifier aggregate information of individual decision trees, created with random subsets of predictors following the concept of bootstrapping (Leist et al., 2022). XGB classifier are based on a sequential ensemble of individual decision trees used to minimise the prediction error in final data partitions (T. Chen & Guestrin, 2016; Leist et al., 2022).

ML algorithms included immediate and delayed recall, individual ADLs/IADLs, and sociodemographic indicators age, education (tertiary/upper secondary/lower secondary), and sex/gender (male/female). Additionally, interviewer-rated variables were included comprising provided reading assistance (yes/no), willingness to answer (good/bad),

clarification/comprehension questions (6-step Likert Scale from Never to Always). In rare circumstances proxies that were present during the interview reported IADLs on behalf of (0.7% of full sample), or together with the respondent (1.7% of full sample). Information on the presence and type of proxy was thus included in ML algorithms (No, Partner, Relative, Helper/Other). The outcome (class) used for model training was self-report physician-diagnosis of dementia.

To address class imbalance (i.e., majority of participants without self-report physician-diagnosis of dementia), three training sets were defined, by random split, downsampling the majority class, or the synthetic minority oversampling technique (SMOTE; Chawla et al., 2002). With SMOTE, new cases are created based on the k-nearest neighbours of the minority class (Chawla et al., 2002).

Hyperparameters of RF and XGB models were tuned using grid search in five-fold cross validation with the area under the receiver operating characteristic curve (AUC) as criterion for selection of the best specification. Sampling weights were derived for GLM based on the inverse of the country-specific prevalence (or 1 minus the country-specific prevalence) for the minority (or majority) class (Kodesh, 2019; OECD, 2018).

Consequently, LW algorithms were compared to three (GLM, RF, XGB) x three (random split, downsampling, SMOTE) + one GLM (weighted) benchmark ML-based algorithms. We will only discuss GLM weighted, RF SMOTE and XGB SMOTE in the following sections.

IV.2.3 Statistical Analysis

Descriptive characteristics of the three training sets (random split, downsampling, SMOTE) and test set were assessed at baseline with Student's t-Tests for continuous and Chi-squared Tests for categorical characteristics.

Model performance for all specifications was assessed in the same test set. In a first step, ML-based algorithms were trained and cutoffs for LW were defined in the training set. Second, classifications for LW and ML-based algorithms were computed for the test set, and performance was assessed comparing self-report physician-diagnosis of dementia to 'probable dementia' with multiple indicators (e.g., AUC, F1, sensitivity, specificity).

Then, country level variation in population-weighted 'probable dementia' prevalence was compared to previously reported estimates. First, per-country prevalence estimates were plotted according to observed dementia status in SHARE and previously reported figures. Then, underreporting across countries when applying classification algorithms was computed. Underreporting for individual countries was calculated as denoted in (1). N_{SHARE} is the number of dementia cases in the test set, based on the population-weighted prevalence according to each algorithm. N_{OECD} is the number of dementia cases in the test set, based on prevalence estimates reported by the OECD.

$$underreporting = 1 - \left(\frac{n_{SHARE}}{n_{OECD}}\right) \quad (1)$$

Prevalence estimates were mapped, to explore geographical patterns. Mean values in further domains of health and cognitive function were compared in 'probable dementia' and self-reported physician-diagnosis of dementia to assess validity of classifications. Finally, performance metrics were stratified by country to inspect fairness of classifications. All analyses were performed in R version 4.2.0 (R Core Team, 2022).

IV.3 Results

Of 77 202 participants in SHARE wave seven, a total of *N*=56 622 (*M* [*SD*] age=71.7 [8.1] years; 56.3% female) from 26 countries were eligible to our analysis of which 2.1%

reported physician-diagnosis of dementia (Figure 4). Baseline characteristics are provided in Table 7.

Figure 4 *Flow Chart Illustrating Sample Size According to Eligibility Criteria*

Note. Numbers refer to sample size after application of the eligibility criteria, denoted in bold font.

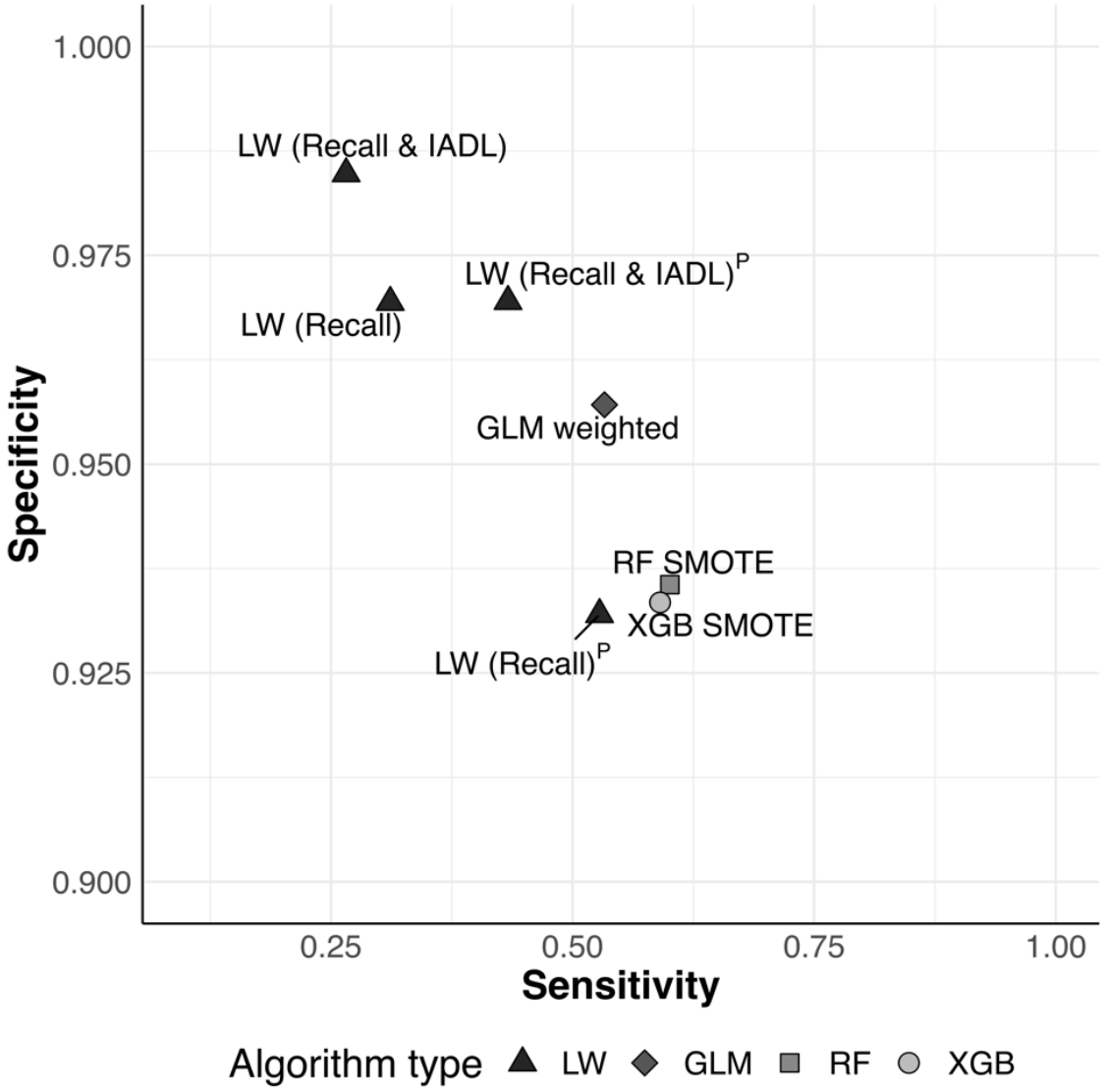
Table 7 *Baseline Characteristics (Study 3)*

Characteristics	Test Set	Training Set		
	(*n*=28 312)	Random Split (*n*=28 310)	DOWN (*n*=1,170)	SMOTE (*n*=4,095)
Age, *M* (*SD*)	71.7 (8.05)	71.7 (8.08)	75.4 (8.89)	73.9 (8.72)
Gender				
Female	15 937 (56.3%)	15 931 (56.3%)	682 (58.3%)	2,321 (56.7%)
Male	12 375 (43.7%)	12 379 (43.7%)	488 (41.7%)	1,774 (43.3%)
Education (ISCED 1997)				
Lower secondary	11 418 (40.3%)	11 376 (40.2%)	589 (50.3%)	1,864 (45.5%)
Upper secondary	9,563 (33.8%)	9,512 (33.6%)	332 (28.4%)	1,280 (31.3%)
Tertiary	7,331 (25.9%)	7,422 (26.2%)	249 (21.3%)	951 (23.2%)
Dementia				
Yes	591 (2.1%)	585 (2.1%)	585 (50.0%)	1,170 (28.6%)
No	27 721 (97.9%)	27 725 (97.9%)	585 (50.0%)	2,925 (71.4%)

Note. Reported *p* values are based on Student's t-Tests for continuous and Chi-squared Tests for categorical characteristics. DOWN=training set created with downsampling; SMOTE=training set created with the synthetic minority oversampling technique.

Model performance was assessed regarding (balanced) accuracy, sensitivity, specificity (Figure 5), precision, F1 and AUC (Leist et al., 2022). All models accurately predicted 'probable dementia' (accuracy=0.83 to 0.98). However, performance varied for metrics that are more robust in imbalanced data (balanced accuracy=0.50 to 0.81; F1=0.01 to 0.30). Discrimination was moderate to good overall (AUC=0.63 to 0.90). For LW, sensitivity was higher with prevalence-based compared to statistically informed Recall cutoffs (Appendix IV Table S2). IADL inclusion in LW (Recall & IADL)^{P} increased specificity and combined good balanced accuracy (0.70), moderate AUC (0.70) and the best F1 across all algorithms (0.30).

Figure 5 *Sensitivity and Specificity of Classification Algorithms in the Test Set.*

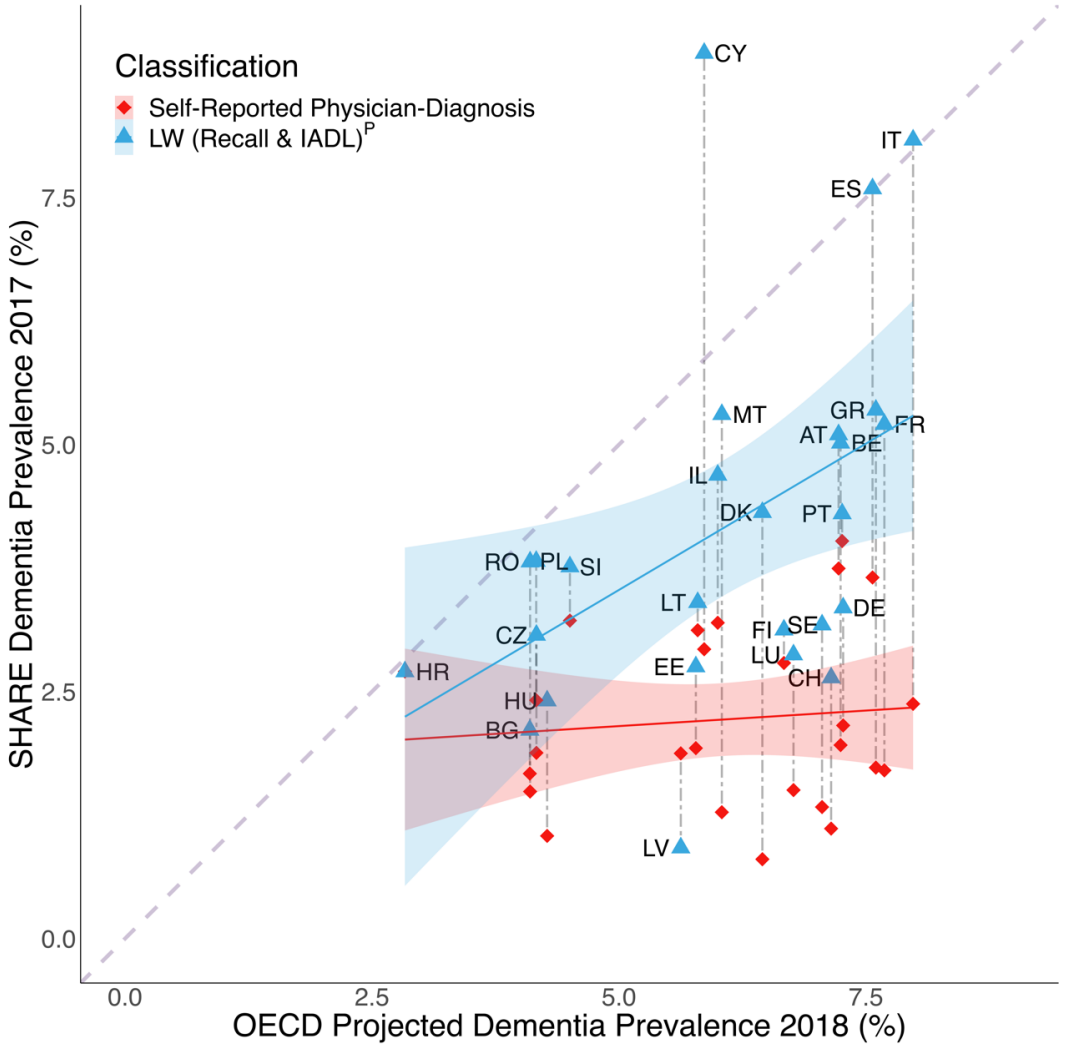


Note. LW (Recall)=Langa-Weir algorithm with a Recall-cutoff reflecting the 2.5th percentile; LW (Recall)$^{P=}$Langa-Weir algorithm with a Recall-cutoff reflecting country-level dementia prevalence; LW (Recall & IADL)=Langa-Weir algorithm based on LW (Recall) with an IADL cutoff reflecting 1.5 IQR above Q3; LW (Recall & IADL)$^{P=}$Langa-Weir algorithm based on LW (Recall)^{P} with an IADL cutoff reflecting 1.5 IQR above Q3; GLM weighted=weighted Logistic Regression, RF SMOTE=Random Forest trained in data created with the synthetic minority oversampling technique; XGB SMOTE=XGBoost trained in data created with the synthetic minority oversampling technique; IADL=Instrumental Activities of Daily Living.

For ML-based algorithms, GLM (weighted), RF SMOTE and XGB SMOTE showed the best performance combining good balanced accuracy (0.75 to 0.77), good AUC (0.86 to 0.89) and the best F1 within their algorithm type (0.26 to 0.30).

Regarding country-level variation in dementia prevalence, estimates based on SHARE with self-reported physician-diagnosis of dementia, or 'probable dementia' were compared to earlier reported country-specific prevalence (Figure 6). LW (Recall & IADL)^{P} 'probable dementia' prevalence was more similar to previous findings, suggesting less underreporting. A steeper slope of the linear fit further suggests less variation in underreporting across countries.

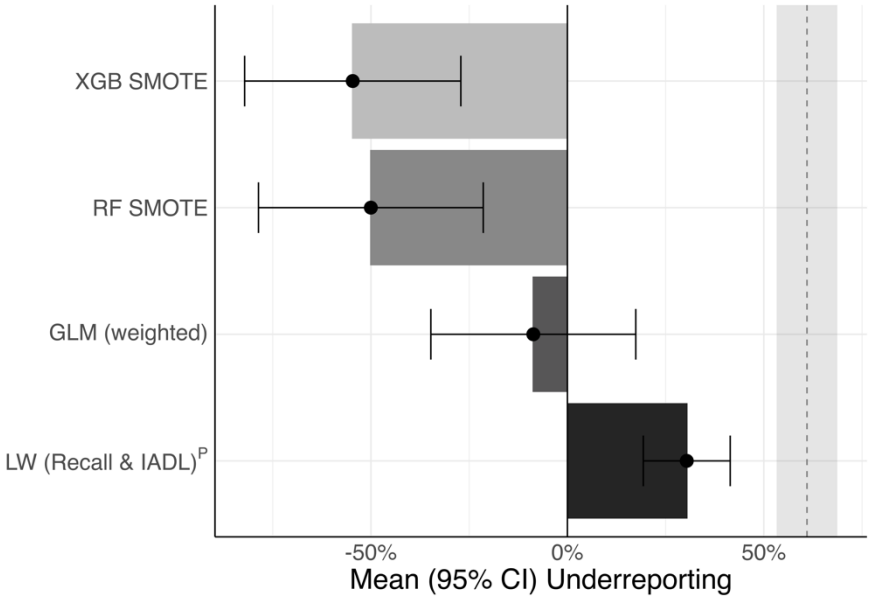
Figure 6 *Dementia Prevalence Across Countries*



Note. Population-weighted dementia prevalence across countries based on projections from the OECD and a population-based study in Israel (Kodesh, 2019; OECD, 2018) on x-axis and SHARE on y-axis. Red rectangles reflect estimates for self-reported physician-diagnosis of dementia. Blue triangles reflect estimates for LW (Recall & IADL)^{P} 'probable dementia'. Vertical dotted lines reflect the discrepancy between (probable) dementia prevalence estimates in SHARE. The diagonal dotted line reflects perfect overlap of estimates in SHARE and OECD. Solid lines and shaded areas reflect linear models and confidence limits for (probable) dementia prevalence-based on self-report physician-diagnosis (red) or LW (Recall & IADL)^{P} (blue). See ISO alpha 2 country codes in Appendix IV Table S3. SHARE=Survey of Health, Ageing and Retirement in Europe; OECD=Organisation for Economic Co-operation and Development; LW (Recall & IADL)^{P}=Langa-Weir algorithm with a Recall-cutoff reflecting country-level dementia prevalence and an IADL cutoff reflecting 1.5 IQR above Q3; IADL=Instrumental Activities of Daily Living.

Underreporting with a prevalence estimate based on self-report physician-diagnosis of dementia was 61.0% (95% CI, 53.3%, 68.7%) on average. Underreporting with a prevalence estimate based on LW (Recall & IADL)^{P} 'probable dementia' was reduced to 30.4% (95% CI, 19.3%, 41.4%) on average (Figure 7, Appendix IV Table S3).

Figure 7 *Mean Underreporting Across Countries With Classification Algorithms*

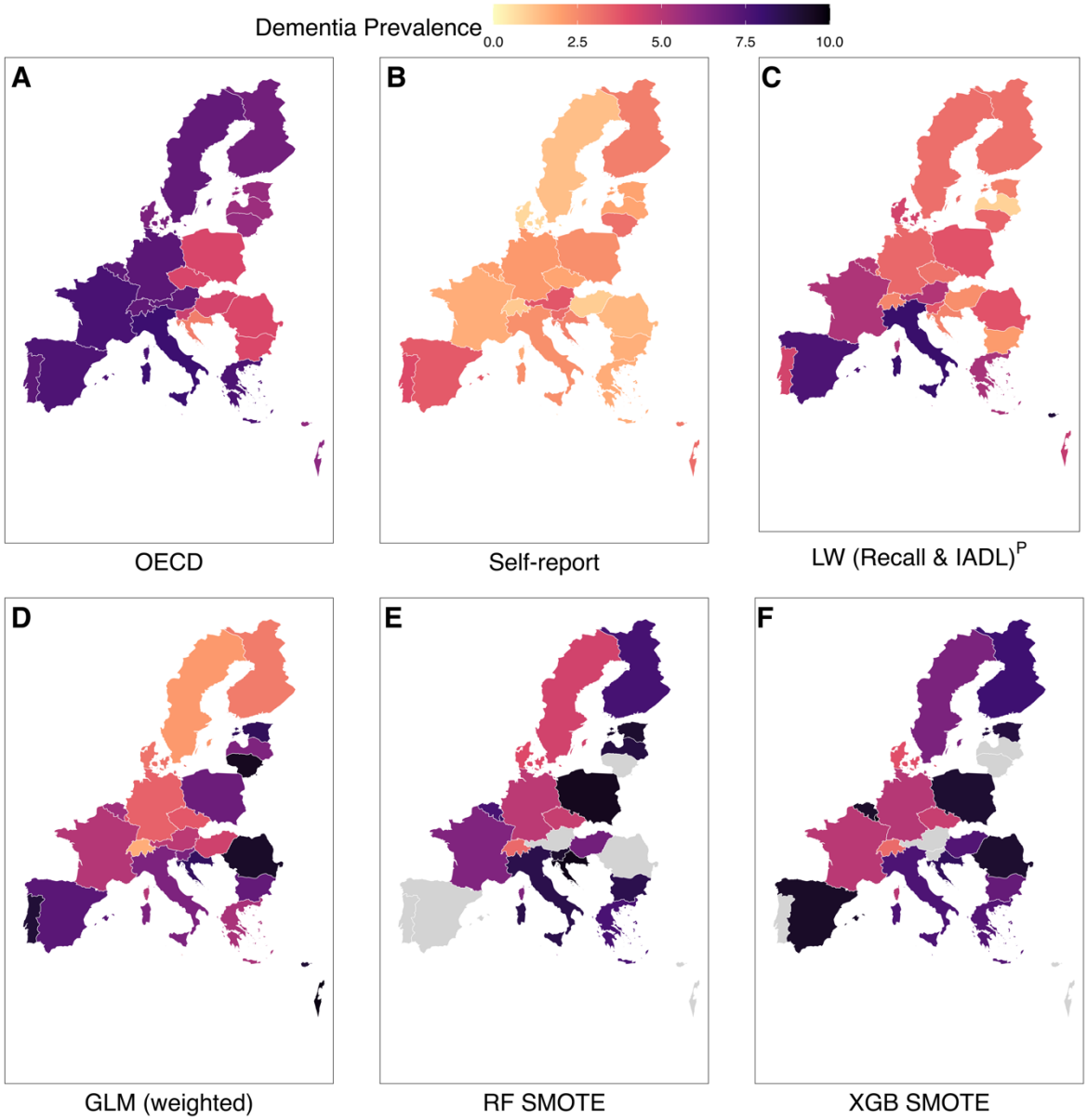


Note. Points indicate mean underreporting (error bars indicate 95% CI) when the number of people living with 'probable dementia' is compared to previously reported estimates (Kodesh, 2019; OECD, 2018). Solid vertical line indicates perfect overlap of the number of people living with 'probable dementia' in the test set when applying classification algorithms and the number of people living with dementia in the test set calculated based on previously reported population-level prevalence estimates (Kodesh, 2019; OECD, 2018). The dashed vertical line indicates mean underreporting (light grey fill indicates 95% CI) when the number of people living with self-reported physician-diagnosis of dementia is compared to previously reported estimates (Kodesh, 2019; OECD, 2018). LW (Recall)^{P}=Langa-Weir algorithm with a Recall-cutoff reflecting country-level dementia prevalence and an IADL cutoff reflecting 1.5 IQR above Q3; GLM weighted=weighted Logistic Regression; RF SMOTE=Random Forest trained in data created with the synthetic minority oversampling technique; XGB SMOTE=XGBoost trained in data created with the synthetic minority oversampling technique; IADL=Instrumental Activities of Daily Living.

Prevalence estimates based on GLM (weighted) suggested higher variation in underreporting and a negative linear trend (results not shown) despite a better reduction in underreporting (*M* [95% CI] underreporting=-8.7% [-34.8%, 17.4%]). Other ML algorithms drastically overestimated prevalence.

Prevalence estimates were further mapped to explore geographical patterns (Figure 8). Whereas previously reported estimates and SHARE estimates based on self-reported physician-diagnosis of dementia suggested overall differences in magnitude, previously reported estimates indicated low variation between neighbouring countries. Prevalence was overall higher with LW (Recall & IADL)^{P} 'probable dementia' compared to self-reported physician-diagnosis of dementia but lower compared to OECD projections. Differences in prevalence between neighbouring countries were smaller with LW (Recall & IADL)^{P} 'probable dementia' compared to self-reported physician-diagnosis of dementia but higher compared to previously reported estimates. GLM (weighted), RF SMOTE and XGB SMOTE reinforced discrepancies between some neighbouring countries and exceeded previously reported prevalence estimates.

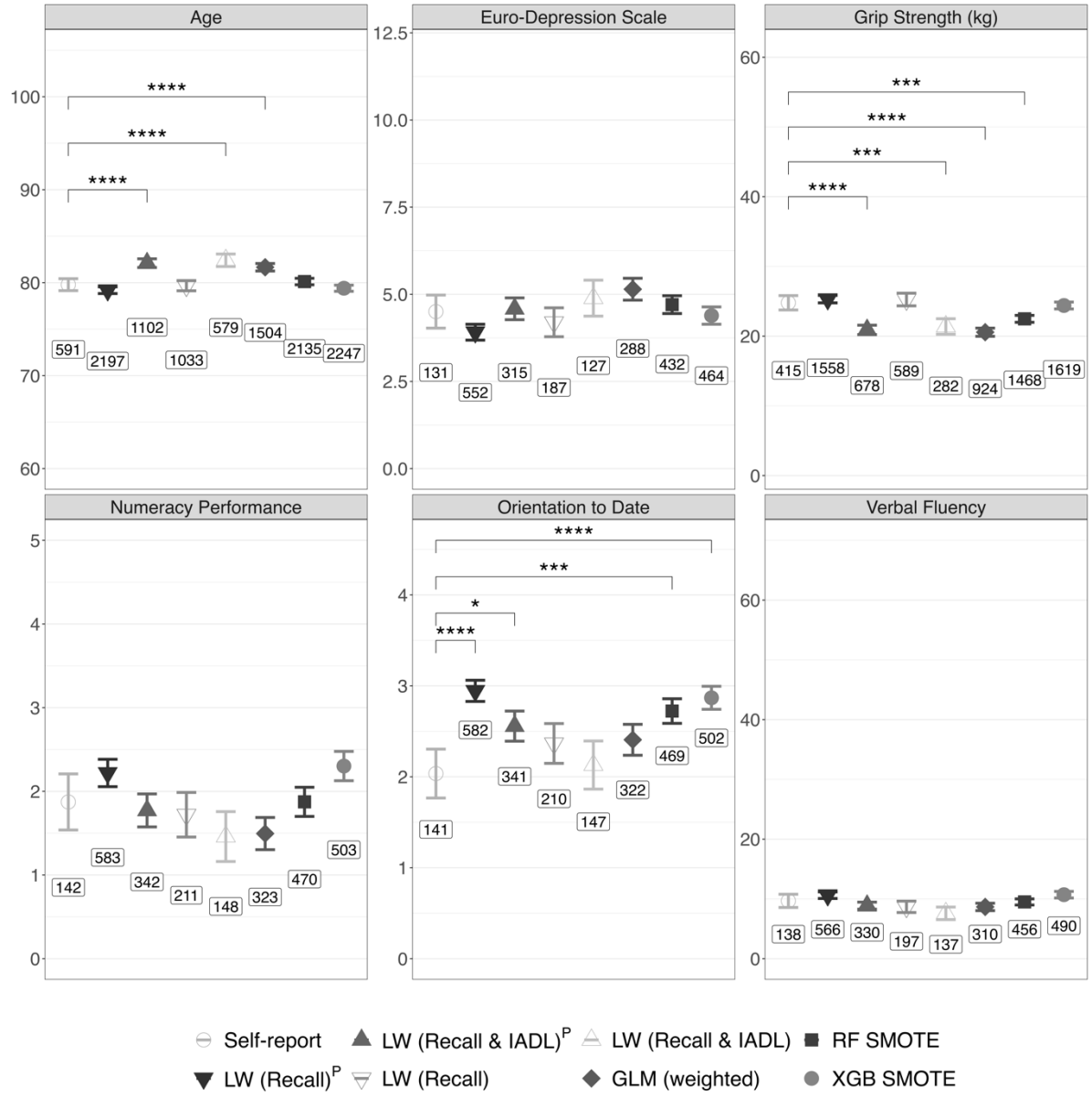
Figure 8 *Dementia Prevalence Maps*



Note. Population weighted dementia prevalence across countries based on A projections from the OECD and a population-based study in Israel (Kodesh, 2019; OECD, 2018), B self-reported physician-diagnosis of dementia, or C-F 'probable dementia' classification. Grey fill indicates prevalence > 10%. LW (Recall)^{P}=Langa-Weir algorithm with a Recall-cutoff reflecting country-level dementia prevalence and an IADL cutoff reflecting 1.5 IQR above Q3; GLM weighted=weighted Logistic Regression; RF SMOTE=Random Forest trained in data created with the synthetic minority oversampling technique; XGB SMOTE=XGBoost trained in data created with the synthetic minority oversampling technique; IADL=Instrumental Activities of Daily Living.

Validity was further assessed comparing mean values in further domains of health and cognitive function between ‘probable dementia’ and self-reported physician-diagnosis of dementia in complete cases (Figure 9). Results suggest good fit overall for depressive symptoms, verbal fluency, and numeracy performance (Prince et al., 1999). Grip strength aligned best with LW (Recall), LW (Recall)^{P}, and XGB SMOTE, just like age. Regarding orientation to date, only LW ‘probable dementia’ algorithms with statistically informed Recall cutoffs and GLM (weighted) overlap with self-reported physician-diagnosis of dementia.

Figure 9 *Validity assessment in further domains of health and cognitive function*



Notes. Means with 95% CIs for in (probable) dementia. Labels refer to the number of participants with (probable) dementia and complete data. LW (Recall)=Langa-Weir algorithm with a Recall-cutoff reflecting the 2.5th percentile; LW (Recall)$^{P=}$Langa-Weir algorithm with a Recall-cutoff reflecting country-level dementia prevalence; LW (Recall & IADL)=Langa-Weir algorithm based on LW (Recall) with an IADL cutoff reflecting 1.5 IQR above Q3; LW (Recall & IADL)$^{P=}$Langa-Weir algorithm based on LW (Recall)^{P} with an IADL cutoff reflecting 1.5 IQR above Q3; GLM weighted=weighted Logistic Regression; RF SMOTE=Random Forest trained in data created with the synthetic minority oversampling technique; XGB SMOTE=XGBoost trained in data created with the synthetic minority oversampling technique; IADL=Instrumental Activities of Daily Living. *p<.05. **p<.01. ***p<.001. ****p<.0001.

Performance metrics were stratified by country to inspect fairness of algorithms (Appendix IV Figure S1). Variation in performance is higher for LW (Recall & IADL)^{P} compared to ML-based algorithms for AUC, F1, precision, and sensitivity, but for balanced accuracy, variation is similar. For accuracy and specificity LW (Recall & IADL)^{P} shows the least variation.

IV.4 Discussion

In this study, we adapted the LW dementia classification algorithm and tested its ability to detect 'probable dementia' in the European context. LW proved useful to detect 'probable dementia' compared to when classification is based entirely on self-report of a physician-diagnosis. In validity checks on the population level, we found that LW based on immediate recall, delayed recall and IADLs with a prevalence-based recall cutoff (LW [Recall & IADL]^{P}) performed best in reducing underreporting across countries. On the individual level, performance profiles in other domains of health and cognitive function such as numeracy in 'probable dementia' matched those in participants who self-report physician-diagnosis of dementia in a subset of countries. Despite higher complexity and a larger number of indicators, ML-based classifiers performed less consistent across countries reinforcing the superiority of the adapted LW classification to help identify 'probable dementia' with a minimal predictor set.

A previous study suggested validity of LW classifications in the US context and performance in line with algorithms additionally incorporating demographic characteristics (Gianattasio et al., 2019). We found similar sensitivity and high specificity of the adapted LW (Recall & IADL)^{P} in the European context, despite a smaller set of indicators (Gianattasio et al., 2019). More recently, other ML-based algorithms classifying 'probable dementia' were evaluated in the European context (Cleret de Langavant et al., 2018, 2020; Gharbi-Meliani et al., 2023; Twait et al., 2023). In line with our findings, a recent study

suggested limited surplus in performance over logistic regression models when using complex ML-based algorithms, and only so with survival analyses (Twait et al., 2023). As an alternative to supervised learning, where models are trained on an a priori labelled class, a recent study applied a previously established unsupervised ML approach to clustering in SHARE, longitudinally (Cleret de Langavant et al., 2018; Gharbi-Meliani et al., 2023). AUC and sensitivity of LW (Recall & IADL)^{P} classification were in line with the clustering based classification in SHARE wave seven, although being marginally lower (Gharbi-Meliani et al., 2023). This suggests similar classification performance with the score cutoff based algorithm. LW (Recall & IADL)^{P} application requires no longitudinal follow-up and hence, compared with longitudinal algorithms, we were able to classify more data points from more countries at a given wave, which however precluded direct comparison, e.g., of the number of newly identified 'probable dementia' (Gharbi-Meliani et al., 2023). Nonetheless, including external information on population-level dementia prevalence suggested that LW (Recall & IADL)^{P} identified an expected additional number of 'probable dementia', thus suggesting reduced potential underreporting, while maintaining high specificity (Kodesh, 2019; OECD, 2018).

We found variation in classification performance across countries. Such variation was also apparent in a recent study in SHARE, and may be due to descriptive differences across countries, emerging e.g., from differences in population structures (Gharbi-Meliani et al., 2023). As an example, sensitivity was lowest in countries with lower mean recall performance, especially with LW (Recall & IADL)^{P}, suggesting floor effects during cutoff definitions. Furthermore, prevalence in countries with distinct distributions of antecedents to dementia, was systematically overestimated with ML-based classifiers, e.g., for Eastern compared to Northern European countries (Alzheimer Europe, 2019). Another reason for country-level variation may be differential association of included indicators with dementia risk, e.g., depending on welfare regimes, or policy across countries. Although we cannot rule

out that emerging biases reduce performance for some countries, inspection of performance metrics when excluding data from one country at a time during training and testing did not alter main findings (results not shown).

Compared with benchmark ML-based algorithms, LW (Recall & IADL)^{P} suggested higher and more consistent specificity across countries. When using ML-based classifiers such as RF/XGB SMOTE to detect 'probable dementia', our results suggest a lack of consistency in prevalence estimates of neighbouring countries. More dramatically, ML-based 'probable dementia' prevalence exceeded population-based projections and GLM (weighted) introduced a negative association between SHARE-based prevalence estimates and those informed by previous findings (Kodesh, 2019; OECD, 2018; Prince et al., 2015). Contrary, with LW (Recall & IADL)^{P}, prevalence in neighbouring countries was more homogeneous and more similar to previously reported estimates, leading to a positive association between prevalence estimates based on SHARE and previously reported estimates.

Critically, projected prevalence estimates used to assess validity of classifications come with considerable uncertainty stemming from oversimplification (e.g., assuming constant age-specific prevalence), varying operationalisations or lacking knowledge about future developments in medicine or policy (Schwarzinger & Dufouil, 2022). Further, estimates of the OECD reflect projections for 2018 based on data from 2015. However, time lag was low and OECD prevalence estimates were generally higher than those based on self-report physician-diagnosis in SHARE. Still, prevalence may be understated due to healthy volunteer bias, or lacking representativeness in underlying studies, or systematic underdiagnosis in low- and middle-income countries (Cleret de Langavant et al., 2020). Critically, receiving a diagnosis given dementia may depend on the severity of symptoms, lacking access to screening tools, or lacking knowledge of or access to treatment and care (Bond et al., 2005). Moreover, stigma evolving around dementia may result in longer times

until diagnosis, with apparent variation in such stigma across European countries, aligning with the availability of specialised care (Vernooij-Dassen et al., 2005). In any case, self-reporting a dementia diagnosis may amplify such biases (Mullin et al., 2023). It is crucial to interpret our findings acknowledging absence of a gold-standard measure of dementia prevalence and thus discrepancies between the number of people living with dementia with or without a diagnosis to self-report. As such, we refer to underreporting resulting from multiple processes encompassing but not limited to failure to self-report a present diagnosis or absence of a diagnosis despite presence of dementia. In absence of clinically valid assessment of dementia inclusion of external information allowed employing the equipercentile equating approach to SHARE and a consequent exploration of mechanisms leading to differences between data sources (Alzheimer's Association, 2010; Crimmins et al., 2011). As such, our findings suggest that LW (Recall & IADL)^{P} efficiently reduced underreporting defined as discrepancy between previously reported estimates and SHARE-based estimates, uniformly across countries.

Internal validation further suggested that LW (Recall & IADL)^{P} 'probable dementia' was similar to self-reported physician-diagnosis of dementia regarding further domains of health and cognitive function in a subset of countries with available markers. GLM (weighted), and LW adaptations including IADLs overstated age and understated grip strength, both reflecting risk factors of dementia (Bai et al., 2021). Our findings suggest these algorithms classify older, physically more impaired participants irrespective of potentially underlying or absent dementia thus increasing noise and deteriorating fairness with respect to ageism. Inclusion of IADLs, conveying information on worsening physical health but not dementia, specifically, may explain this. Interestingly, IADL inclusion had a positive effect on specificity, possibly by accounting for floor effects in recall measures. Whereas XGB SMOTE and GLM (weighted) 'probable dementia' fit well to self-reported physician-based

dementia diagnosis, prevalence estimates were highly overstated with XGB SMOTE, and biased across countries with GLM (weighted). Our findings further suggest calibration may be negatively affected in algorithms trained with SMOTE (van den Goorbergh et al., 2022). Verbal fluency, depressive symptoms, and numeracy performance were similar in self-report physician-diagnosis of dementia and 'probable dementia' across algorithms (Alzheimer's Association, 2010; Crimmins et al., 2011; Sutin et al., 2019). Critically, depressive symptoms may play a role as early sign or risk factor of dementia, or relate to recall performance (r=-.28, p<.001) and IADL reporting (r=.37, p<.001) irrespective of dementia (Demnitz et al., 2020; Livingston et al., 2020). Orientation to date was not well reflected by LW classifications with a prevalence-based recall cutoff, or RF/XGB SMOTE, potentially due to the categorical operationalisation. In sum, our results support similarity of 'probable dementia' and self-reported physician-diagnosis of dementia in most algorithms.

IV.4.1 Limitations

This study systematically investigated a range of dementia classification algorithms to adjust for underreporting of dementia in a large European ageing survey, using internally derived and externally validated prevalence estimates. Some limitations need to be considered when interpreting our findings. First, we could not train models on CIND classification and thus participants with mild limitations may be misclassified 'without probable dementia'. Second, dementia rates were lower in our sample than in previous studies reducing statistical power to detect 'probable dementia'. Further, a smaller number of participants self-reporting dementia limits generalisability of the validation procedure (Gianattasio et al., 2019). Third, discrepancies in dementia prevalence which we interpreted as potential underreporting may be due to selection bias, or due to diagnoses being based on self-reports, both of which could lead to misclassification following stricter cutoff definitions. Related, models trained on self-reported physician-diagnosis of dementia, which is less

reliable than formal diagnosis, may miss prevalent cases due to reduced statistical power during training (Cigolle et al., 2018). Fourth, participants self-reporting limitations may systematically differ from those not disclosing such information impeding generalisability of our findings (Gamble et al., 2022). Fifth, LW was adapted to a reduced set of indicators and self-reported IADLs reducing discriminatory power. However, recall scores contributed most to LW (20/27 points) and mean Serial 7's scores for LW (Recall & IADL)^{P} 'probable dementia' and self-report physician-diagnosed dementia did not differ significantly in a subset of the data, suggesting limited added value of including Serial 7's for classification. Sixth, a previous study suggested the need for model stratification (Gianattasio et al., 2020). However, class imbalance, sample size and lacking diversity prohibited fairness evaluation of classifications in stratified samples. Seventh, discussed algorithms were applied to cross-sectional data, and may misclassify participants with outlying low performance. Further, LW (Recall & IADL)^{P} cannot differentiate prevalent or incident 'probable dementia'. Eighth, participants in our study were younger (age 60 and older) compared to the Aging, Demographics, and Memory Study (age 70 and older), likely healthier (complete case, community-dwelling) and proxy-ratings were not available, which potentially reduced power to detect cases and yielded more conservative cutoffs (Langa et al., 2005). We thus call for the inclusion of proxy assessments to bolster research relating to cognitive ageing and dementia.

IV.4.2 Conclusions

In absence of clinically validated dementia assessment in observational studies, classification algorithms such as LW can be adapted to cross-national cohort surveys such as SHARE to reduce underreporting of dementia. In this study, LW (Recall & IADL)^{P} identified 'probable dementia' with high validity compared to ML-based classifiers. Many large ageing surveys provide recall items or IADLs (Banks et al., 2021; HCAP Network, 2024). We thus

provide a transparent and transportable classification with a minimal predictor set, based on the pre-established LW algorithm. While ‘probable dementia’ does not reflect a diagnosis, we hope to empower dementia researchers in several ways. First, the present work may facilitate uptake of dementia classification algorithms for research in SHARE. Additionally, we provide knowledge to transport classifications into other applications, since cutoffs used for classification are directly interpretable and adaptable across settings. Second, classifications may be used to inform sampling strategies. Finally, a ‘probable dementia’ indicator may improve statistical power, offering means to assess sensitivity in a multitude of research applications. Future research may offer opportunities to validate our findings with the Harmonized Cognitive Assessment Protocol and compare performance across sister studies of SHARE and HRS (HCAP Network, 2024). This may also yield the potential to investigate algorithm performance in subgroups for fairness evaluations and disparities research.

General Discussion

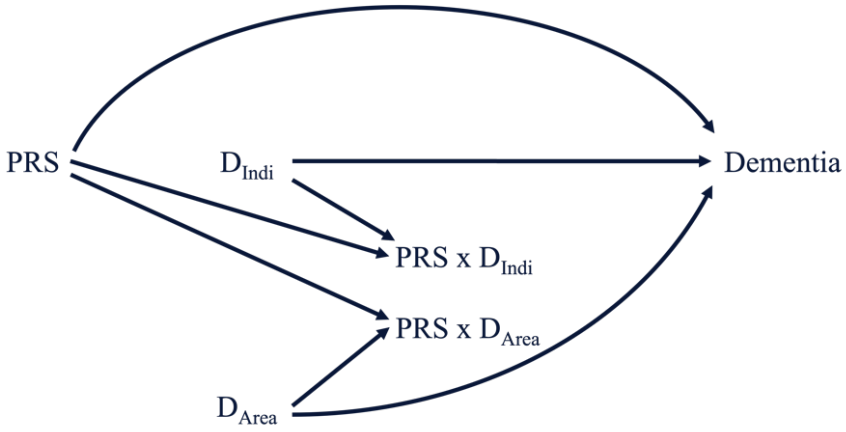
Chapter V – Discussion of Main Findings

V.1 Summary of Findings

This thesis covered three individual studies investigating potentially modifiable risk factors, theoretically derived underlying mechanisms, as well as the application of a classification algorithm for 'probable dementia' to a multi-country ageing survey. The following section summarises the main findings.

Building on the AD Exposome, which provides a framework for interactive paths of exogenous (macrolevel or individual) and endogenous factors associated with the aetiology of dementia, we investigated how living environments characterised by fewer socioeconomic resources may capture individual dementia risk (Finch & Kulminski, 2019). In doing so, we additionally considered individual genetic predisposition and varying degrees of individual-level socioeconomic resources. Our findings suggest that individuals living in areas with fewer socioeconomic resources are at greater risk of developing all-cause dementia, even if their individual socioeconomic resources are considered (Figure 10). Extending on previous findings that suggested lower dementia risk when following a lifestyle characterised by e.g., engagement in physical activity, we could show that even if genetic risk of dementia was intermediate or high, fewer socioeconomic resources in the neighbourhood were associated with increased risk of dementia (Lourida et al., 2019). While there was no interaction of exogenous macrolevel area-level socioeconomic deprivation with endogenous polygenic dementia risk, highest dementia incidence in those that lived in areas with fewer socioeconomic resources and had high genetic risk suggested additivity of detrimental factors towards risk of dementia.

Figure 10 *Directed Acyclic Graph Summarising Main Findings Reported in Chapter II*

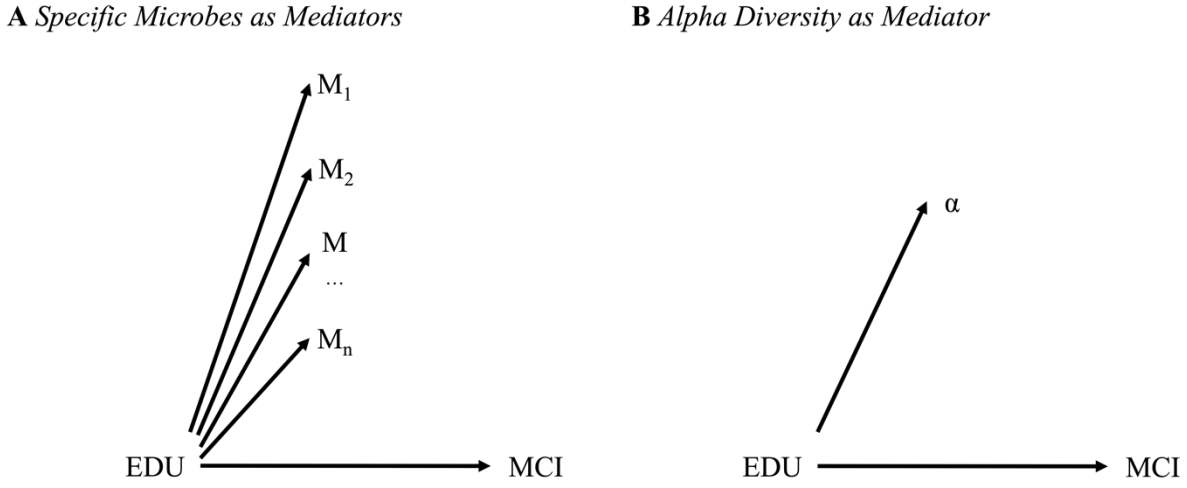


Note. Area- and individual-level socioeconomic deprivation were associated with time to incident all-cause dementia independent of polygenic risk. Note that this summary does not illustrate observed/unobserved confounding considered in the formal analysis and discussion of findings. Interaction terms were depicted according to a proposal of Attia et al. (2022). PRS=Polygenic risk score quantifying risk of dementia; D_{Indi}=Individual-level socioeconomic deprivation; D_{Area}=Area-level socioeconomic deprivation; x=Multiplicative interaction.

Further proceeding to investigate potential mechanisms that link SES to cognitive function in later life, we examined direct and indirect effects of a marker of SES, namely years of education, on MCI risk that would be mediated by metrics of another endogenous characteristic, the gut microbiome. Results were in line with previous reviews on the association of education and cognitive function across the life course in that they suggested higher MCI risk in lower education (Lövdén et al., 2020). However, results reported in Chapter III suggest that gut microbiome features (alpha diversity, composition) do not significantly mediate the association of lower education (i.e., 0-10 years of education) with higher MCI risk (Figure 11). Instead, there was an association of higher education, defined as more than ten years of formal education, with alpha diversity, a metric indicating richness of microbes, as well as an education-related taxonomic signature in line with a previously identified AD-related signature. However, as in Chapter II, the simultaneous expression of

endogenous characteristics and exogenous risk factors were associated with the risk of cognitive impairment in an additive manner.

Figure 11 *Directed Acyclic Graph Summarising Main Findings Reported in Chapter III*

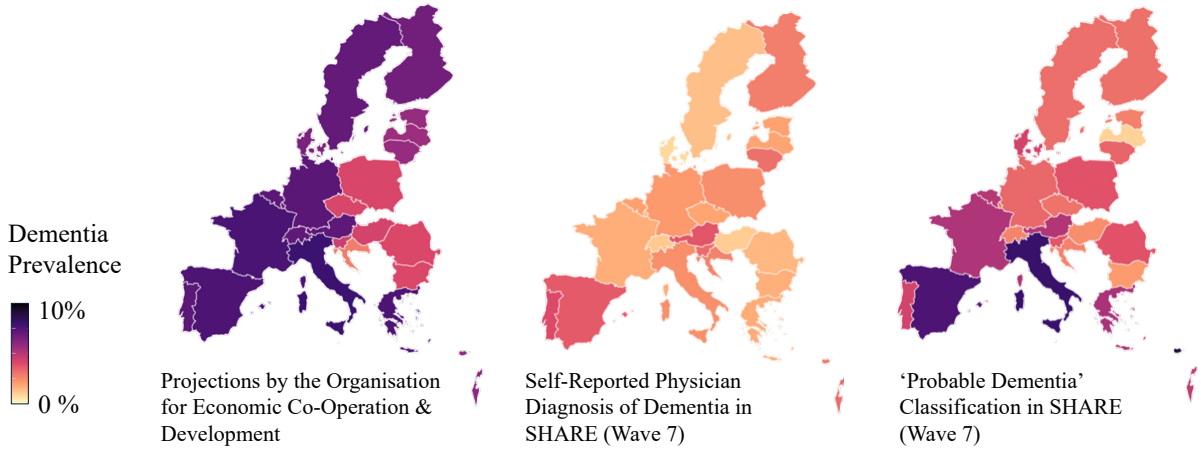


Note. The association of years of education with mild cognitive impairment was neither mediated by A specific microbes nor B metrics of alpha diversity. Note that this summary does not illustrate observed/unobserved confounding considered in the formal analysis and discussion of findings. EDU=Education operationalised as groups of self-reported years of education (0-10, 11-16, 16+); MCI=Mild cognitive impairment; $M_{1...n}$=Abundance of individual microbes; α=Alpha diversity metric.

While the examination of risk factors and biological mechanisms linking individual modifiable risk factors and SDoH to dementia aetiology was the scope of Chapters II and III, delivery of prevention programmes targeting said factors requires accurate and early identification of individuals at-risk. Critically, dementia case ascertainment is resource demanding in cohort surveys and further biased by individual socioeconomic factors determining access to healthcare and consequent likelihood of diagnosis. In Chapter IV we investigated the validity of predictive algorithms based on readily available cognitive test items and reported limitations regarding IADL across multiple European countries. We

adapted an existing ‘probable dementia’ classification algorithm and tested its predictive ability against benchmark ML-based algorithms. We could show that the application of country-level cutoffs for reported IADL and word list recall performance based on externally validated dementia prevalence estimates helped to identify ‘probable dementia’ with high specificity and reasonable sensitivity (Figure 12). This transparent and transportable approach performed similar to ML-based approaches and suggested less country-specific bias. Importantly, those classified with ‘probable dementia’ performed similar to those that self-reported physician-diagnosis of dementia in related domains of health and cognitive function.

Figure 12 *Graph Summarising Main Findings Reported in Chapter IV*



Note. Algorithmic classification of ‘probable dementia’ reduced underreporting defined as discrepancy between prevalence estimates reported by the Organisation for Economic Co-operation & Development (left panel) and prevalence estimates based on self-reported physician-diagnosis of dementia (middle panel) or algorithmically classified ‘probable dementia’ status (right panel) in SHARE. Darker colours indicate higher prevalence. SHARE=Survey of Health, Ageing and Retirement in Europe.

V.2 Contributions to the Research Field

The following section will discuss how the presented findings can be integrated into the current dementia prevention research landscape given key challenges concerning dementia prevention, identified in section I.5.

V.2.1 Dementia (Precision) Prevention in Areas With Fewer Socioeconomic Resources

The first aim of the thesis was to extend on research suggesting health disparities regarding dementia risk along a socioeconomic gradient.

To conceptualise the impact of SDoH and individual-level modifiable factors on dementia risk, the theoretical framework of the AD Exposome allowed us to examine interactions of risk factors embedded in different levels of proximity, i.e., exogenous, and endogenous (Finch & Kulminski, 2019). As such we extended on previous investigations of exogenous macrolevel (area-level) and exogenous individual (individual-level) socioeconomic deprivation by modelling interactions with endogenous genetic predisposition for AD. Our findings underline the existence of socioeconomic patterns of dementia risk, in line with a vast research landscape encompassing epidemiological evidence suggesting constituents of socioeconomic standing to be associated with dementia risk (Livingston et al., 2020; Ranson et al., 2020; Weiss et al., 2020). Causal paths linking SDoH and other SES-related theoretically modifiable risk factors to dementia, are still challenged. However, some researchers have argued that assumptions underlying Mendelian Randomisation, frequently considered as source of causal evidence, are not easily satisfied for complex exposures such as education (Anderson et al., 2024; European Alzheimer's & Dementia Biobank Mendelian Randomization Collaboration et al., 2023; McMartin & Conley, 2020). Generally, broader SDoH, or related factors such as SES or sex/gender, reflect categories in which other modifiable risk factors are clustering. As such, a higher prevalence of risk factors may drive differences in dementia risk observed across e.g., SES strata (Deckers et al., 2019; Geraets &

Leist, 2023; Peters, Booth, et al., 2019). Importantly, ADRD incidence in later life could be driven by SDoH even if the nature of the association with dementia risk was indirect. We identified associations of area-level socioeconomic deprivation with WMH burden, even after controlling for individual-level socioeconomic deprivation. In fact, associations of socioeconomic deprivation with greater WMH burden did apply to both operationalisation strategies and suggest vascular contributions to increased dementia risk. In line with previous findings, this reinforces a perspective on SDoH contributing to dementia risk indirectly, by altering cardiometabolic disease burden (Kivimäki et al., 2019).

We further showed that associations of socioeconomic deprivation with dementia risk did not vary based on the underlying genetic risk. Despite the importance of targeted interventions, e.g., aiming at behaviour change in at-risk populations, our findings (indicating no interaction with polygenic risk) do not support individualised (e.g., genetically informed) approaches to dementia prevention regarding socioeconomic deprivation. This is in line with previous findings relating e.g., to lifestyle, education, or cardiometabolic multimorbidity, albeit some ambiguity stemming from a previous study suggesting no association of lifestyle with dementia risk given less favourable genetic predisposition (Finch & Kulminski, 2019; Licher, Darweesh, et al., 2019; Lourida et al., 2019; Ranson et al., 2020; Tai et al., 2022). In absence of an interaction, reported explorative findings suggested global detriments to brain health in higher area-level socioeconomic deprivation, e.g., reflected in lower grey matter volume, rather than alterations in signature regions of AD such as hippocampal volume. Reported associations may be due to the employed area-level measure capturing detrimental environmental features. As an example, air pollution may confer brain damage not necessarily along a causal path involving a genetic predisposition but via exposure to particulate matter (Peters, Ee, et al., 2019). Further, features (e.g., walkability) may pertain

generalisable associations with dementia in terms of similar association strengths across individuals with different endogenous diatheses (Yu et al., 2023).

Hence, our findings suggest delivery of preventive measures may yield reduced ADRD incidence if applied on a neighbourhood or area-based scale. It could be argued that for efficient improvement of public health, more emphasis should be put on altering living environments for communities in addition to more targeted interventions (Walsh et al., 2023). This is especially important in light of intricacies, invasiveness, agency, and resource intensity entailed in e.g., risk modification through lifestyle changes, as well as an increasing need to accommodate demographic and environmental changes and their economic impact on health care systems worldwide (Wimo et al., 2023). However, Chapter II also calls for a careful evaluation of associations of SDoH with dementia risk considering employed operationalisation strategies, especially when deducing potential working mechanisms, and related prevention efforts. Of note, previous observational findings suggest limited effect modification of the relationship of modifiable social and behavioural risk factors with dementia across e.g., broader categories such as sex/gender, or SES (Geraets & Leist, 2023). Conversely, other studies suggest potentially greater detriments to brain health conferred by depression or hypertension, depending on constructs such as ethnicity or social deprivation (Hofbauer & Rodriguez, 2023; Mukadam et al., 2023). In general, individualised approaches to prevention entail adaptation of contents, delivery, or timing to populations and individuals, and are thus likely to improve dementia risk reduction efforts on a population-level, even in absence of effect modification (Stephen et al., 2021).

Of note, the provision of individualised preventive measures informed by quantifications of an individual's risk, may be more, or less important depending on the level of proximity to the individual, in which operationalisations of SDoH were embedded. So far, we have discussed mainly factors constituting, or behaviours interacting with physical

features of the living environment, i.e., conceptualised as individual exposure rather than structural determinant. For more specific behaviours however, e.g., smoking, previous findings suggest interaction with endogenous factors, such as known AD-related risk alleles (i.e., in *APOE* ε4 carriers), supporting individualised, precision prevention efforts (N. Zhang et al., 2021). Of note, neither area-level nor individual-level socioeconomic deprivation reflect narrowly defined risk behaviours. Still, we argue that individual behaviours may relate to individual- and area-level socioeconomic deprivation and thus reinforce potentials to prevention efforts on a greater scale. In example, one's living environment characterised by fewer socioeconomic resources may alter the likelihood of individuals to engage in certain activities and behaviours, e.g., according to descriptive norms, neighbourhood disamenities, or perceptions thereof (Sevi et al., 2024; Yu et al., 2023). As such, individual risk behaviours, e.g., associated with downstream vascular damage, may be affected by contextual factors limiting opportunity space or one's ability to engage in a physically active lifestyle, thus indirectly impeding cardiovascular health. As another example, even when providing targeted information and assistance in helping individuals to alter their diet, limited socioeconomic resources or availability of adequately priced stores may render compliance almost irrational. Findings from the UK suggest that in the lowest income decile, following the Eatwell Guide would reflect spending over 70% of the disposable income (Marmot, 2020).

Our findings underline the need for a holistic approach to dementia prevention, combining individualised and structural approaches. Critically, higher dementia risk in higher socioeconomic deprivation does not necessarily translate into preventive potential. While microsimulations indicate decreasing dementia prevalence in high income countries, possibly due to increasing access to healthcare, improved management of chronic diseases, or educational opportunities, these models simplify complex living environments and only alter specific components (Brück et al., 2022). To adequately address broad SDoH, social roles

and systemic differences in, as well as options to adhere to a healthy lifestyle need to be considered. Further, emphasis needs to be put on disentangling contributions of co-occurring risk factors, e.g., when investigating racial or SES-related disparities. This is especially important for estimating total contributions of certain risk factors to dementia incidence. However previously applied methods to adjust for clustering of risk factors, may underestimate total contributions and do not consider sequential manifestation times (Welberry et al., 2023). More complex modelling strategies and longitudinal follow-up data are required to estimate population-level absolute contribution of risk factors to dementia incidence, which may then be used to inform policy or public health decision making (Allore et al., 2015).

In sum, our findings transport a positive message for public health research in the domain of dementia. As such, Chapter II suggests features of the environment to capture risk factors associated with dementia risk that may be intervened on, irrespective of the genetically determined predisposition or individual socioeconomic standing. This reinforces the implicit assumption that dementia risk may be systemically, at least partly modifiable by intervening on SDoH. The first aim consisted in elucidating the socioeconomic gradient of health disparities in a more comprehensive setting. Incorporating endogenous and their potential interactions with exogenous characteristics, allowed us to update our hypotheses about working mechanisms potentially underlying observational findings of increased dementia risk in less favourable SDoH strata. More precisely, area-level socioeconomic deprivation suggested to combine an aggregated association of individual-level socioeconomic deprivation with WMH burden alongside an environmental association of global detriments to brain health. These findings suggest that the efficacy of preventive actions will require varying degrees of personalisation, depending on the precisely defined exposure and the related working mechanisms.

V.2.2 Preventing Cognitive Impairment by Targeting Underpinnings of Resilience

The second aim was to identify working mechanisms of the association of an exogenous individual risk factor, education, with the risk of cognitive impairment, to further elucidate potential targets for dementia risk reduction in later life.

To investigate the impact of modifiable social and behavioural risk factors on later life health, the adoption of a life course perspective incorporating resilience as a key determinant of healthy ageing has been proposed previously (Finch & Kulminski, 2019; Kuh et al., 2003; Oosterhuis et al., 2023). However, criteria for positive trials investigating dementia prevention by intervening on modifiable risk factors most often still are operationalised as prolonged times to reach a clinical endpoint or deceleration of cognitive decline. As such, build-up of pathology preceding but ultimately leading to cognitive decline or dementia are rarely targeted. As an example, instead of targeting distal cognitive decline one may measure e.g., alpha diversity or other candidate features, potentially preceding and associated with later life cognitive health. Barriers may be associated with the resource intensity of collecting life course follow-up data, or technical limitations to measurement of and evidence supporting biological antecedents of dementia. Still, interventions are frequently tested in individuals at-risk (e.g., defined by age above 65 or presence of genetic risk markers/exposure to risk factors such as lower education), which likely have been exposed to less favourable living conditions or risk behaviours and thus accumulated disadvantage over decades.

Such accumulated exposure to risk factors resulting in detriments to health in later life is implied by the high intercorrelation of different age-related diseases. Age-related disease clusters have e.g., been identified by application of machine-learning methods to large-scale electronic health records and could be examined to identify communalities and shared aetiological processes (Kuan et al., 2021). Such observations underline the necessity to

understand how and which factors collectively drive the pace of ageing (Ferrucci et al., 2020). Conceptualising the ageing process contrasting accumulation of damage and potential to resilience, with accelerated ageing denoting greater risk of dementia and other age-related non-communicable diseases, may help to inform and update theories on how modifiable social and behavioural risk factors interfere with health in later life.

As a recapitulation, our findings presented in Chapter II extended on previous literature suggesting increased dementia risk in those living with fewer socioeconomic resources, likely due to downstream vascular damage which could be interpreted as an increased pace of ageing. Unfortunately, we did not have retrospective data on socioeconomic deprivation, which is a common caveat in observational or registry-based studies, and hence we could not differentiate participants' time exposed to more, or less favourable area-level or individual-level socioeconomic characteristics. However, it would be less plausible to assume WMH burden had led to varying area-level socioeconomic deprivation. It would rather be likely, that pathological alterations in brain health resulted from continued exposure to detrimental environmental characteristics.

Given wide agreement to prioritise intervention in later life and in individuals at greatest risk, working mechanisms of risk propagation still at play in later life require further investigation (Livingston & Costafreda, 2023). In that, assuming the pace of ageing to alter times to dementia diagnosis or other age-related diseases, windows of opportunity for intervention need to be identified. Moreover, given a multifactorial process determining dementia risk, trials investigating single-domain interventions targeting modifiable social and behavioural risk factors will likely be of limited efficacy, especially with ADRD as clinical endpoint (WHO, 2022a). This is reinforced by findings from recent dementia prevention trials targeting e.g., diet, hearing loss, or physical activity in later life (Barnes et al., 2023; Buckinx & Aubertin-Leheudre, 2021; Ciria et al., 2023; Dawes & Völter, 2023; Yeo et al.,

2023). More holistic, multi-domain interventions were more effective in altering cognitive function, albeit low specificity and more pronounced associations with secondary outcomes such as quality of life, e.g., decreasing depressive symptoms (Hafdi et al., 2021; Ngandu et al., 2015; Wittmann et al., 2024). Mechanistically, one may assume prevention efforts targeting modifiable risk factors to exert their impact via (a) prevention of downstream vascular damage or neurodegeneration, discussed in Chapter II, (b) slowing of or recovery from pre-existing vascular damage or neurodegeneration and related clinical symptoms, or (c) reinforcement and build-up of resilience to vascular damage or neurodegeneration.

Regarding (a), the likelihood of or the time until dementia may depend on exposure to modifiable social and behavioural risk factors, themselves conferring vascular damage, or exacerbating neurodegeneration (Bir et al., 2021; Cermakova et al., 2015). As such, primary prevention via intervening on e.g., nutrition or lifestyle, would lead to reduced dementia risk if delivered prior to the manifestation of vascular damage or by reducing its extent early in the trajectory. Notably, this assumed mechanism is implied by increased rates of AD in hypertension, suggesting vascular contribution to AD risk, and supported by clinical trials suggesting reduced dementia risk when providing antihypertensive medication (Adesuyan et al., 2023; J. R. Smith et al., 2023; Thunell et al., 2021). Critically, under (a), proving efficacy of early intervention on modifiable risk factors may require follow-up starting in midlife since dementia-related pathological changes may occur as early as two decades prior to clinical diagnosis (Zetterberg & Bendlin, 2021). Our findings presented in Chapter III support orientation towards primary prevention with intervention in early life. We found associations of gut microbiome characteristics with years of formal education, as well as associations of education with MCI risk, but no significant mediation. Identified overlap in gut microbiome signatures of AD or MCI, and lower (0-10 years) education suggest education to relate to

early life gut microbiome maturation and development, with limited potential to modifying MCI risk by intervention on gut microbiome diversity in later life.

(b) and (c), i.e., reflecting recovery or resilience, aim at secondary or tertiary prevention via intervention on modifiable risk factors. Our findings presented in Chapters II and III suggest that intervention on environmental features or education may reduce the risk of cognitive impairment – if delivered prior to damage – but do not necessarily extend to suggested recovery or effective reinforcement of resilience – if delivered in later life. Critically, factors such as education may reinforce resilience only when translating into e.g., higher income or socioeconomic position, which may require policy intervention (Lövdén et al., 2020; Seblova et al., 2021). Given the findings presented in Chapter II one may argue that area-level socioeconomic deprivation may proxy such manifest socioeconomic position.

(b), specifically thrives on the idea of a feedback loop between risk factors and antecedents to dementia risk, with potential recovery along a causal path. This presupposition is challenged empirically, e.g., reflected in limited efficacy of targeting modifiable dementia risk factors in later life (Hafdi et al., 2021). In line with previous findings, results presented in Chapter III did not provide evidence for a biological mechanism linking years of education to cognitive impairment (i.e., via gut microbiome composition or diversity). More generally however, intervention in later life may still extend to reducing risk of more sever IADL limitations, e.g., by maintaining global cognition (Ngandu et al., 2015; Rosenberg et al., 2018). While potential working mechanisms are less intuitive compared to (a), one could reason that e.g., engaging in physical activity or caloric restriction in later life may improve neuroplasticity or neurogenesis, in turn lowering dementia risk in line with (b) or (c), respectively (Bettio et al., 2017; Bieri et al., 2023; Cotman et al., 2007; Piscopo et al., 2022; Witte et al., 2009). This would implicate potential for secondary or tertiary prevention, by intervention on modifiable risk factors.

Likely, risk reduction efforts targeting SDoH, or individual-level modifiable risk factors require long periods to detect potentially conferred beneficial effects. Targeting primary prevention, lifestyles may have manifested in earlier life and continued to contribute to dementia risk for longer periods throughout an individual's life history (Lövdén et al., 2020; Seblova et al., 2021). Critically, feasibility constraints limit follow-up periods of (multi-domain) trials targeting modifiable risk factors. As such, research as in Chapter III, targeting late-life risk factors exacerbating, adding to, or mediating detrimental and beneficial associations of SDoH and individual-level modifiable risk factors with i.e., MCI risk, may be more pragmatic and relevant for interventions in later life, at-risk populations (Livingston & Costafreda, 2023). Our study showcased how working mechanisms such as mediation may be investigated in cross-sectional data, by application of counterfactual imputation (Valeri & VanderWeele, 2013; VanderWeele, 2014).

Results reported in Chapter III suggested a pattern of microbial taxa that were less abundant in lower education, and previously identified less abundant in MCI and dementia. Further, we found that counterfactual intervention on gut microbiome diversity eliminated ~20% of the association of higher education with MCI, in a cross-sectional study. While our findings suggest only a small portion of the total effect of years of education not being of a direct nature, one may argue that biological correlates of increased or reduced dementia risk are associated with individual modifiable risk factors and SDoH. While it is less clear, if a potential reduction in the total effect (i.e., suggested by PE) points to an additive, or interactive (resilience) mechanism, our findings underline the importance of considering SDoH and related individual risk factors in studies targeting biological pathways associated with AD pathogenesis.

In Chapters II and III, we were able to investigate potential interaction of exogenous risk factors, with endogenous factors relating to later life health and the risk of cognitive

impairment. Both results further elucidated the potential role of SDoH and related individual modifiable risk factors for the risk of adverse outcomes such as MCI and dementia, pointing to early intervention and (a). Viewing dementia as an age-related disease belonging to a cluster of other age- and lifestyle-related diseases is supported by the fact that despite the application of increasingly complex prediction models and utilising deeply-phenotyped data, age remains the most important risk factor and hence predictor of dementia (Marinescu et al., 2021). As such, age likely reflects an exposure history relating to increased dementia risk, e.g., via neurodegeneration, vascular damage or resilience to such (Ferrucci et al., 2020).

In sum, the investigation of working mechanisms of the association of education with cognitive function in later life, suggested that there are synergies in taxonomic signatures of lower education and those identified in AD and MCI. In absence of significant mediation of the association of years of education with cognitive impairment, lessons learned may be subsumed in (a) the importance of early delivery of prevention to effectively shape health trajectories throughout the life course and (b)/(c) the need to further elucidate biological mechanisms underlying the associations of SDoH and individual-level modifiable risk factors with the risk of cognitive impairment in later life.

V.2.3 Upscaling 'Probable Dementia' Classification

The third aim was to facilitate classification of 'probable dementia' based on a minimal predictor set, in cohorts without linkage to health records, or clinical case ascertainment.

Recent prevention research focussing on e.g., hearing aids, or multi-domain interventions, suggests protective effects in high risk groups rather than the whole population (F. R. Lin et al., 2023; Livingston & Costafreda, 2023). Consequently, targeting of individual risk factors may not aid the overall population or in other terms, not every individual may equally benefit from a specific intervention. Associations may not be identifiable in

convenience samples or low-risk populations, but only in those with diatheses (e.g., *APOE* ε4 carriers) or those with a less favourable lifestyle (e.g., sedentary lifestyle) to begin with. This reinforces the need to identify at-risk populations and to deliver targeted approaches to dementia prevention.

The need for risk stratification is further reflected in the continued efforts of the scientific community to develop easily accessible, transportable and transparent algorithms to quantify dementia risk (e.g., CAIDE, LIBRA), or ascribe 'probable dementia' status based on cognitive and function data (Crimmins et al., 2011; Deckers et al., 2019; Herzog & Wallace, 1997; Hurd et al., 2013; Pekkala et al., 2017). To arrive at generalisable conclusions about the contribution of modifiable risk factors of interest to dementia risk, one further needs to consider that roles and SDoH reflecting broader categories may vary in their association with dementia risk globally. As an example, SES or obesity, associated with increased dementia risk in high income countries, may not or even inversely relate to dementia risk in low- and middle-income countries (Daran et al., 2023). With risk prediction as a crucial task to stratify dementia prevention efforts, our findings presented in Chapter IV showcased strong variation in sensitivity and specificity of examined 'probable dementia' classification algorithms across European countries. While some proposed dementia classification algorithms consider e.g., demographic characteristics, our results suggest cross-country variation in related social roles, reporting behaviour or task difficulty to yield differences in predictive accuracy, even if classification is based on standardised cognitive tests and IADL limitations, and if cutoffs used for classification are stratified by country (Crimmins et al., 2011; Herzog & Wallace, 1997; Hurd et al., 2013; Q. Wu et al., 2013).

One potential application for dementia classification algorithms is oversampling of at-risk individuals for large cohort surveys. Our findings add to existing literature showcasing that baseline differences in sample structure may not be fully addressed by inclusion of

sociodemographic characteristics and that fair classification across countries is at stake (Gianattasio et al., 2020). Findings of Chapter IV instead support classification based on objective cognitive test scores and functional limitations to achieve high specificity across samples from different populations.

Such classification is in line with a clinical approach to testing individuals rather than prediction of risk based on relative measures of association strength, e.g., for macrolevel socioeconomic standing. However, broader SDoH categories, e.g., reflected in educational profiles or gender roles may bias seemingly objective measures of physical limitations (i.e., IADLs) that were utilised in our adapted classification algorithm (Bishop et al., 2016). In this regard other items, assessing e.g., mobility limitations, such as climbing a flight of stairs are related to IADL, but operationalise more serious disablement, i.e., physical limitations, while not being as dependent on environmental factors, access to assistive technology or adopted social roles (Bishop et al., 2016). Moreover, such items are more sensitive to early deterioration in physical limitations and resulting autarchy, thus being better suited for identifying at-risk populations for primary prevention early in the disease trajectory and reflecting quality of life as a patient-centred outcome.

Still, we could show that inclusion of functional limitations increases specificity, underscoring that distribution-based thresholds for cognitive function may in part capture normative ageing and that further information is necessary to approach classification of a ‘probable dementia’ phenotype. This is in line with conceptualising normative ageing with regard to maintaining dual functionality and underscores the utility of evaluating cognitive health with more patient-centred outcomes that additionally capture meaningful detriments to quality of later life (Ferraro et al., 2023).

While ‘probable dementia’ classification as presented in Chapter IV is not reflecting clinical diagnosis, application of such algorithms in longitudinal cohort surveys may allow

researchers to classify participants in line with diagnostic criteria proposed by the National Institute on Aging – Alzheimer's Association (NIA-AA) workgroups on diagnostic guidelines for AD (McKhann et al., 2011). As such, insidious onset of cognitive symptoms as well as a history of worsening cognitive function may be detected when applied longitudinally. Of note, exhaustive testing, i.e., across multiple cognitive domains, over time is limited by survey designs and available instruments. Such real-world limitations owing to resource intensity of dementia case ascertainment in representative cohort surveys hamper the opportunity to formally assess validity of existing classification algorithms in absence of a 'ground truth' dementia status. Still, our findings presented in Chapter IV show that utilising a minimal predictor set with a threshold-based classification allows a reduction of eligibility criteria (i.e., availability of data over time) and hence to classify the complete sample, opposed to more sophisticated algorithms drawing on individual follow-up over a large number of indicators to apply prediction algorithms, or unsupervised clustering methods (Gharbi-Meliani et al., 2023).

In sum, the third aim examining classification of 'probable dementia' led to an adapted, previously established classification algorithm that is transparent and transportable. The presented algorithm may swiftly be adapted to other data, granted availability of a minimal set of predictors and availability of snapshots of cognitive performance and functional limitations for individuals. Applications may extend to inform oversampling of at-risk individuals or testing sensitivity of risk factor research, by reducing class imbalance with 'probable' versus self-reported dementia status. Furthermore, the classification presented, based on cutoffs for interpretable scales allows prompt assessment of fairness and performance variation in sample. Given utmost importance of accurate risk prediction, these findings enable research to account for underreporting or underdiagnosis of dementia in

observational cohorts and to maximise utility of cognitive function measures complying with real world limitations relating to case ascertainment and data collection.

V.3 General Limitations

Despite thorough design of the individual studies constituting Chapters II to IV, some limitations need to be considered interpreting the collection of findings presented.

First, low response rates, limited representativeness, and selective attrition (e.g., drop-out due to worsening cognitive function and related institutionalisation) may affect generalisability of findings in terms of a healthy volunteer bias. Of note, this limitation is ubiquitous to observational research, and potential biases were thoroughly investigated with a broad range of sensitivity analyses.

Second, investigation of biological mechanisms, i.e., with volumetric measures in MRI data, polygenic risk scores, or metrics of gut microbiome diversity and taxonomic signatures, may inflate type II errors given likely involvement of specific genes, molecules, or biomarkers for neurodegeneration.

Third, reverse causation cannot be ruled out given early pathophysiological alterations of dementia as well as early-life or innate abilities likely not fully captured by available follow-up. Hence, the associative nature of reported findings needs to be considered, implying the need to further test efficacy of prevention efforts targeting discussed risk factors and mechanisms.

Fourth, sensitivity of dementia/MCI case ascertainment may vary e.g., depending on person-level characteristics, available resources, access to healthcare, or mode (e.g., self-reported/health records). Cross-sectional dementia/MCI ascertainment further violates clinical diagnostic criteria presupposing deterioration of cognitive function over time, which may lead to differential misclassification.

Fifth, despite an increasing understanding of the complex intertwining of somatic conditions (e.g., cardiovascular disease burden), individual risk factors/SDoH and aetiologies of dementia subtypes, the individual studies focussed on ADRD and MCI. In absence of e.g., biomarker-informed differential diagnosis, interpretation of findings is not necessarily generalisable to individuals living with a specific subtype of dementia.

Sixth, limited diversity precluded formal assessment of fairness in prediction models and the generalisability of findings across subgroups characterised by race/ethnicity. Critically, distributions of antecedents to dementia risk as well as varying association strength coupled with potential structural racism and related access to healthcare may affect reported associations.

V.4 General Strengths

This thesis entails three individual studies jointly providing strong evidence for socioeconomically patterned dementia risk profiles, elucidating potential working mechanisms linking modifiable risk factors and risk of cognitive impairment, and potentials to identify 'probable dementia' across countries. This section will point to key strengths of the conducted studies.

Individual studies were drawing on large data sets with comprehensive assessments of risk factors, and cognitive function, including deeply-phenotyped cohorts characterised by linkages to health records, death registries, clinical assessments, and biological markers involving MRI scans, genetic and microbiome data. This allowed a holistic approach to investigating the role of modifiable social and behavioural factors for dementia risk on three levels of proximity, as well as potential interactions denoting targets for personalised prevention efforts. Moreover, reported findings are based on application of sophisticated modelling paradigms to test mechanistic hypotheses and properties of predictive models entailing a wide array of risk factors. As such, a triangulation of evidence was performed,

informed by methodological approaches from fields of epidemiology, medicine, computer science as well as causal inference.

More specifically, Chapter II examined effect modification across a comprehensive set of individual- and area-level socioeconomic characteristics, additionally considering individual genetic predisposition in over 200 000 individuals. ADRD was ascertained in clinical records and death registry data, reflecting gold standard diagnostic and near-complete follow-up. Moreover, exploration of imaging markers in a large subset informed discussion of potential underlying working mechanisms.

Chapter III added to an in-depth investigation of potential mechanisms linking education, to MCI via potentially modifiable, biological mechanisms and applied two sets of mediation analysis allowing a distinction of statistically, as well as biologically defined mediational mechanisms. Again, clinically informed MCI ascertainment established internal validity of analyses.

Chapter IV adapted and validated a previously established algorithm to classify 'probable dementia' in a multi-country ageing survey with over 50 000 individuals, drawing on externally validated country-level prevalence estimates. Performance to reducing underreporting of dementia was compared against a broad set of well-performing ML-based models with algorithms to improve class imbalance. Comprehensiveness of the large panel survey allowed extensive cross-validation and enriched inspection of internal validity regarding the identified phenotype.

In sum, this thesis responded to major challenges in the field of dementia prevention research, adds to the nomological network of interactions and working mechanisms underlying contributions of SDoH and theoretically modifiable social and behavioural risk factors to individual dementia risk, provides actionable means to classifying cognitive phenotypes in line with 'probable dementia' in observational data, and delivers important

leads to future research projects with respect to modifiable risk factors and their role for individual dementia risk and related precision prevention efforts, which will be discussed in the following sections.

V.5 Reflection, Outlook, and Avenues to Future Research

Findings presented in Chapters II to IV respond to key challenges in the field of dementia prevention. The following sections present outlooks and avenues for future research, based on reported findings and focussing on deepening our understanding of modifiable risk factors and their life course contribution to dementia risk.

V.5.1 Outlook Chapter II:

Findings presented in Chapter II suggest delving deeper into examining dementia risk patterns characterised by individual- and area-level socioeconomic deprivation and are based on explorative findings.

First, the association of individual-level socioeconomic deprivation with incident all-cause dementia was more pronounced than for area-level socioeconomic deprivation. Explorative findings however suggested detriments to brain health to be conferred via greater burden of WMH, with both area-level and individual-level socioeconomic deprivation. Previous findings suggested socioeconomically patterned dementia risk in part due to increased exposure to and prevalence of less healthy lifestyles, and related cardiovascular diseases (X. Chen et al., 2022; Deckers et al., 2019; Geraets & Leist, 2023; Yu et al., 2023). As such, our findings may point to both environmental/structural and individual-level socioeconomic resources to be associated with dementia risk, e.g., along a cardiovascular pathway involving cerebrovascular insult. Drawing on the rich UK Biobank, a next step would be to investigate potential mediation of identified associations of area-level and individual-level socioeconomic deprivation with dementia by cardiovascular disease burden. Related, follow-up imaging data is becoming increasingly available for a subset of the UK

Biobank, which would allow to investigate mediating paths not only with respect to dementia as a clinical endpoint, but also revisiting assumed directions of reported explorative findings. Importantly, when refining analysis based on the UK Biobank, one may use sampling weights to overcome a frequently discussed limitation of the UK Biobank data, the low response rate, presented in a recent study (van Alten et al., 2022).

Second, a recent twin study suggested that genetic predisposition may underlie the association of cardiometabolic multimorbidity with dementia (Dove et al., 2023). Interestingly, we also report a limited role of socioeconomic deprivation for the time to incident all-cause dementia in low polygenic risk. As such, a more focussed interaction analysis, e.g., with *APOE* ε4 status (e.g., involved in cholesterol metabolism and Aβ clearance), or other fine grained genetic markers relating to the association of cardiometabolic conditions and dementia may further elucidate the potential to personalised approaches for ADRD prevention by intervention on socioeconomic factors (Martins et al., 2006).

Lastly, our findings regarding area-level socioeconomic deprivation and dementia risk motivate future studies to examine potential causal paths as well as involved exposures captured with area-level opposed to individual-level indices of socioeconomic deprivation. In that, differential associations identified across operationalisations of socioeconomic characteristics suggest the existence of geographically patterned, or community-specific dementia risk. Extending on previously discussed environmental exposures such as air pollution or structural barriers to engagement in a healthy lifestyle, one may investigate the role of descriptive norms for health behaviours (X. Chen et al., 2022; Peters, Ee, et al., 2019). Such norms could be assessed in qualitative studies or via proxy measures, e.g., how likely are participants engaging in health-related activities depending on their residency/community, or analytical approaches such as network analyses.

V.5.2 Outlook Chapter III:

Findings reported in Chapter III suggest education-related taxonomic signatures in line with an AD-related phenotype, but no significant mediation by gut microbiome diversity or individual taxa. One potential explanation may be that lower education denotes higher MCI risk and is thus more prevalent in the MCI subgroup. Consequently, identified synergies in taxonomic signatures may not indicate pathophysiological alterations but be due to neglecting systematically different levels of education in AD/MCI groups in previous studies (e.g., L. Chen et al., 2022). In line with outlooks presented for Chapter II, longitudinal assessments of the gut microbiome composition could be leveraged to ascertain temporality. As such, future studies may examine differential abundance of AD/MCI-related microbes across education groups over time. This will allow revisiting the identified communality investigating if evolutions align with education, and if those are in turn associated with decreasing cognitive function. Prospective longitudinal designs to observe age-related changes in gut microbiome signatures (in older adults at risk of developing dementia), targeting mediational mechanism, are needed and have been proposed recently (Koblinsky et al., 2023; Phillips et al., 2022).

Future studies may further measure lifestyle-related behaviours at the intersection of SES, education, and health-related outcomes more comprehensively, e.g., with Food Frequency Questionnaires to test assumed causal paths involving specific behaviours likely mediating associations of SES-related modifiable risk factors with endogenous factors. In that, researchers may take advantage of technological advancements in the analysis of omics (i.e., metagenomics, metabolomics, metaproteomics, metatranscriptomics) and related availability of meta-omic datasets (Heintz-Buschart & Wilmes, 2018). Such data allows more granular investigation of assumed mechanisms and to e.g., dissect functional pathways from gut microbiome metrics to health-related phenotypes (Heintz-Buschart & Wilmes, 2018).

Research targeting biological pathways is necessary to explain why observed education-related gut microbiome signatures did not translate into altered risk of impaired cognition in Chapter III and thus to assess potential for prevention by intervention targeting the gut microbiome.

V.5.3 Outlook Chapter IV:

With respect to findings reported in Chapter IV, recent developments provide potential to updating and further testing the adapted dementia classification algorithm. Cross validation of provided algorithms is yet to be conducted and given collaborative proceedings to curate and harmonise data on cognitive function in later life globally, a next step would be to replicate our study in such harmonised data to be available for SHARE (Gross et al., 2023; Kobayashi et al., 2024; Langa et al., 2019). Moreover, the now available SHARE wave eight encompasses a more extensive battery of cognitive tests which may be leveraged to increase sensitivity of algorithmic classifications and a better differentiation of normative cognitive decline from dementia-related phenotype. In doing so, application of a normative sampling approach which has been proposed recently, could be utilised to increase internal validity of the threshold derivation procedure (Manly et al., 2022). Furthermore, wave eight may be leveraged to assess re-test reliability of classifications at wave seven.

Moreover, differentiating underreporting from underdiagnosis is of utmost interest. In example, we operationalised underreporting as discrepancy between externally validated and in-sample self-reported dementia prevalence estimates. We could thus not distinguish underreporting, i.e., participants not reporting a present diagnosis, from underdiagnosis, i.e., participants not having received a diagnosis despite cognitive impairment. While harmonisation of data provides the opportunity to validate algorithms in extraneous data, replication in data sources encompassing medical claims or consensus diagnosis of experts would allow to test performance of the dementia classification algorithm given near-

complete, clinically validated dementia ascertainment. Alternatively, one may apply the provided algorithms in data sources that entail neurological markers relating to hallmark indicators of e.g., AD, which would allow to investigate specificity of classifications further.

Since we proposed application of dementia classification algorithms to inform e.g., sampling, evaluation of fairness regarding apparent racial or gender biases resulting from pronounced data health gaps are still outstanding. Critically, biases in data bases used to train algorithms, e.g., driven by a gender-health gap or lacking diversity are likely reflected in predictive algorithms and further problematise their implementation in diagnostic and care settings, but also in research (Gianattasio et al., 2020). In pursuit of precision prevention, one-size-fits-all cutoffs for diagnosis may hamper accurate risk stratification. As such, a further avenue for research may be to adopt the country-level thresholding procedure to meaningful subgroups and to revisit classification performance and bias in nationally representative data.

V.5.4 Further Outlooks – Updating the Scope

Investigating theoretically modifiable social and behavioural risk factors was at the core of the thesis. Conducting the individual studies however, the scope grew broader, and additional challenges were identified, which were not restricted to a priori formulated research questions. These avenues are related to questions of causality, working mechanisms and risk factors, but extend to quantification of the contribution of risk factors to dementia incidence on a population level, and to modelling the accumulation of risk with respect to a life course perspective. This section aims to point to further avenues emerging from Chapters II to IV but exceeding the initial scope of the thesis.

Findings presented in Chapters II and III were based on SDoH and individual modifiable risk factors such as education and socioeconomic deprivation, which are related to SES inherently. While previous findings suggest that e.g., education manifests in early life (in

part transmitted at birth) and may contribute to build up of resilience or exposure to risk throughout the life course, there is ambiguity with regard to the efficacy of education-related interventions, which may be explained by timing (Gutierrez et al., 2023; Lövdén et al., 2020; Marden et al., 2017; Rogers et al., 2009; Seblova et al., 2021). As such, windows of opportunity for potential interventions are of question and relevance of exposures at a specific time or during a specified period require further attention. As an example, within a life course approach to risk propagation, cognitive function batteries (i.e., used to classify 'probable dementia' in Chapter IV) capture current performance but fail to uncover processes (being detrimental/beneficial) and developments that have led to the levels measured in later life. Critically, some exposures (e.g., income/wealth) may vary over the life course and thus, exposure histories need to be modelled explicitly to examine risk propagation and relevance of risk factors over time.

Recently proposed models allow to specify the accumulation of exposure to risk factors with functional regression, thereby accounting for conventional, discrete measurement protocols, and to probabilistically test competing hypothesis of life course models of risk, i.e., accumulation versus critical/sensitive periods (Bodelet et al., 2024; Chumbley et al., 2021; Madathil et al., 2018). Such models could be used to test potential accumulation of disadvantage, conferred e.g., by time-varying exposure to depressive symptoms (Klee, Bodelet, et al., Manuscript in preparation).

Moreover, with recent advancements in distinguishing chronological age from the biological underpinnings of an ageing process, build-up of pathophysiological alterations may be observed at an earlier stage and in turn, the detection of windows of opportunity for the provision of respective preventive measures may be facilitated. Biological underpinnings of dementia are acknowledged e.g., in the research framework provided by the NIA-AA, illustrating opportunities to investigate the impact of risk factors on biologically defined AD,

with respect to amyloid deposition, pathologic tau, and neurodegeneration (Jack et al., 2018). With this biological definition proposed, testable theoretical models may be derived to investigate early pathological markers and distinct pathways, and as such intervention targets. While beyond the scope of this thesis, the research framework is translated into recommendations for diagnosis and staging at the time of writing (Alzheimer's Association Workgroup, 2023; The Lancet Neurology, 2024). In line with the updated 2018 NIA-AA Research Framework, future research may be able to examine potentially causal paths, e.g., by ascertaining imaging or fluid biomarkers (Aβ deposition, pathologic tau, neurodegeneration), to identify subgroups with specific pathophysiological alterations (Jack et al., 2018, 2024). As such, working mechanisms, relating modifiable risk factors to build-up of pathology could be investigated early in the disease trajectory, prior to later cognitive impairment likely due to mixed neuropathology, or even distinct pathways e.g., associated with depression, or fatigues. As an example, ascertaining blood-based biomarkers of neurodegeneration in adolescents with less favourable lifestyles, or areas with fewer socioeconomic resources (in line with geographical patterning of dementia risk alluded to in the outlook for Chapter II), could be leveraged to examine associations of SDoH and individual modifiable risk factors with life course risk. Thus, accumulation of risk may be investigated mechanistically, albeit current limitations in individual risk prediction based on fluid biomarkers (Jack et al., 2024). Using different biological operationalisations, this was aimed at in a previous study investigating allostatic load, a quantification of stress-induced biological risk, in area-level socioeconomic deprivation (Gustafsson et al., 2014; Szanton et al., 2005).

Given findings presented e.g., in Chapter II, providing information for risk stratification and profiling, dementia prevention efforts are inherently tied to efficient resource allocation which in addition to timing, should be informed by quantification of

population-level absolute contribution of risk factors to ADRD incidence. However, modifiable risk factors relating to dementia risk are rarely occurring in isolation and within a life course model of risk, estimating their total contribution to dementia incidence requires to account for the heterogeneity in co-occurring risk factors as well as their sequential build up over time. Methods such as the longitudinal extension of the average attributable fraction (LE-AAF) offer means to derive population-level estimates for attributable fractions due to an individual or collection of SDoH and related modifiable risk factors, that are valid across emerging sequences and account for times at risk (Allore et al., 2016; Murphy et al., 2012). LE-AAF may be applied to modelling PAF of e.g., cardiovascular disease combinations, that frequently co-occur and denote higher dementia risk (Dove et al., 2023; Guo et al., 2024; Klee, Markwardt, et al., Manuscript in preparation).

Lastly, Chapter III provided insights in how counterfactual imputation can be leveraged for causal mediation analysis in observational data (Valeri & VanderWeele, 2013; VanderWeele, 2014). Related to the question of windows of opportunities to intervene on modifiable risk factors, and absolute contributions to population-level dementia incidence, limitations inherent to observational, population-representative studies measuring modifiable risk factors and cognitive function over larger time spans, e.g., communality/co-occurrence of risk factors, appeal to benefits of novel analytical paradigms. As an example, the target trial framework has been proposed, which entails a protocol for comparative effectiveness research in observational data, as an alternative to testing causal effects when RCTs with sufficient sample sizes, or follow-up periods are not feasible (Hernán & Robins, 2016; Hernán & Taubman, 2008). Drawing on the potential-outcomes framework, existing repositories may be revisited and analysed with research goals formulated in line with those of RCTs but based on observational data.

V.6 General Conclusion

In summary, presented studies confirm pre-established risk factors of MCI and dementia, and extend on previous findings by suggesting that environmental and individual modifiable risk factors as well as SDoH are associated with brain health, reflect key modifiers of endogenous characteristics associated with a phenotype of cognitive impairment, and constitute clusters of further risk factors potentially contributing to neurodegeneration, vascular damage, and impaired cognition.

Findings presented in Chapter II suggest considering more thoroughly shaping of the living environments in which people live, work, and age as dementia prevention strategy. This may entail limiting exposure to risk factors or increasing exposure to protective or resilience factors, by reducing barriers to e.g., engaging in more favourable lifestyles. Structural efforts to dementia prevention, e.g., applied to areas with fewer socioeconomic resources, are likely beneficial in addition to individualised efforts to precision prevention. Of note, further research is needed to identify drivers of the reported associations with dementia over the life course, reflecting potential intervention targets. Against this background, presented findings reinforce the notion of modifiability of individual dementia risk, irrespective of the presence of more, or less favourable endogenous characteristics.

Moreover, findings presented in Chapters II and III emphasise the need to further explore endogenous intermediate factors potentially contributing to and thus reflecting an entry point to managing individual dementia risk, in addition to environmental and in part structural determinants. In that, especially findings reported in Chapter III underline the importance of delivering preventive measures early in life but point to potentials in leveraging highly granular data to explore windows of opportunity and translational research applications.

Finally, while the ‘probable dementia’ classification algorithm presented in Chapter IV, is by no means intended to reflect a diagnosis, our findings suggest validity of the approach to identifying participants that self-report living with dementia as well as those with a similar health and function profile. Given the growing importance of identifying at-risk populations and individuals, such easily transportable, adaptable, and transparent approaches to risk stratification may offer manifold enrichment of endeavours relating to both research and prevention applications, e.g., by increasing statistical power, or informing sampling schemes.

In conclusion, reported results emphasise the need to nuancing and contextualising targeted or precision prevention efforts. When examining mechanisms of SDoH and related individual modifiable risk factors potentially conferring dementia risk, our findings strongly emphasise the necessity to critically reflect on operationalisation and analytical strategies, at best aiming at a triangulation of evidence with different model specifications, measures, and methods, in diverse cohorts. It is of utmost importance to draw on increasingly available deeply-phenotyped data and increasingly accessible sophisticated analytical paradigms to examine interactive mechanisms linking SDoH and individual modifiable social and behavioural risk factors with dementia risk on a societal, behavioural, genetic, and molecular level of abstraction. Only then, timely, and fair identification of individuals and populations at risk may be followed by the provision of support aiding the collective of stakeholders to respond to the challenges imposed by the dementia syndrome, globally.

References

Abolhasani, E., Hachinski, V., Ghazaleh, N., Azarpazhooh, M. R., Mokhber, N., & Martin, J. (2023). Air Pollution and Incidence of Dementia. *Neurology*, *100*(2), e242–e254. https://doi.org/10.1212/WNL.0000000000201419

Ackley, S. F., Zimmerman, S. C., Brenowitz, W. D., Tchetgen Tchetgen, E. J., Gold, A. L., Manly, J. J., Mayeda, E. R., Filshtein, T. J., Power, M. C., Elahi, F. M., Brickman, A. M., & Glymour, M. M. (2021). Effect of reductions in amyloid levels on cognitive change in randomized trials: Instrumental variable meta-analysis. *BMJ (Clinical Research Ed.)*, *372*, n156. https://doi.org/10.1136/bmj.n156

Adesuyan, M., Jani, Y. H., Alsugeir, D., Howard, R., Wong, I. C. K., Wei, L., & Brauer, R. (2023). Trends in the incidence of dementia in people with hypertension in the UK 2000 to 2021. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *15*(3), e12466. https://doi.org/10.1002/dad2.12466

Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *44*(2), 139–160. https://doi.org/10.1111/j.2517-6161.1982.tb01195.x

Ajnakina, O., Cadar, D., & Steptoe, A. (2020). Interplay between Socioeconomic Markers and Polygenic Predisposition on Timing of Dementia Diagnosis. *Journal of the American Geriatrics Society*, *68*(7), 1529–1536. https://doi.org/10.1111/jgs.16406

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., Sotiropoulos, S. N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P., Rorden, C., Daducci, A., Alexander, D. C., Zhang, H., Dragonu, I., Matthews, P. M., … Smith, S. M. (2018). Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage*, *166*, 400–424. https://doi.org/10.1016/j.neuroimage.2017.10.034

Alfaro-Almagro, F., McCarthy, P., Afyouni, S., Andersson, J. L. R., Bastiani, M., Miller, K. L., Nichols, T. E., & Smith, S. M. (2021). Confound modelling in UK Biobank brain imaging. *NeuroImage*, *224*, 117002. https://doi.org/10.1016/j.neuroimage.2020.117002

Alkasir, R., Li, J., Li, X., Jin, M., & Zhu, B. (2017). Human gut microbiota: The links with dementia development. *Protein & Cell*, *8*(2), 90–102. https://doi.org/10.1007/s13238-016-0338-6

Allore, H. G., Zhan, Y., Cohen, A. B., Tinetti, M. E., Trentalange, M., & McAvay, G. (2016). Methodology to Estimate the Longitudinal Average Attributable Fraction of Guideline-recommended Medications for Death in Older Adults With Multiple Chronic Conditions. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *71*(8), 1113–1116. https://doi.org/10.1093/gerona/glv223

Allore, H. G., Zhan, Y., Tinetti, M., Trentalange, M., & McAvay, G. (2015). Longitudinal Average Attributable Fraction as a Method for Studying Time-Varying Conditions and Treatments on Recurrent Self-rated Health: The Case of Medications in Older Adults with Multiple Chronic Conditions. *Annals of Epidemiology*, *25*(9), 681-686.e4. https://doi.org/10.1016/j.annepidem.2015.03.022

Alzheimer, A. (1907). Über eine eigenartige Erkrankung der Hirnrinde. *Allgemeine Zeitschrift Für Psychiatrie Und Psychisch-Gerichtliche Medizin*, *64*, 146–148.

Alzheimer Europe. (2019). *Dementia in Europe Yearbook 2019: Estimating the prevalence of dementia in Europe*. https://www.alzheimer-europe.org/resources/publications/dementia-europe-yearbook-2019-estimating-prevalence-dementia-europe

Alzheimer's Association. (2010). 2010 Alzheimer's disease facts and figures. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *6*(2), 158–194. https://doi.org/10.1016/j.jalz.2010.01.009

Alzheimer's Association. (2022). 2022 Alzheimer's disease facts and figures. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *18*(4), 700–789. https://doi.org/10.1002/alz.12638

Alzheimer's Association. (2023). 2023 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *19*(4), 1598–1695. https://doi.org/10.1002/alz.13016

Alzheimer's Association Workgroup. (2023, October 9). *Revised Criteria for Diagnosis and Staging of Alzheimer's Disease*. alz.org/DiagnosticCriteria

Alzheimer's Disease International, Acosta, D., Krishnamoorthy, E. S., Wortmann, M., Prince, M., Saxena, S., Chair, T. S. S., & World Health Organization. (2012). *Dementia: A public health priority*. https://www.alzint.org/resource/dementia-a-public-health-priority/

American Psychiatric Association, DSM-5 Task Force. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed*. American Psychiatric Publishing, Inc. https://doi.org/10.1176/appi.books.9780890425596

Anderson, E. L., Davies, N. M., Korologou-Linden, R., & Kivimäki, M. (2024). Dementia prevention: The Mendelian randomisation perspective. *Journal of Neurology, Neurosurgery, and Psychiatry*, *95*(4), 384–390. https://doi.org/10.1136/jnnp-2023-332293

Andrieu, S., Guyonnet, S., Coley, N., Cantet, C., Bonnefoy, M., Bordes, S., Bories, L., Cufi, M.-N., Dantoine, T., Dartigues, J.-F., Desclaux, F., Gabelle, A., Gasnier, Y., Pesce, A., Sudres, K., Touchon, J., Robert, P., Rouaud, O., Legrand, P., … Olivier-Abbal, P. (2017). Effect of long-term omega 3 polyunsaturated fatty acid supplementation with

or without multidomain intervention on cognitive function in elderly adults with memory complaints (MAPT): A randomised, placebo-controlled trial. *The Lancet Neurology*, *16*(5), 377–389. https://doi.org/10.1016/S1474-4422(17)30040-6

Ansart, M., Epelbaum, S., Bassignana, G., Bône, A., Bottani, S., Cattai, T., Couronné, R., Faouzi, J., Koval, I., Louis, M., Thibeau-Sutre, E., Wen, J., Wild, A., Burgos, N., Dormont, D., Colliot, O., & Durrleman, S. (2021). Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review. *Medical Image Analysis*, *67*, 101848. https://doi.org/10.1016/j.media.2020.101848

Attia, J., Holliday, E., & Oldmeadow, C. (2022). A proposal for capturing interaction and effect modification using DAGs. *International Journal of Epidemiology*, *51*(4), 1047–1053. https://doi.org/10.1093/ije/dyac126

Avgerinos, K. I., Ferrucci, L., & Kapogiannis, D. (2021). Effects of monoclonal antibodies against amyloid-β on clinical and biomarker outcomes and adverse event risks: A systematic review and meta-analysis of phase III RCTs in Alzheimer's disease. *Ageing Research Reviews*, *68*, 101339. https://doi.org/10.1016/j.arr.2021.101339

Bai, G., Wang, Y., Kuja-Halkola, R., Li, X., Tomata, Y., Karlsson, I. K., Pedersen, N. L., Hägg, S., & Jylhävä, J. (2021). Frailty and the risk of dementia: Is the association explained by shared environmental and genetic factors? *BMC Medicine*, *19*(1), 248. https://doi.org/10.1186/s12916-021-02104-3

Baldini, F., Hertel, J., Sandt, E., Thinnes, C. C., Neuberger-Castillo, L., Pavelka, L., Betsou, F., Krüger, R., Thiele, I., Aguayo, G., Allen, D., Ammerlann, W., Aurich, M., Balling, R., Banda, P., Beaumont, K., Becker, R., Berg, D., Binck, S., … on behalf of the NCER-PD Consortium. (2020). Parkinson's disease-associated alterations of the gut

microbiome predict disease-relevant changes in metabolic functions. *BMC Biology*, *18*(1), 62. https://doi.org/10.1186/s12915-020-00775-7

Banks, J., Batty, G. D., Breedvelt, J. J. F., Coughlin, K., Crawford, R., Marmot, M., Nazroo, J., Oldfield, Z., Steel, N., Steptoe, A., Wood, M., & Zaninotto, P. (2021). *English Longitudinal Study of Ageing: Waves 0-9, 1998-2019* (36th edn) [Data collection]. UK Data Service. https://doi.org/10.5255/UKDA-SN-5050-23

Barbera, M., Perera, D., Matton, A., Mangialasche, F., Rosenberg, A., Middleton, L., Ngandu, T., Solomon, A., & Kivipelto, M. (2023). Multimodal Precision Prevention—A New Direction in Alzheimer's Disease. *The Journal of Prevention of Alzheimer's Disease*, *10*(4), 718–728. https://doi.org/10.14283/jpad.2023.114

Barnes, L. L., Dhana, K., Liu, X., Carey, V. J., Ventrelle, J., Johnson, K., Hollings, C. S., Bishop, L., Laranjo, N., Stubbs, B. J., Reilly, X., Agarwal, P., Zhang, S., Grodstein, F., Tangney, C. C., Holland, T. M., Aggarwal, N. T., Arfanakis, K., Morris, M. C., & Sacks, F. M. (2023). Trial of the MIND Diet for Prevention of Cognitive Decline in Older Persons. *The New England Journal of Medicine*, *389*(7), 602–611. https://doi.org/10.1056/NEJMoa2302368

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Batty, G. D., Gale, C. R., Kivimäki, M., Deary, I. J., & Bell, S. (2020). Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: Prospective cohort study and individual participant meta-analysis. *BMJ*, *368*. https://doi.org/10.1136/bmj.m131

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An Inventory for Measuring Depression. *Archives of General Psychiatry*, *4*(6), 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004

Bergmann, M., Kneip, T., De Luca, G., & Scherpenzeel, A. (2019). *Survey participation in the Survey of Health, Ageing and Retirement in Europe (SHARE), Wave 1-7. Based on Release 7.0.0. 41-2019* (SHARE Working Paper Series). MEA, Max Planck Institute for Social Law and Social Policy.

Bergmann, M., Scherpenzeel, A., & Börsch-Supan, A. (2019). *SHARE Wave 7 Methodology: Panel Innovations and Life Histories*. MEA, Max Planck Institute for Social Law and Social Policy.

Bettio, L. E. B., Rajendran, L., & Gil-Mohapel, J. (2017). The effects of aging in the hippocampus and cognitive decline. *Neuroscience & Biobehavioral Reviews*, *79*, 66–86. https://doi.org/10.1016/j.neubiorev.2017.04.030

Biagi, E., Candela, M., Turroni, S., Garagnani, P., Franceschi, C., & Brigidi, P. (2013). Ageing and gut microbes: Perspectives for health maintenance and longevity. *Pharmacological Research*, *69*(1), 11–20. https://doi.org/10.1016/j.phrs.2012.10.005

Bieri, G., Schroer, A. B., & Villeda, S. A. (2023). Blood-to-brain communication in aging and rejuvenation. *Nature Neuroscience*, *26*(3), 379–393. https://doi.org/10.1038/s41593-022-01238-8

Bir, S. C., Khan, M. W., Javalkar, V., Toledo, E. G., & Kelley, R. E. (2021). Emerging Concepts in Vascular Dementia: A Review. *Journal of Stroke and Cerebrovascular Diseases*, *30*(8), 105864. https://doi.org/10.1016/j.jstrokecerebrovasdis.2021.105864

Bishop, N. J., Eggum-Wilkens, N. D., Haas, S. A., & Kronenfeld, J. J. (2016). Estimating the Co-Development of Cognitive Decline and Physical Mobility Limitations in Older

U.S. Adults. *Demography*, *53*(2), 337–364. https://doi.org/10.1007/s13524-016-0458-x

Bocancea, D. I., van Loenhoud, A. C., Groot, C., Barkhof, F., van der Flier, W. M., & Ossenkoppele, R. (2021). Measuring Resilience and Resistance in Aging and Alzheimer Disease Using Residual Methods. *Neurology*, *97*(10), 474–488. https://doi.org/10.1212/WNL.0000000000012499

Bodelet, J., Potente, C., Blanc, G., Chumbley, J., Imeri, H., Hofer, S., Harris, K. M., Muniz-Terrera, G., & Shanahan, M. (2024). A Bayesian functional approach to test models of life course epidemiology over continuous time. *International Journal of Epidemiology*, *53*(1), dyad190. https://doi.org/10.1093/ije/dyad190

Bond, J., Stave, C., Sganga, A., O'Connell, B., & Stanley, R. L. (2005). Inequalities in dementia care across Europe: Key findings of the Facing Dementia Survey. *International Journal of Clinical Practice. Supplement*, *146*, 8–14. https://doi.org/10.1111/j.1368-504x.2005.00480.x

Börsch-Supan, A. (2022). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 7. Release version: 8.0.0. SHARE-ERIC. Data set.* https://doi.org/10.6103/SHARE.w7.800

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., & Zuber, on behalf of the S. C. C. T., Sabrina. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, *42*(4), 992–1001. https://doi.org/10.1093/ije/dyt088

Bowyer, R. C. E., Jackson, M. A., Le Roy, C. I., Ni Lochlainn, M., Spector, T. D., Dowd, J. B., & Steves, C. J. (2019). Socioeconomic Status and the Gut Microbiome: A

TwinsUK Cohort Study. *Microorganisms*, *7*(1), 17. https://doi.org/10.3390/microorganisms7010017

Bradley, V., & Nichols, T. E. (2022). *Addressing selection bias in the UK Biobank neurological imaging cohort*. medRxiv. https://doi.org/10.1101/2022.01.13.22269266

Bradshaw, A. C., Georges, J., & Alzheimer Europe Board. (2024). Anti-Amyloid Therapies for Alzheimer's Disease: An Alzheimer Europe Position Paper and Call to Action. *The Journal of Prevention of Alzheimer's Disease*, *11*(2), 265–273. https://doi.org/10.14283/jpad.2024.37

Brenowitz, W. D. (2021). Invited Commentary: Body Mass Index and Risk of Dementia—Potential Explanations for Life-Course Differences in Risk Estimates and Future Research Directions. *American Journal of Epidemiology*, *190*(12), 2511–2514. https://doi.org/10.1093/aje/kwab095

Brenowitz, W. D., Zimmerman, S. C., Filshtein, T. J., Yaffe, K., Walter, S., Hoffmann, T. J., Jorgenson, E., Whitmer, R. A., & Glymour, M. M. (2021). Extension of Mendelian Randomization to Identify Earliest Manifestations of Alzheimer Disease: Association of Genetic Risk Score for Alzheimer Disease With Lower Body Mass Index by Age 50 Years. *American Journal of Epidemiology*, *190*(10), 2163. https://doi.org/10.1093/aje/kwab103

Brodaty, H., & Donkin, M. (2009). Family caregivers of people with dementia. *Dialogues in Clinical Neuroscience*, *11*(2), 217–228. https://doi.org/10.31887/DCNS.2009.11.2/hbrodaty

Brück, C. C., Wolters, F. J., Ikram, M. A., & de Kok, I. M. C. M. (2022). Projected prevalence and incidence of dementia accounting for secular trends and birth cohort effects: A population-based microsimulation study. *European Journal of Epidemiology*, *37*(8), 807–814. https://doi.org/10.1007/s10654-022-00878-1

Brugulat-Serrat, A., Salvadó, G., Sudre, C. H., Grau-Rivera, O., Suárez-Calvet, M., Falcon, C., Sánchez-Benavides, G., Gramunt, N., Fauria, K., Cardoso, M. J., Barkhof, F., Molinuevo, J. L., Gispert, J. D., Camí, J., Cacciaglia, R., Operto, G., Skouras, S., Minguillón, C., Polo, A., … for the ALFA Study. (2020). Patterns of white matter hyperintensities associated with cognition in middle-aged cognitively healthy individuals. *Brain Imaging and Behavior*, *14*(5), 2012–2023. https://doi.org/10.1007/s11682-019-00151-2

Buckinx, F., & Aubertin-Leheudre, M. (2021). Nutrition to Prevent or Treat Cognitive Impairment in Older Adults: A GRADE Recommendation. *The Journal of Prevention of Alzheimer's Disease*, *8*(1), 110–116. https://doi.org/10.14283/jpad.2020.40

Bunnik, E. M., Smedinga, M., Milne, R., Georges, J., Richard, E., & Schermer, M. H. N. (2022). Ethical Frameworks for Disclosure of Alzheimer Disease Biomarkers to Research Participants: Conflicting Norms and a Nuanced Policy. *Ethics & Human Research*, *44*(6), 2–13. https://doi.org/10.1002/eahr.500146

Bussel, E. F. van, Richard, E., Arts, D. L., Nooyens, A. C. J., Coloma, P. M., Waal, M. W. M. de, Akker, M. van den, Biermans, M. C. J., Nielen, M. M. J., Boven, K. van, Smeets, H., Matthews, F. E., Brayne, C., Busschers, W. B., Gool, W. A. van, & Charante, E. P. M. van. (2017). Dementia incidence trend over 1992-2014 in the Netherlands: Analysis of primary care data. *PLOS Medicine*, *14*(3), e1002235. https://doi.org/10.1371/journal.pmed.1002235

Cabrera, C., Vicens, P., & Torrente, M. (2021). Modifiable Risk Factors for Dementia: The Role of Gut Microbiota. *Current Alzheimer Research*, *18*(13), 993–1009. https://doi.org/10.2174/1567205018666211215152411

Cadar, D., Lassale, C., Davies, H., Llewellyn, D. J., Batty, G. D., & Steptoe, A. (2018). Individual and Area-Based Socioeconomic Factors Associated With Dementia

Incidence in England: Evidence From a 12-Year Follow-up in the English Longitudinal Study of Ageing. *JAMA Psychiatry*, *75*(7), 723–732. https://doi.org/10.1001/jamapsychiatry.2018.1012

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

CentERdata, Institute for data collection and research, & SHARE Central, Munich Center for the Economics of Aging. (2024). *SHARE Data & Documentation Tool* [Application]. https://www.share-datadocutool.org/

Cermakova, P., Eriksdotter, M., Lund, L. H., Winblad, B., Religa, P., & Religa, D. (2015). Heart failure and Alzheimer's disease. *Journal of Internal Medicine*, *277*(4), 406–425. https://doi.org/10.1111/joim.12287

Chao, A. (1984). Nonparametric Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics*, *11*(4), 265–270. https://www.jstor.org/stable/4615964

Chapko, D., McCormack, R., Black, C., Staff, R., & Murray, A. (2018). Life-course determinants of cognitive reserve (CR) in cognitive aging and dementia - a systematic literature review: Aging & Mental Health. *Aging & Mental Health*, *22*(8), 915–926. https://doi.org/10.1080/13607863.2017.1348471

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, L., Xu, X., Wu, X., Cao, H., Li, X., Hou, Z., Wang, B., Liu, J., Ji, X., Zhang, P., & Li, H. (2022). A comparison of the composition and functions of the oral and gut

microbiotas in Alzheimer's patients. *Frontiers in Cellular and Infection Microbiology*, *12*, 942460. https://doi.org/10.3389/fcimb.2022.942460

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, X., Lee, C., & Huang, H. (2022). Neighborhood built environment associated with cognition and dementia risk among older adults: A systematic literature review. *Social Science & Medicine*, *292*, 114560. https://doi.org/10.1016/j.socscimed.2021.114560

Chumbley, J., Xu, W., Potente, C., Harris, K. M., & Shanahan, M. (2021). A Bayesian approach to comparing common models of life-course epidemiology. *International Journal of Epidemiology*, *50*(5), 1660–1670. https://doi.org/10.1093/ije/dyab073

Cigolle, C. T., Nagel, C. L., Blaum, C. S., Liang, J., & Quiñones, A. R. (2018). Inconsistency in the Self-report of Chronic Diseases in Panel Surveys: Developing an Adjudication Method for the Health and Retirement Study. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *73*(5), 901–912. https://doi.org/10.1093/geronb/gbw063

Ciria, L. F., Román-Caballero, R., Vadillo, M. A., Holgado, D., Luque-Casado, A., Perakakis, P., & Sanabria, D. (2023). An umbrella review of randomized control trials on the effects of physical exercise on cognition. *Nature Human Behaviour*, *7*(6), 928–941. https://doi.org/10.1038/s41562-023-01554-4

Clarke, P. J., Ailshire, J. A., House, J. S., Morenoff, J. D., King, K., Melendez, R., & Langa, K. M. (2012). Cognitive function in the community setting: The neighbourhood as a source of 'cognitive reserve'? *Journal of Epidemiology and Community Health*, *66*(8), 730–736. https://doi.org/10.1136/jech.2010.128116

Clarke, P. J., Weuve, J., Barnes, L., Evans, D. A., & Mendes de Leon, C. F. (2015). Cognitive decline and the neighborhood environment. *Annals of Epidemiology*, *25*(11), 849–854. https://doi.org/10.1016/j.annepidem.2015.07.001

Cleret de Langavant, L., Bayen, E., Bachoud-Lévi, A., & Yaffe, K. (2020). Approximating dementia prevalence in population-based surveys of aging worldwide: An unsupervised machine learning approach. *Alzheimer's & Dementia : Translational Research & Clinical Interventions*, *6*(1), e12074. https://doi.org/10.1002/trc2.12074

Cleret de Langavant, L., Bayen, E., & Yaffe, K. (2018). Unsupervised Machine Learning to Identify High Likelihood of Dementia in Population-Based Surveys: Development and Validation Study. *Journal of Medical Internet Research*, *20*(7), e10493. https://doi.org/10.2196/10493

Cole, S. R., & Hernán, M. A. (2008). Constructing Inverse Probability Weights for Marginal Structural Models. *American Journal of Epidemiology*, *168*(6), 656–664. https://doi.org/10.1093/aje/kwn164

Cotman, C. W., Berchtold, N. C., & Christie, L.-A. (2007). Exercise builds brain health: Key roles of growth factor cascades and inflammation. *Trends in Neurosciences*, *30*(9), 464–472. https://doi.org/10.1016/j.tins.2007.06.011

Cox, S. R., Lyall, D. M., Ritchie, S. J., Bastin, M. E., Harris, M. A., Buchanan, C. R., Fawns-Ritchie, C., Barbu, M. C., de Nooij, L., Reus, L. M., Alloza, C., Shen, X., Neilson, E., Alderson, H. L., Hunter, S., Liewald, D. C., Whalley, H. C., McIntosh, A. M., Lawrie, S. M., … Deary, I. J. (2019). Associations between vascular risk factors and brain MRI indices in UK Biobank. *European Heart Journal*, *40*(28), 2290–2300. https://doi.org/10.1093/eurheartj/ehz100

Crimmins, E. M., Kim, J. K., Langa, K. M., & Weir, D. R. (2011). Assessment of cognition using surveys and neuropsychological assessment: The Health and Retirement Study

and the Aging, Demographics, and Memory Study. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *66*(Suppl 1), i162-171. https://doi.org/10.1093/geronb/gbr048

Cummings, J. (2023). Anti-Amyloid Monoclonal Antibodies are Transformative Treatments that Redefine Alzheimer's Disease Therapeutics. *Drugs*, *83*(7), 569–576. https://doi.org/10.1007/s40265-023-01858-9

Cummings, J., Apostolova, L., Rabinovici, G. D., Atri, A., Aisen, P., Greenberg, S., Hendrix, S., Selkoe, D., Weiner, M., Petersen, R. C., Salloway, S., & For the Alzheimer's Disease and Related Disorders Therapeutics Work Group. (2023). Lecanemab: Appropriate Use Recommendations. *The Journal of Prevention of Alzheimer's Disease*, *10*(3), 362–377. https://doi.org/10.14283/jpad.2023.30

d'Uva, T. B., Lindeboom, M., O'Donnell, O., & van Doorslaer, E. (2011). Education-related inequity in healthcare with heterogeneous reporting of health. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, *174*(3), 639–664. https://doi.org/10.1111/j.1467-985X.2011.00706.x

Daran, B., Levasseur, P., & Clément, M. (2023). Updating the association between socioeconomic status and obesity in low-income and lower-middle-income sub-Saharan African countries: A literature review. *Obesity Reviews*, *24*(10), e13601. https://doi.org/10.1111/obr.13601

Dawes, P., & Völter, C. (2023). Do hearing loss interventions prevent dementia? *Zeitschrift Fur Gerontologie Und Geriatrie*, *56*(4), 261–268. https://doi.org/10.1007/s00391-023-02178-z

de Keijzer, C., Bauwelinck, M., & Dadvand, P. (2020). Long-Term Exposure to Residential Greenspace and Healthy Ageing: A Systematic Review. *Current Environmental Health Reports*, *7*(1), 65–88. https://doi.org/10.1007/s40572-020-00264-7

de Lange, A.-M. G., Kaufmann, T., Quintana, D. S., Winterton, A., Andreassen, O. A., Westlye, L. T., & Ebmeier, K. P. (2021). Prominent health problems, socioeconomic deprivation, and higher brain age in lonely and isolated individuals: A population-based study. *Behavioural Brain Research*, *414*, 113510. https://doi.org/10.1016/j.bbr.2021.113510

De Rubeis, V., Bayat, S., Griffith, L. E., Smith, B. T., & Anderson, L. N. (2019). Validity of self-reported recall of anthropometric measures in early life: A systematic review and meta-analysis. *Obesity Reviews*, *20*(10), 1426–1440. https://doi.org/10.1111/obr.12881

Debette, S., Schilling, S., Duperron, M.-G., Larsson, S. C., & Markus, H. S. (2019). Clinical Significance of Magnetic Resonance Imaging Markers of Vascular Brain Injury. *JAMA Neurology*, *76*(1), 81–94. https://doi.org/10.1001/jamaneurol.2018.3122

Deckers, K., Cadar, D., van Boxtel, M. P. J., Verhey, F. R. J., Steptoe, A., & Köhler, S. (2019). Modifiable Risk Factors Explain Socioeconomic Inequalities in Dementia Risk: Evidence from a Population-Based Prospective Cohort Study. *Journal of Alzheimer's Disease*, *71*(2), 549–557. https://doi.org/10.3233/JAD-190541

DeJong, N. R., Jansen, J. F. A., van Boxtel, M. P. J., Schram, M. T., Stehouwer, C. D. A., Dagnelie, P. C., van der Kallen, C. J. H., Kroon, A. A., Wesselius, A., Koster, A., Backes, W. H., & Köhler, S. (2023). Cognitive resilience depends on white matter connectivity: The Maastricht Study. *Alzheimer's & Dementia*, *19*(4), 1164–1174. https://doi.org/10.1002/alz.12758

Delpino, F. M., Figueiredo, L. M., Bielemann, R. M., Da Silva, B. G. C., Dos Santos, F. S., Mintem, G. C., Flores, T. R., Arcêncio, R. A., & Nunes, B. P. (2022). Ultra-processed food and risk of type 2 diabetes: A systematic review and meta-analysis of

longitudinal studies. *International Journal of Epidemiology*, *51*(4), 1120–1141. https://doi.org/10.1093/ije/dyab247

Demnitz, N., Anatürk, M., Allan, C. L., Filippini, N., Griffanti, L., Mackay, C. E., Mahmood, A., Sexton, C. E., Suri, S., Topiwala, A. G., Zsoldos, E., Kivimäki, M., Singh-Manoux, A., & Ebmeier, K. P. (2020). Association of trajectories of depressive symptoms with vascular risk, cognitive function and adverse brain outcomes: The Whitehall II MRI sub-study. *Journal of Psychiatric Research*, *131*, 85–93. https://doi.org/10.1016/j.jpsychires.2020.09.005

Desai, R., John, A., Saunders, R., Marchant, N. L., Buckman, J. E. J., Charlesworth, G., Zuber, V., & Stott, J. (2023). Examining the Lancet Commission risk factors for dementia using Mendelian randomisation. *BMJ Mental Health*, *26*(1), e300555. https://doi.org/10.1136/bmjment-2022-300555

Ding, T., & Schloss, P. D. (2014). Dynamics and associations of microbial community types across the human body. *Nature*, *509*(7500), 357–360. https://doi.org/10.1038/nature13178

Donders, A. R. T., van der Heijden, G. J. M. G., Stijnen, T., & Moons, K. G. M. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, *59*(10), 1087–1091. https://doi.org/10.1016/j.jclinepi.2006.01.014

Doove, L. L., Van Buuren, S., & Dusseldorp, E. (2014). Recursive partitioning for missing data imputation in the presence of interaction effects. *Computational Statistics & Data Analysis*, *72*(C), 92–104. https://doi.org/10.1016/j.csda.2013.10.025.

Dove, A., Guo, J., Marseglia, A., Fastbom, J., Vetrano, D. L., Fratiglioni, L., Pedersen, N. L., & Xu, W. (2023). Cardiometabolic multimorbidity and incident dementia: The Swedish twin registry. *European Heart Journal*, *44*(7), 573–582. https://doi.org/10.1093/eurheartj/ehac744

Dove, A., & Xu, W. (2023). Cardiometabolic multimorbidity and cognitive decline. *The Lancet Healthy Longevity*, *4*(6), e241–e242. https://doi.org/10.1016/S2666-7568(23)00053-3

Elman, J. A., Vogel, J. W., Bocancea, D. I., Ossenkoppele, R., van Loenhoud, A. C., Tu, X. M., Kremen, W. S., & the Alzheimer's Disease Neuroimaging Initiative. (2022). Issues and recommendations for the residual approach to quantifying cognitive resilience and reserve. *Alzheimer's Research & Therapy*, *14*(1), 102. https://doi.org/10.1186/s13195-022-01049-w

Espeland, M. A., Luchsinger, J. A., Baker, L. D., Neiberg, R., Kahn, S. E., Arnold, S. E., Wing, R. R., Blackburn, G. L., Bray, G., Evans, M., Hazuda, H. P., Jeffery, R. W., Wilson, V. M., Clark, J. M., Coday, M., Demos-McDermott, K., Foreyt, J. P., Greenway, F., Hill, J. O., … For the Look AHEAD Study Group. (2017). Effect of a long-term intensive lifestyle intervention on prevalence of cognitive impairment. *Neurology*, *88*(21), 2026–2035. https://doi.org/10.1212/WNL.0000000000003955

European Alzheimer's & Dementia Biobank Mendelian Randomization Collaboration, Luo, J., Thomassen, J. Q., Bellenguez, C., Grenier-Boley, B., de Rojas, I., Castillo, A., Parveen, K., Küçükali, F., Nicolas, A., Peters, O., Schneider, A., Dichgans, M., Rujescu, D., Scherbaum, N., Jürgen, D., Riedel-Heller, S., Hausner, L., Porcel, L. M., … Frikke-Schmidt, R. (2023). Genetic Associations Between Modifiable Risk Factors and Alzheimer Disease. *JAMA Network Open*, *6*(5), e2313734. https://doi.org/10.1001/jamanetworkopen.2023.13734

Faith, D. P., Minchin, P. R., & Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, *69*(1), 57–68. https://doi.org/10.1007/BF00038687

Ferraro, K. F., Bauldry, S., Sauerteig-Rolston, M. R., & Thomas, P. A. (2023). Dual Functionality in Later Life. *The Gerontologist*, *63*(7), 1110–1116. https://doi.org/10.1093/geront/gnad031

Ferrucci, L., Gonzalez-Freire, M., Fabbri, E., Simonsick, E., Tanaka, T., Moore, Z., Salimi, S., Sierra, F., & de Cabo, R. (2020). Measuring biological aging in humans: A quest. *Aging Cell*, *19*(2), e13080. https://doi.org/10.1111/acel.13080

Fierini, F. (2020). Mixed dementia: Neglected clinical entity or nosographic artifice? *Journal of the Neurological Sciences*, *410*, 116662. https://doi.org/10.1016/j.jns.2019.116662

Finch, C. E., & Kulminski, A. M. (2019). The Alzheimer's Disease Exposome. *Alzheimer's & Dementia*, *15*(9), 1123–1132. https://doi.org/10.1016/j.jalz.2019.06.3914

First, M. B., Gaebel, W., Maj, M., Stein, D. J., Kogan, C. S., Saunders, J. B., Poznyak, V. B., Gureje, O., Lewis-Fernández, R., Maercker, A., Brewin, C. R., Cloitre, M., Claudino, A., Pike, K. M., Baird, G., Skuse, D., Krueger, R. B., Briken, P., Burke, J. D., … Reed, G. M. (2021). An organization- and category-level comparison of diagnostic requirements for mental disorders in ICD-11 and DSM-5. *World Psychiatry*, *20*(1), 34–51. https://doi.org/10.1002/wps.20825

Flier, W. M. van der, & Scheltens, P. (2005). Epidemiology and risk factors of dementia. *Journal of Neurology, Neurosurgery & Psychiatry*, *76*(suppl 5), v2–v7. https://doi.org/10.1136/jnnp.2005.082867

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

Formánek, T., Kagstrom, A., Winkler, P., & Cermakova, P. (2019). Differences in cognitive performance and cognitive decline across European regions: A population-based

prospective cohort study. *European Psychiatry*, *58*, 80–86. https://doi.org/10.1016/j.eurpsy.2019.03.001

Franciotti, R., Nardini, D., Russo, M., Onofrj, M., & Sensi, S. L. (2023). Comparison of Machine Learning-based Approaches to Predict the Conversion to Alzheimer's Disease from Mild Cognitive Impairment. *Neuroscience*, *514*, 143–152. https://doi.org/10.1016/j.neuroscience.2023.01.029

Franco-Marina, F., García-González, J. J., Wagner-Echeagaray, F., Gallo, J., Ugalde, O., Sánchez-García, S., Espinel-Bermúdez, C., Juárez-Cedillo, T., Rodríguez, M. Á. V., & García-Peña, C. (2010). The Mini-mental State Examination revisited: Ceiling and floor effects after score adjustment for educational level in an aging Mexican population. *International Psychogeriatrics*, *22*(1), 72–81. https://doi.org/10.1017/S1041610209990822

Frisoni, G. B., Altomare, D., Thal, D. R., Ribaldi, F., van der Kant, R., Ossenkoppele, R., Blennow, K., Cummings, J., van Duijn, C., Nilsson, P. M., Dietrich, P.-Y., Scheltens, P., & Dubois, B. (2022). The probabilistic model of Alzheimer disease: The amyloid hypothesis revised. *Nature Reviews Neuroscience*, *23*(1), 53–66. https://doi.org/10.1038/s41583-021-00533-w

Gamble, L. D., Matthews, F. E., Jones, I. R., Hillman, A. E., Woods, B., Macleod, C. A., Martyr, A., Collins, R., Pentecost, C., Rusted, J. M., & Clare, L. (2022). Characteristics of people living with undiagnosed dementia: Findings from the CFAS Wales study. *BMC Geriatrics*, *22*(1), 409. https://doi.org/10.1186/s12877-022-03086-4

GBD 2019 Collaborators, Nichols, E., Abd-Allah, F., Abdoli, A., Abosetugn, A. E., Abrha, W. A., Abualhasan, A., Abu-Gharbieh, E., Akinyemi, R. O., & Alahdab, F. (2021). Global mortality from dementia: Application of a new method and results from the

Global Burden of Disease Study 2019. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *7*(1), e12200. https://doi.org/10.1002/trc2.12200

Geraets, A. F. J., & Leist, A. K. (2023). Sex/gender and socioeconomic differences in modifiable risk factors for dementia. *Scientific Reports*, *13*(1), 80. https://doi.org/10.1038/s41598-022-27368-4

Gharbi-Meliani, A., Husson, F., Vandendriessche, H., Bayen, E., Yaffe, K., Bachoud-Lévi, A.-C., & Cleret de Langavant, L. (2023). Identification of high likelihood of dementia in population-based surveys using unsupervised clustering: A longitudinal analysis. *Alzheimer's Research & Therapy*, *15*(1), 209. https://doi.org/10.1186/s13195-023-01357-9

Gianattasio, K. Z., Ciarleglio, A., & Power, M. C. (2020). Development of Algorithmic Dementia Ascertainment for Racial/Ethnic Disparities Research in the US Health and Retirement Study. *Epidemiology (Cambridge, Mass.)*, *31*(1), 126–133. https://doi.org/10.1097/EDE.0000000000001101

Gianattasio, K. Z., Wu, Q., Glymour, M. M., & Power, M. C. (2019). Comparison of Methods for Algorithmic Classification of Dementia Status in the Health and Retirement Study. *Epidemiology (Cambridge, Mass.)*, *30*(2), 291–302. https://doi.org/10.1097/EDE.0000000000000945

Gillman, M. W. (2015). Primordial prevention of cardiovascular disease. *Circulation*, *131*(7), 599–601. https://doi.org/10.1161/CIRCULATIONAHA.115.014849

Glymour, M. M., Kawachi, I., Jencks, C. S., & Berkman, L. F. (2008). Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. *Journal of Epidemiology & Community Health*, *62*(6), 532–537. https://doi.org/10.1136/jech.2006.059469

Gómez-Isla, T., & Frosch, M. P. (2019). The Challenge of Defining Alzheimer Disease Based on Biomarkers in the Absence of Symptoms. *JAMA Neurology*, *76*(10), 1143–1144. https://doi.org/10.1001/jamaneurol.2019.1667

Gómez-Isla, T., & Frosch, M. P. (2022). Lesions without symptoms: Understanding resilience to Alzheimer disease neuropathological changes. *Nature Reviews Neurology*, *18*(6), 323–332. https://doi.org/10.1038/s41582-022-00642-9

Gorber, S. C., Tremblay, M., Moher, D., & Gorber, B. (2007). A comparison of direct vs. self-report measures for assessing height, weight and body mass index: A systematic review. *Obesity Reviews*, *8*(4), 307–326. https://doi.org/10.1111/j.1467-789X.2007.00347.x

Gottesman, R. F., Schneider, A. L. C., Zhou, Y., Coresh, J., Green, E., Gupta, N., Knopman, D. S., Mintz, A., Rahmim, A., Sharrett, A. R., Wagenknecht, L. E., Wong, D. F., & Mosley, T. H. (2017). Association Between Midlife Vascular Risk Factors and Estimated Brain Amyloid Deposition. *JAMA*, *317*(14), 1443–1450. https://doi.org/10.1001/jama.2017.3090

Grande, G., Marengoni, A., Vetrano, D. L., Roso-Llorach, A., Rizzuto, D., Zucchelli, A., Qiu, C., Fratiglioni, L., & Calderón-Larrañaga, A. (2021). Multimorbidity burden and dementia risk in older adults: The role of inflammation and genetics. *Alzheimer's & Dementia*, *17*(5), 768–776. https://doi.org/10.1002/alz.12237

Greenhalgh, K., Meyer, K. M., Aagaard, K. M., & Wilmes, P. (2016). The human gut microbiome in health: Establishment and resilience of microbiota over a lifetime. *Environmental Microbiology*, *18*(7), 2103–2116. https://doi.org/10.1111/1462-2920.13318

Grill, J. D., & Karlawish, J. (2022). Disclosing Alzheimer Disease Biomarker Results to Research Participants. *JAMA Neurology*, *79*(7), 645–646. https://doi.org/10.1001/jamaneurol.2022.1307

Gross, A. L., Li, C., Briceño, E. M., Rentería, M. A., Jones, R. N., Langa, K. M., Manly, J. J., Nichols, E., Weir, D., Wong, R., Berkman, L., Lee, J., & Kobayashi, L. C. (2023). Harmonisation of later-life cognitive function across national contexts: Results from the Harmonized Cognitive Assessment Protocols. *The Lancet Healthy Longevity*, *4*(10), e573–e583. https://doi.org/10.1016/S2666-7568(23)00170-8

Guo, J., Gao, B., Huang, Y., & Song, S. (2024). Trajectory of multimorbidity before dementia: A 24-year follow-up study. *Alzheimer's & Dementia : Diagnosis, Assessment & Disease Monitoring*, *16*(1), e12523. https://doi.org/10.1002/dad2.12523

Gustafsson, P. E., San Sebastian, M., Janlert, U., Theorell, T., Westerlund, H., & Hammarström, A. (2014). Life-Course Accumulation of Neighborhood Disadvantage and Allostatic Load: Empirical Integration of Three Social Determinants of Health Frameworks. *American Journal of Public Health*, *104*(5), 904–910. https://doi.org/10.2105/AJPH.2013.301707

Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2018). Worldwide trends in insufficient physical activity from 2001 to 2016: A pooled analysis of 358 population-based surveys with 1·9 million participants. *The Lancet Global Health*, *6*(10), e1077–e1086. https://doi.org/10.1016/S2214-109X(18)30357-7

Gutierrez, S., Meza, E., Glymour, M. M., & Torres, J. M. (2023). My Parent, Myself, or My Child: Whose Education Matters Most for Trajectories of Cognitive Aging in Middle Age? *American Journal of Epidemiology*, kwad108. https://doi.org/10.1093/aje/kwad108

Hafdi, M., Hoevenaar-Blom, M. P., & Richard, E. (2021). Multi-domain interventions for the prevention of dementia and cognitive decline. *The Cochrane Database of Systematic Reviews*, *2021*(11), CD013572. https://doi.org/10.1002/14651858.CD013572.pub2

Hardy, J. A., & Higgins, G. A. (1992). Alzheimer's Disease: The Amyloid Cascade Hypothesis. *Science*, *256*(5054), 184–185. https://doi.org/10.1126/science.1566067

Harrison, K. L., Hunt, L. J., Ritchie, C. S., & Yaffe, K. (2019). Dying with Dementia: Under-recognized and Stigmatized. *Journal of the American Geriatrics Society*, *67*(8), 1548–1551. https://doi.org/10.1111/jgs.15895

HCAP Network. (2024, January 19). *Harmonized Cognitive Assessment Protocol Network.* Understanding Cognition Through Harmonized Data. https://hcap.isr.umich.edu/

Heck, K. E., Braveman, P., Cubbin, C., Chávez, G. F., & Kiely, J. L. (2006). Socioeconomic Status and Breastfeeding Initiation Among California Mothers. *Public Health Reports*, *121*(1), 51. https://doi.org/10.1177/003335490612100111

Heintz-Buschart, A., & Wilmes, P. (2018). Human Gut Microbiome: Function Matters. *Trends in Microbiology*, *26*(7), 563–574. https://doi.org/10.1016/j.tim.2017.11.002

Hendriks, S., Peetoom, K., Bakker, C., van der Flier, W. M., Papma, J. M., Koopmans, R., Verhey, F. R. J., de Vugt, M., Köhler, S., & Young-Onset Dementia Epidemiology Study Group. (2021). Global Prevalence of Young-Onset Dementia: A Systematic Review and Meta-analysis. *JAMA Neurology*, *78*(9), 1080–1090. https://doi.org/10.1001/jamaneurol.2021.2161

Hendriks, S., Ranson, J. M., Peetoom, K., Lourida, I., Tai, X. Y., de Vugt, M., Llewellyn, D. J., & Köhler, S. (2024). Risk Factors for Young-Onset Dementia in the UK Biobank. *JAMA Neurology*, *81*(2), 134–142. https://doi.org/10.1001/jamaneurol.2023.4929

Henney, A. E., Gillespie, C. S., Alam, U., Hydes, T. J., Mackay, C. E., & Cuthbertson, D. J. (2024). High intake of ultra-processed food is associated with dementia in adults: A

systematic review and meta-analysis of observational studies. *Journal of Neurology*, *271*(1), 198–210. https://doi.org/10.1007/s00415-023-12033-1

Hernán, M. A., & Robins, J. M. (2016). Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *American Journal of Epidemiology*, *183*(8), 758–764. https://doi.org/10.1093/aje/kwv254

Hernán, M. A., & Taubman, S. L. (2008). Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *International Journal of Obesity (2005)*, *32*(Suppl 3), S8-14. https://doi.org/10.1038/ijo.2008.82

Herzog, A. R., & Wallace, R. B. (1997). Measures of cognitive functioning in the AHEAD Study. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *52 Spec No*, 37–48. https://doi.org/10.1093/geronb/52b.special_issue.37

Hill, T. C. J., Walsh, K. A., Harris, J. A., & Moffett, B. F. (2003). Using ecological diversity measures with bacterial communities. *FEMS Microbiology Ecology*, *43*(1), 1–11. https://doi.org/10.1111/j.1574-6941.2003.tb01040.x

Hipp, G., Vaillant, M., Diederich, N. J., Roomp, K., Satagopam, V. P., Banda, P., Sandt, E., Mommaerts, K., Schmitz, S. K., Longhino, L., Schweicher, A., Hanff, A.-M., Nicolai, B., Kolber, P., Reiter, D., Pavelka, L., Binck, S., Pauly, C., Geffers, L., … Krüger, R. (2018). The Luxembourg Parkinson's Study: A Comprehensive Approach for Stratification and Early Diagnosis. *Frontiers in Aging Neuroscience*, *10*, 326. https://doi.org/10.3389/fnagi.2018.00326

Hofbauer, L. M., & Rodriguez, F. S. (2023). The role of social deprivation and depression in dementia risk: Findings from the longitudinal survey of health, ageing and retirement in Europe. *Epidemiology and Psychiatric Sciences*, *32*, e10. https://doi.org/10.1017/S2045796023000033

Honig, M. S., & Bock, T. (2017). Luxembourg – ECEC Workforce Profile. In P. Oberhuemer & I. Schreyer (Eds.), *Workforce Profiles in Systems of Early Childhood Education and Care in Europe*. http://www.seepro.eu/English/pdfs/LUXEMBOURG_Workforce.pdf

Hoops, S., Nazem, S., Siderowf, A. D., Duda, J. E., Xie, S. X., Stern, M. B., & Weintraub, D. (2009). Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology*, *73*(21), 1738–1745. https://doi.org/10.1212/WNL.0b013e3181c34b47

Hu, Y.-J., & Satten, G. A. (2020). Testing hypotheses about the microbiome using the linear decomposition model (LDM). *Bioinformatics (Oxford, England)*, *36*(14), 4106–4115. https://doi.org/10.1093/bioinformatics/btaa260

Hunt, J. F. V., Vogt, N. M., Jonaitis, E. M., Buckingham, W. R., Koscik, R. L., Zuelsdorff, M., Clark, L. R., Gleason, C. E., Yu, M., Okonkwo, O., Johnson, S. C., Asthana, S., Bendlin, B. B., & Kind, A. J. H. (2021). Association of Neighborhood Context, Cognitive Decline, and Cortical Change in an Unimpaired Cohort. *Neurology*, *96*(20), e2500–e2512. https://doi.org/10.1212/WNL.0000000000011918

Hurd, M. D., Martorell, P., Delavande, A., Mullen, K. J., & Langa, K. M. (2013). Monetary Costs of Dementia in the United States. *New England Journal of Medicine*, *368*(14), 1326–1334. https://doi.org/10.1056/NEJMsa1204629

Jack, C. R., Jr, Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., … Silverberg, N. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

Jack, C. R., Jr, Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., Shaw, L. M., Vemuri, P., Wiste, H. J., Weigand, S. D., Lesnick, T. G., Pankratz, V. S., Donohue, M. C., & Trojanowski, J. Q. (2013). Update on hypothetical model of Alzheimer's disease biomarkers. *Lancet Neurology*, *12*(2), 207–216. https://doi.org/10.1016/S1474-4422(12)70291-0

Jack, C. R., Jr, Therneau, T. M., Weigand, S. D., Wiste, H. J., Knopman, D. S., Vemuri, P., Lowe, V. J., Mielke, M. M., Roberts, R. O., Machulda, M. M., Graff-Radford, J., Jones, D. T., Schwarz, C. G., Gunter, J. L., Senjem, M. L., Rocca, W. A., & Petersen, R. C. (2019). Prevalence of Biologically vs Clinically Defined Alzheimer Spectrum Entities Using the National Institute on Aging–Alzheimer's Association Research Framework. *JAMA Neurology*, *76*(10), 1174–1183. https://doi.org/10.1001/jamaneurol.2019.1971

Jack, C. R., Jr, Wiste, H. J., Algeciras-Schimnich, A., Weigand, S. D., Figdore, D. J., Lowe, V. J., Vemuri, P., Graff-Radford, J., Ramanan, V. K., Knopman, D. S., Mielke, M. M., Machulda, M. M., Fields, J., Schwarz, C. G., Cogswell, P. M., Senjem, M. L., Therneau, T. M., & Petersen, R. C. (2024). Comparison of plasma biomarkers and amyloid PET for predicting memory decline in cognitively unimpaired individuals. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *20*(3), 2143–2154. https://doi.org/10.1002/alz.13651

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., Sealock, J., Karlsson, I. K., Hägg, S., Athanasiu, L., Voyle, N., Proitsi, P., Witoelar, A., Stringer, S., Aarsland, D., Almdahl, I. S., Andersen, F., Bergh, S., Bettella, F., … Posthuma, D. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics*, *51*(3), 404–413. https://doi.org/10.1038/s41588-018-0311-9

Jin, Y., Liang, J., Hong, C., Liang, R., & Luo, Y. (2023). Cardiometabolic multimorbidity, lifestyle behaviours, and cognitive function: A multicohort study. *The Lancet. Healthy Longevity*, *4*(6), e265–e273. https://doi.org/10.1016/S2666-7568(23)00054-5

Jönsson, L., Wimo, A., Handels, R., Johansson, G., Boada, M., Engelborghs, S., Frölich, L., Jessen, F., Kehoe, P. G., Kramberger, M., de Mendonça, A., Ousset, P. J., Scarmeas, N., Visser, P. J., Waldemar, G., & Winblad, B. (2023). The affordability of lecanemab, an amyloid-targeting therapy for Alzheimer's disease: An EADC-EC viewpoint. *The Lancet Regional Health - Europe*, *29*, 100657. https://doi.org/10.1016/j.lanepe.2023.100657

Kaul, A., Mandal, S., Davidov, O., & Peddada, S. D. (2017). Analysis of Microbiome Data in the Presence of Excess Zeros. *Frontiers in Microbiology*, *8*, 2114. https://doi.org/10.3389/fmicb.2017.02114

Kim, S., Egerter, S., Cubbin, C., Takahashi, E. R., & Braveman, P. (2007). Potential Implications of Missing Income Data in Population-Based Surveys: An Example from a Postpartum Survey in California. *Public Health Reports*, *122*(6), 753–763. https://doi.org/10.1177/003335490712200607

Kivimäki, M., Luukkonen, R., Batty, G. D., Ferrie, J. E., Pentti, J., Nyberg, S. T., Shipley, M. J., Alfredsson, L., Fransson, E. I., Goldberg, M., Knutsson, A., Koskenvuo, M., Kuosma, E., Nordin, M., Suominen, S. B., Theorell, T., Vuoksimaa, E., Westerholm, P., Westerlund, H., … Jokela, M. (2018). Body mass index and risk of dementia: Analysis of individual-level data from 1.3 million individuals. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *14*(5), 601–609. https://doi.org/10.1016/j.jalz.2017.09.016

Kivimäki, M., Singh-Manoux, A., Pentti, J., Sabia, S., Nyberg, S. T., Alfredsson, L., Goldberg, M., Knutsson, A., Koskenvuo, M., Koskinen, A., Kouvonen, A., Nordin,

M., Oksanen, T., Strandberg, T., Suominen, S. B., Theorell, T., Vahtera, J., Väänänen, A., Virtanen, M., … IPD-Work consortium. (2019). Physical inactivity, cardiometabolic disease, and risk of dementia: An individual-participant meta-analysis. *BMJ (Clinical Research Ed.)*, *365*, l1495. https://doi.org/10.1136/bmj.l1495

Kivipelto, M., Mangialasche, F., & Ngandu, T. (2018). Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. *Nature Reviews Neurology*, *14*(11), 653–666. https://doi.org/10.1038/s41582-018-0070-3

Kivipelto, M., Solomon, A., Ahtiluoto, S., Ngandu, T., Lehtisalo, J., Antikainen, R., Bäckman, L., Hänninen, T., Jula, A., Laatikainen, T., Lindström, J., Mangialasche, F., Nissinen, A., Paajanen, T., Pajala, S., Peltonen, M., Rauramaa, R., Stigsdotter-Neely, A., Strandberg, T., … Soininen, H. (2013). The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER): Study design and progress. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *9*(6), 657–665. https://doi.org/10.1016/j.jalz.2012.09.012

Kobayashi, L. C., Jones, R. N., Briceño, E. M., Rentería, M. A., Zhang, Y., Meijer, E., Langa, K. M., Lee, J., & Gross, A. L. (2024). Cross-national comparisons of later-life cognitive function using data from the Harmonized Cognitive Assessment Protocol (HCAP): Considerations and recommended best practices. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *20*(3), 2273–2281. https://doi.org/10.1002/alz.13694

Koblinsky, N. D., Power, K. A., Middleton, L., Ferland, G., & Anderson, N. D. (2023). The Role of the Gut Microbiome in Diet and Exercise Effects on Cognition: A Review of the Intervention Literature. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *78*(2), 195–205. https://doi.org/10.1093/gerona/glac166

Kodesh, A. (2019). Prevalence and comorbidities of dementia in Israel: A nationally representative cohort study. *International Psychogeriatrics*, *31*(7), 1059–1063. https://doi.org/10.1017/S1041610218001461

Kuan, V., Fraser, H. C., Hingorani, M., Denaxas, S., Gonzalez-Izquierdo, A., Direk, K., Nitsch, D., Mathur, R., Parisinos, C. A., Lumbers, R. T., Sofat, R., Wong, I. C. K., Casas, J. P., Thornton, J. M., Hemingway, H., Partridge, L., & Hingorani, A. D. (2021). Data-driven identification of ageing-related diseases from electronic health records. *Scientific Reports*, *11*(1), 2938. https://doi.org/10.1038/s41598-021-82459-y

Kuh, D., Ben-Shlomo, Y., Lynch, J., Hallqvist, J., & Power, C. (2003). Life course epidemiology. *Journal of Epidemiology & Community Health*, *57*(10), 778–783. https://doi.org/10.1136/jech.57.10.778

Kulmala, J., Ngandu, T., Havulinna, S., Levälahti, E., Lehtisalo, J., Solomon, A., Antikainen, R., Laatikainen, T., Pippola, P., Peltonen, M., Rauramaa, R., Soininen, H., Strandberg, T., Tuomilehto, J., & Kivipelto, M. (2019). The Effect of Multidomain Lifestyle Intervention on Daily Functioning in Older People. *Journal of the American Geriatrics Society*, *67*(6), 1138–1144. https://doi.org/10.1111/jgs.15837

Kuźma, E., Hannon, E., Zhou, A., Lourida, I., Bethel, A., Levine, D. A., Lunnon, K., Thompson-Coon, J., Hyppönen, E., & Llewellyn, D. J. (2018). Which Risk Factors Causally Influence Dementia? A Systematic Review of Mendelian Randomization Studies. *Journal of Alzheimer's Disease*, *64*(1), 181–193. https://doi.org/10.3233/JAD-180013

Lambert, J. C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., DeStafano, A. L., Bis, J. C., Beecham, G. W., Grenier-Boley, B., Russo, G., Thorton-Wells, T. A., Jones, N., Smith, A. V., Chouraki, V., Thomas, C., Ikram, M. A., Zelenika, D., Vardarajan, B. N., … Amouyel, P. (2013). Meta-analysis of 74,046

individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics*, *45*(12), 1452–1458. https://doi.org/10.1038/ng.2802

Langa, K. M., Plassman, B. L., Wallace, R. B., Herzog, A. R., Heeringa, S. G., Ofstedal, M. B., Burke, J. R., Fisher, G. G., Fultz, N. H., Hurd, M. D., Potter, G. G., Rodgers, W. L., Steffens, D. C., Weir, D. R., & Willis, R. J. (2005). The Aging, Demographics, and Memory Study: Study design and methods. *Neuroepidemiology*, *25*(4), 181–191. https://doi.org/10.1159/000087448

Langa, K. M., Ryan, L. H., McCammon, R. J., Jones, R. N., Manly, J. J., Levine, D. A., Sonnega, A., Farron, M., & Weir, D. R. (2019). The Health and Retirement Study Harmonized Cognitive Assessment Protocol Project: Study Design and Methods. *Neuroepidemiology*, *54*(1), 64–74. https://doi.org/10.1159/000503004

Lee, J., Howard, R. S., & Schneider, L. S. (2022). The Current Landscape of Prevention Trials in Dementia. *Neurotherapeutics*, *19*(1), 228–247. https://doi.org/10.1007/s13311-022-01236-5

Lee, S. A., Lim, J. Y., Kim, B.-S., Cho, S. J., Kim, N. Y., Kim, O. B., & Kim, Y. (2015). Comparison of the gut microbiota profile in breast-fed and formula-fed Korean infants using pyrosequencing. *Nutrition Research and Practice*, *9*(3), 242–248. https://doi.org/10.4162/nrp.2015.9.3.242

Legendre, P., & De Cáceres, M. (2013). Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecology Letters*, *16*(8), 951–963. https://doi.org/10.1111/ele.12141

Lei, X., Matovic, D., Leung, W.-Y., Viju, A., & Wuthrich, V. M. (2024). The relationship between social media use and psychosocial outcomes in older adults: A systematic review. *International Psychogeriatrics*, 1–33. https://doi.org/10.1017/S1041610223004519

Leist, A. K., Bar-Haim, E., & Chauvel, L. (2021). Inequality of educational opportunity at time of schooling predicts cognitive functioning in later adulthood. *SSM - Population Health*, *15*, 100837. https://doi.org/10.1016/j.ssmph.2021.100837

Leist, A. K., Klee, M., Kim, J. H., Rehkopf, D. H., Bordas, S. P. A., Muniz-Terrera, G., & Wade, S. (2022). Mapping of machine learning approaches for description, prediction, and causal inference in the social and health sciences. *Science Advances*, *8*(42), eabk1942. https://doi.org/10.1126/sciadv.abk1942

Leonenko, G., Baker, E., Stevenson-Hoare, J., Sierksma, A., Fiers, M., Williams, J., de Strooper, B., & Escott-Price, V. (2021). Identifying individuals with high risk of Alzheimer's disease using polygenic risk scores. *Nature Communications*, *12*(1), 4506. https://doi.org/10.1038/s41467-021-24082-z

Li, J., Joshi, P., Ang, T. F. A., Liu, C., Auerbach, S., Devine, S., & Au, R. (2021). Mid- to Late-Life Body Mass Index and Dementia Risk: 38 Years of Follow-up of the Framingham Study. *American Journal of Epidemiology*, *190*(12), 2503. https://doi.org/10.1093/aje/kwab096

Li, Z., Zhou, J., Liang, H., Ye, L., Lan, L., Lu, F., Wang, Q., Lei, T., Yang, X., Cui, P., & Huang, J. (2022). Differences in Alpha Diversity of Gut Microbiota in Neurological Diseases. *Frontiers in Neuroscience*, *16*, 879318. https://doi.org/10.3389/fnins.2022.879318

Licher, S., Ahmad, S., Karamujić-Čomić, H., Voortman, T., Leening, M. J. G., Ikram, M. A., & Ikram, M. K. (2019). Genetic predisposition, modifiable-risk-factor profile and long-term dementia risk in the general population. *Nature Medicine*, *25*(9), 1364–1369. https://doi.org/10.1038/s41591-019-0547-7

Licher, S., Darweesh, S. K. L., Wolters, F. J., Fani, L., Heshmatollah, A., Mutlu, U., Koudstaal, P. J., Heeringa, J., Leening, M. J. G., Ikram, M. K., & Ikram, M. A.

(2019). Lifetime risk of common neurological diseases in the elderly population. *Journal of Neurology, Neurosurgery & Psychiatry*, *90*(2), 148–156. https://doi.org/10.1136/jnnp-2018-318650

Lin, F. R., Pike, J. R., Albert, M. S., Arnold, M., Burgard, S., Chisolm, T., Couper, D., Deal, J. A., Goman, A. M., Glynn, N. W., Gmelin, T., Gravens-Mueller, L., Hayden, K. M., Huang, A. R., Knopman, D., Mitchell, C. M., Mosley, T., Pankow, J. S., Reed, N. S., … Coresh, J. (2023). Hearing intervention versus health education control to reduce cognitive decline in older adults with hearing loss in the USA (ACHIEVE): A multicentre, randomised controlled trial. *Lancet (London, England)*, *402*(10404), 786–797. https://doi.org/10.1016/S0140-6736(23)01406-X

Lin, H., & Peddada, S. D. (2020). Analysis of compositions of microbiomes with bias correction. *Nature Communications*, *11*(1), 3514. https://doi.org/10.1038/s41467-020-17041-7

Lindeza, P., Rodrigues, M., Costa, J., Guerreiro, M., & Rosa, M. M. (2020). Impact of dementia on informal care: A systematic review of family caregivers' perceptions. *BMJ Supportive & Palliative Care*, bmjspcare-2020-002242. https://doi.org/10.1136/bmjspcare-2020-002242

Littlejohns, T. J., Holliday, J., Gibson, L. M., Garratt, S., Oesingmann, N., Alfaro-Almagro, F., Bell, J. D., Boultwood, C., Collins, R., Conroy, M. C., Crabtree, N., Doherty, N., Frangi, A. F., Harvey, N. C., Leeson, P., Miller, K. L., Neubauer, S., Petersen, S. E., Sellors, J., … Allen, N. E. (2020). The UK Biobank imaging enhancement of 100,000 participants: Rationale, data collection, management and future directions. *Nature Communications*, *11*(1), 2624. https://doi.org/10.1038/s41467-020-15948-9

Liu, P., Wu, L., Peng, G., Han, Y., Tang, R., Ge, J., Zhang, L., Jia, L., Yue, S., Zhou, K., Li, L., Luo, B., & Wang, B. (2019). Altered microbiomes distinguish Alzheimer's disease

from amnestic mild cognitive impairment and health in a Chinese cohort. *Brain, Behavior, and Immunity*, *80*, 633–643. https://doi.org/10.1016/j.bbi.2019.05.008

Livingston, G., & Costafreda, S. G. (2023). Interventions to prevent dementia should target those at high risk. *Lancet (London, England)*, *402*(10404), 750–751. https://doi.org/10.1016/S0140-6736(23)01472-1

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S. G., Dias, A., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Ogunniyi, A., … Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet (London, England)*, *396*(10248), 413–446. https://doi.org/10.1016/S0140-6736(20)30367-6

Lourida, I., Hannon, E., Littlejohns, T. J., Langa, K. M., Hyppönen, E., Kuźma, E., & Llewellyn, D. J. (2019). Association of Lifestyle and Genetic Risk With Incidence of Dementia. *JAMA*, *322*(5), 430–437. https://doi.org/10.1001/jama.2019.9879

Lövdén, M., Fratiglioni, L., Glymour, M. M., Lindenberger, U., & Tucker-Drob, E. M. (2020). Education and Cognitive Functioning Across the Life Span. *Psychological Science in the Public Interest*, *21*(1), 6–41. https://doi.org/10.1177/1529100620920576

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lyall, D. M., Cox, S. R., Lyall, L. M., Celis-Morales, C., Cullen, B., Mackay, D. F., Ward, J., Strawbridge, R. J., McIntosh, A. M., Sattar, N., Smith, D. J., Cavanagh, J., Deary, I. J., & Pell, J. P. (2020). Association between APOE e4 and white matter hyperintensity volume, but not total brain volume or white matter integrity. *Brain*

Imaging and Behavior, *14*(5), 1468–1476. PubMed. https://doi.org/10.1007/s11682-019-00069-9

Madathil, S., Joseph, L., Hardy, R., Rousseau, M.-C., & Nicolau, B. (2018). A Bayesian approach to investigate life course hypotheses involving continuous exposures. *International Journal of Epidemiology*, *47*(5), 1623–1635. https://doi.org/10.1093/ije/dyy107

Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., & Peddada, S. D. (2015). Analysis of composition of microbiomes: A novel method for studying microbial composition. *Microbial Ecology in Health and Disease*, *26*(1), 27663. https://doi.org/10.3402/mehd.v26.27663

Manly, J. J., Jones, R. N., Langa, K. M., Ryan, L. H., Levine, D. A., McCammon, R., Heeringa, S. G., & Weir, D. (2022). Estimating the Prevalence of Dementia and Mild Cognitive Impairment in the US: The 2016 Health and Retirement Study Harmonized Cognitive Assessment Protocol Project. *JAMA Neurology*, *79*(12), 1242–1249. https://doi.org/10.1001/jamaneurol.2022.3543

Marden, J. R., Tchetgen Tchetgen, E. J., Kawachi, I., & Glymour, M. M. (2017). Contribution of Socioeconomic Status at 3 Life-Course Periods to Late-Life Memory Function and Decline: Early and Late Predictors of Dementia Risk. *American Journal of Epidemiology*, *186*(7), 805–814. https://doi.org/10.1093/aje/kwx155

Marinescu, R. V., Oxtoby, N. P., Young, A. L., Bron, E. E., Toga, A. W., Weiner, M. W., Barkhof, F., Fox, N. C., Eshaghi, A., Toni, T., Salaterski, M., Lunina, V., Ansart, M., Durrleman, S., Lu, P., Iddi, S., Li, D., Thompson, W. K., Donohue, M. C., … The Alzheimer's Disease Neuroimaging Initiative. (2021). The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge: Results after 1 Year

Follow-up. *Machine Learning for Biomedical Imaging*, *1*(December 2021), 1–60. https://doi.org/10.59275/j.melba.2021-2dcc

Marmot, M. (2020). Health equity in England: The Marmot review 10 years on. *BMJ*, *368*, m693. https://doi.org/10.1136/bmj.m693

Marmot, M., & Wilkinson, R. G. (2005). Social determinants of health in older age. In M. Marmot & R. G. Wilkinson (Eds.), *Social Determinants of Health* (2nd edn, pp. 267–296). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198565895.003.13

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Martins, I. J., Hone, E., Foster, J. K., Sünram-Lea, S. I., Gnjec, A., Fuller, S. J., Nolan, D., Gandy, S. E., & Martins, R. N. (2006). Apolipoprotein E, cholesterol metabolism, diabetes, and the convergence of risk factors for Alzheimer's disease and cardiovascular disease. *Molecular Psychiatry*, *11*(8), 721–736. https://doi.org/10.1038/sj.mp.4001854

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr, Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (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), 263–269. https://doi.org/10.1016/j.jalz.2011.03.005

McMartin, A., & Conley, D. (2020). Commentary: Mendelian randomization and education-Challenges remain. *International Journal of Epidemiology*, *49*(4), 1193–1206. https://doi.org/10.1093/ije/dyaa160

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

Milcent, C., & Zbiri, S. (2018). Prenatal care and socioeconomic status: Effect on cesarean delivery. *Health Economics Review*, *8*(1), 7. https://doi.org/10.1186/s13561-018-0190-x

Miller, G. E., Engen, P. A., Gillevet, P. M., Shaikh, M., Sikaroodi, M., Forsyth, C. B., Mutlu, E., & Keshavarzian, A. (2016). Lower Neighborhood Socioeconomic Status Associated with Reduced Diversity of the Colonic Microbiota in Healthy Adults. *PloS One*, *11*(2), e0148952. https://doi.org/10.1371/journal.pone.0148952

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., Bartsch, A. J., Jbabdi, S., Sotiropoulos, S. N., Andersson, J. L. R., Griffanti, L., Douaud, G., Okell, T. W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., … Smith, S. M. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, *19*(11), 1523–1536. https://doi.org/10.1038/nn.4393

Mjørud, M., Selbæk, G., Bjertness, E., Edwin, T. H., Engedal, K., Knapskog, A.-B., & Strand, B. H. (2020). Time from dementia diagnosis to nursing-home admission and death among persons with dementia: A multistate survival analysis. *PLOS ONE*, *15*(12), e0243513. https://doi.org/10.1371/journal.pone.0243513

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). Sustainable data analysis with Snakemake. *F1000Research*, *10*, 33. https://doi.org/10.12688/f1000research.29032.2

Mole, J. P., Fasano, F., Evans, J., Sims, R., Hamilton, D. A., Kidd, E., & Metzler-Baddeley, C. (2020). Genetic risk of dementia modifies obesity effects on white matter myelin in cognitively healthy adults. *Neurobiology of Aging*, *94*, 298–310. https://doi.org/10.1016/j.neurobiolaging.2020.06.014

Morais, L. H., Schreiber, H. L., & Mazmanian, S. K. (2021). The gut microbiota–brain axis in behaviour and brain disorders. *Nature Reviews Microbiology*, *19*(4), 241–255. https://doi.org/10.1038/s41579-020-00460-0

Mukadam, N., Marston, L., Lewis, G., Mathur, R., Lowther, E., Rait, G., & Livingston, G. (2023). South Asian, Black and White ethnicity and the effect of potentially modifiable risk factors for dementia: A study in English electronic health records. *PLOS ONE*, *18*(10), e0289893. https://doi.org/10.1371/journal.pone.0289893

Mukadam, N., Sommerlad, A., Huntley, J., & Livingston, G. (2019). Population attributable fractions for risk factors for dementia in low-income and middle-income countries: An analysis using cross-sectional survey data. *The Lancet Global Health*, *7*(5), e596–e603. https://doi.org/10.1016/S2214-109X(19)30074-9

Mullin, D. S., Stirland, L. E., Buchanan, E., Convery, C.-A., Cox, S. R., Deary, I. J., Giuntoli, C., Greer, H., Page, D., Robertson, E., Shenkin, S. D., Szalek, A., Taylor, A., Weatherdon, G., Wilkinson, T., & Russ, T. C. (2023). Identifying dementia using medical data linkage in a longitudinal cohort study: Lothian Birth Cohort 1936. *BMC Psychiatry*, *23*(1), 303. https://doi.org/10.1186/s12888-023-04797-7

Murphy, T. E., McAvay, G., Carriero, N. J., Gross, C. P., Tinetti, M. E., Allore, H. G., & Lin, H. (2012). Deaths observed in Medicare beneficiaries: Average attributable fraction and its longitudinal extension for many diseases. *Statistics in Medicine*, *31*(27), 3313–3319. https://doi.org/10.1002/sim.5337

Nagpal, R., Neth, B. J., Wang, S., Craft, S., & Yadav, H. (2019). Modified Mediterranean-ketogenic diet modulates gut microbiome and short-chain fatty acids in association with Alzheimer's disease markers in subjects with mild cognitive impairment. *EBioMedicine*, *47*, 529–542. https://doi.org/10.1016/j.ebiom.2019.08.032

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nasreddine, Z. S., Rossetti, H., Phillips, N., Chertkow, H., Lacritz, L., Cullum, M., & Weiner, M. (2012). Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sampleAuthor Response. *Neurology*, *78*(10), 765–766. https://doi.org/10.1212/01.wnl.0000413072.54070.a3

Ngandu, T., Lehtisalo, J., Solomon, A., Levälahti, E., Ahtiluoto, S., Antikainen, R., Bäckman, L., Hänninen, T., Jula, A., Laatikainen, T., Lindström, J., Mangialasche, F., Paajanen, T., Pajala, S., Peltonen, M., Rauramaa, R., Stigsdotter-Neely, A., Strandberg, T., Tuomilehto, J., … Kivipelto, M. (2015). A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): A randomised controlled trial. *Lancet (London, England)*, *385*(9984), 2255–2263. https://doi.org/10.1016/S0140-6736(15)60461-5

Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T. T., Al Hamad, H., Alahdab, F., Alanezi, F. M., Alipour, V., Almustanyir, S., Amu, H., Ansari, I., Arabloo, J., Ashraf, T., … Vos, T. (2022). Estimation of the global prevalence of dementia in 2019 and

forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, *7*(2), e105–e125. https://doi.org/10.1016/S2468-2667(21)00249-8

Nobis, L., Manohar, S. G., Smith, S. M., Alfaro-Almagro, F., Jenkinson, M., Mackay, C. E., & Husain, M. (2019). Hippocampal volume across age: Nomograms derived from over 19,700 people in UK Biobank. *NeuroImage: Clinical*, *23*, 101904. https://doi.org/10.1016/j.nicl.2019.101904

OECD. (2018). *Health at a Glance: Europe 2018: State of Health in the EU Cycle*. OECD Publishing. https://doi.org/10.1787/health_glance_eur-2018-en

Oksanen J., Simpson G. L., Blanchet F. G., Kindt R., Legendre P., Minchin P., O'Hara R., Solymos P., Stevens M., Szoecs E., Wagner H., Barbour M., Bedward M., Bolker B., Borcard D., Carvalho G., Chirico M., De Caceres M, Durand S., … Weedon J. (2022). *vegan: Community Ecology Package* (R package version 2.6-2) [Computer software].

Oosterhuis, E. J., Slade, K., May, P. J. C., & Nuttall, H. E. (2023). Toward an Understanding of Healthy Cognitive Aging: The Importance of Lifestyle in Cognitive Reserve and the Scaffolding Theory of Aging and Cognition. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *78*(5), 777–788. https://doi.org/10.1093/geronb/gbac197

Pan, K.-Y., Xu, W., Mangialasche, F., Grande, G., Fratiglioni, L., & Wang, H.-X. (2020). The role of Apolipoprotein E epsilon4 in the association between psychosocial working conditions and dementia. *Aging (Albany NY)*, *12*(4), 3730–3746. https://doi.org/10.18632/aging.102843

Pang, M., Zhu, L., Gabelle, A., Gafson, A. R., Platt, R. W., Galvin, J. E., Krolak-Salmon, P., Rubino, I., de Moor, C., Belachew, S., & Shen, C. (2023). Effect of reduction in brain

amyloid levels on change in cognitive and functional decline in randomized clinical trials: An instrumental variable meta-analysis. *Alzheimer's & Dementia*, *19*(4), 1292–1299. https://doi.org/10.1002/alz.12768

Pearl, J. (2001). Direct and indirect effects. In J. Breese & D. Koller (Eds.), *Uncertainty in artificial intelligence, proceedings of the seventeenth conference* (pp. 411–420). Morgan Kaufmann. https://doi.org/10.48550/arXiv.1301.2300

Pekkala, T., Hall, A., Lötjönen, J., Mattila, J., Soininen, H., Ngandu, T., Laatikainen, T., Kivipelto, M., & Solomon, A. (2017). Development of a Late-Life Dementia Prediction Index with Supervised Machine Learning in the Population-Based CAIDE Study. *Journal of Alzheimer's Disease: JAD*, *55*(3), 1055–1067. https://doi.org/10.3233/JAD-160560

Perez-Nievas, B. G., Stein, T. D., Tai, H.-C., Dols-Icardo, O., Scotton, T. C., Barroeta-Espar, I., Fernandez-Carballo, L., de Munain, E. L., Perez, J., Marquie, M., Serrano-Pozo, A., Frosch, M. P., Lowe, V., Parisi, J. E., Petersen, R. C., Ikonomovic, M. D., López, O. L., Klunk, W., Hyman, B. T., & Gómez-Isla, T. (2013). Dissecting phenotypic traits linked to human resilience to Alzheimer's pathology. *Brain*, *136*(8), 2510–2526. https://doi.org/10.1093/brain/awt171

Peters, R., Booth, A., Rockwood, K., Peters, J., D'Este, C., & Anstey, K. J. (2019). Combining modifiable risk factors and risk of dementia: A systematic review and meta-analysis. *BMJ Open*, *9*(1), e022846. https://doi.org/10.1136/bmjopen-2018-022846

Peters, R., Ee, N., Peters, J., Booth, A., Mudway, I., & Anstey, K. J. (2019). Air Pollution and Dementia: A Systematic Review. *Journal of Alzheimer's Disease*, *70*(s1), 145–163. https://doi.org/10.3233/JAD-180631

Petersen, R. C., Lopez, O., Armstrong, M. J., Getchius, T. S. D., Ganguli, M., Gloss, D., Gronseth, G. S., Marson, D., Pringsheim, T., Day, G. S., Sager, M., Stevens, J., & Rae-Grant, A. (2018). Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Neurology*, *90*(3), 126–135. https://doi.org/10.1212/WNL.0000000000004826

Phillips, S., Watt, R., Atkinson, T., Rajan, S., Hayhoe, A., Savva, G. M., Hornberger, M., Burton, B. J. L., Saada, J., Cambell-Kelly, M., Rushbrook, S., & Carding, S. R. (2022). A protocol paper for the MOTION Study-A longitudinal study in a cohort aged 60 years and older to obtain mechanistic knowledge of the role of the gut microbiome during normal healthy ageing in order to develop strategies that will improve lifelong health and wellbeing. *PloS One*, *17*(11), e0276118. https://doi.org/10.1371/journal.pone.0276118

Piscopo, P., Crestini, A., Carbone, E., Rivabene, R., Ancidoni, A., Lo Giudice, M., Corbo, M., Vanacore, N., & Lacorte, E. (2022). A systematic review on drugs for synaptic plasticity in the treatment of dementia. *Ageing Research Reviews*, *81*, 101726. https://doi.org/10.1016/j.arr.2022.101726

Porsteinsson, A. P., Isaacson, R. S., Knox, S., Sabbagh, M. N., & Rubino, I. (2021). Diagnosis of Early Alzheimer's Disease: Clinical Practice in 2021. *The Journal of Prevention of Alzheimer's Disease*, *8*(3), 371–386. https://doi.org/10.14283/jpad.2021.23

Powell, W. R., Buckingham, W. R., Larson, J. L., Vilen, L., Yu, M., Salamat, M. S., Bendlin, B. B., Rissman, R. A., & Kind, A. J. H. (2020). Association of Neighborhood-Level Disadvantage With Alzheimer Disease Neuropathology. *JAMA Network Open*, *3*(6), e207559. https://doi.org/10.1001/jamanetworkopen.2020.7559

Prince, M. J., Ali, G.-C., Guerchet, M., Prina, A. M., Albanese, E., & Wu, Y.-T. (2016). Recent global trends in the prevalence and incidence of dementia, and survival with dementia. *Alzheimer's Research & Therapy*, *8*(1), 23. https://doi.org/10.1186/s13195-016-0188-8

Prince, M. J., Knapp, M., Guerchet, M., McCrone, P., Prina, M., Comas-Herrera, A., Wittenberg, R., Adelaja, B., Hu, B., King, D., Rehill, A., & Salimkumar, D. (2014). *Dementia UK: Update (Second edition)*. Alzheimer's Society. https://www.alzheimers.org.uk/sites/default/files/migrate/downloads/dementia_uk_update.pdf

Prince, M. J., Reischies, F., Beekman, A. T. F., Fuhrer, R., Jonker, C., Kivela, S.-L., Lawlor, B. A., Lobo, A., Magnusson, H., Fichter, M., Oyen, H. V., Roelands, M., Skoog, I., Turrina, C., & Copeland, J. R. M. (1999). Development of the EURO–D scale – a European Union initiative to compare symptoms of depression in 14 European centres. *The British Journal of Psychiatry*, *174*(4), 330–338. https://doi.org/10.1192/bjp.174.4.330

Prince, M. J., Wimo, A., Guerchet, M., Ali, G.-C., Wu, Y.-T., Prina, M., & Alzheimer's Disease International. (2015). *World Alzheimer Report 2015. The Global Impact of Dementia. An Analysis of Prevalence, Incidence, Cost and Trends*. Alzheimer's Disease International. https://www.alzint.org/u/WorldAlzheimerReport2015.pdf

Qiu, C., & Fratiglioni, L. (2015). A major role for cardiovascular burden in age-related cognitive decline. *Nature Reviews Cardiology*, *12*(5), 267–277. https://doi.org/10.1038/nrcardio.2014.223

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data

processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team. (2022). *R: A Language and Environment for Statistical Computing* (R version 4.2.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Radjabzadeh, D., Bosch, J. A., Uitterlinden, A. G., Zwinderman, A. H., Ikram, M. A., van Meurs, J. B. J., Luik, A. I., Nieuwdorp, M., Lok, A., van Duijn, C. M., Kraaij, R., & Amin, N. (2022). Gut microbiome-wide association study of depressive symptoms. *Nature Communications*, *13*(1), 7128. https://doi.org/10.1038/s41467-022-34502-3

Ranson, J. M., Lourida, I., Hannon, E., Littlejohns, T. J., Ballard, C., Langa, K. M., Hyppönen, E., Kuzma, E., & Llewellyn, D. J. (2020). Genetic risk, education and incidence of dementia. *Alzheimer's & Dementia*, *16*(S10), e045903. https://doi.org/10.1002/alz.045903

Richiardi, L., Bellocco, R., & Zugna, D. (2013). Mediation analysis in epidemiology: Methods, interpretation and bias. *International Journal of Epidemiology*, *42*(5), 1511–1519. https://doi.org/10.1093/ije/dyt127

Robinson, J. L., Lee, E. B., Xie, S. X., Rennert, L., Suh, E., Bredenberg, C., Caswell, C., Van Deerlin, V. M., Yan, N., Yousef, A., Hurtig, H. I., Siderowf, A., Grossman, M., McMillan, C. T., Miller, B., Duda, J. E., Irwin, D. J., Wolk, D., Elman, L., … Trojanowski, J. Q. (2018). Neurodegenerative disease concomitant proteinopathies are prevalent, age-related and APOE4-associated. *Brain*, *141*(7), 2181–2193. https://doi.org/10.1093/brain/awy146

Rogers, M. A. M., Plassman, B. L., Kabeto, M., Fisher, G. G., McArdle, J. J., Llewellyn, D. J., Potter, G. G., & Langa, K. M. (2009). Parental education and late-life dementia in

the United States. *Journal of Geriatric Psychiatry and Neurology*, *22*(1), 71–80. https://doi.org/10.1177/0891988708328220

Rojas-Saunero, L. P., Young, J. G., Didelez, V., Ikram, M. A., & Swanson, S. A. (2021). *Choosing questions before methods in dementia research with competing events and causal goals*. medRxiv. https://doi.org/10.1101/2021.06.01.21258142

Rosenberg, A., Ngandu, T., Rusanen, M., Antikainen, R., Bäckman, L., Havulinna, S., Hänninen, T., Laatikainen, T., Lehtisalo, J., Levälahti, E., Lindström, J., Paajanen, T., Peltonen, M., Soininen, H., Stigsdotter-Neely, A., Strandberg, T., Tuomilehto, J., Solomon, A., & Kivipelto, M. (2018). Multidomain lifestyle intervention benefits a large elderly population at risk for cognitive decline and dementia regardless of baseline characteristics: The FINGER trial. *Alzheimer's & Dementia*, *14*(3), 263–270. https://doi.org/10.1016/j.jalz.2017.09.006

Rossetti, H. C., Lacritz, L. H., Cullum, C. M., & Weiner, M. F. (2011). Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. *Neurology*, *77*(13), 1272–1275. https://doi.org/10.1212/WNL.0b013e318230208a

Saji, N., Murotani, K., Hisada, T., Tsuduki, T., Sugimoto, T., Kimura, A., Niida, S., Toba, K., & Sakurai, T. (2019). The relationship between the gut microbiome and mild cognitive impairment in patients without dementia: A cross-sectional study conducted in Japan. *Scientific Reports*, *9*(1), 19227. https://doi.org/10.1038/s41598-019-55851-y

Saji, N., Niida, S., Murotani, K., Hisada, T., Tsuduki, T., Sugimoto, T., Kimura, A., Toba, K., & Sakurai, T. (2019). Analysis of the relationship between the gut microbiome and dementia: A cross-sectional study conducted in Japan. *Scientific Reports*, *9*(1), 1008. https://doi.org/10.1038/s41598-018-38218-7

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B.,

Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. https://doi.org/10.1128/AEM.01541-09

Schneeweis, N., Skirbekk, V., & Winter-Ebmer, R. (2014). Does Education Improve Cognitive Performance Four Decades After School Completion? *Demography*, *51*(2), 619–643. https://doi.org/10.1007/s13524-014-0281-1

Schneider, J. A., Arvanitakis, Z., Bang, W., & Bennett, D. A. (2007). Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology*, *69*(24), 2197–2204. https://doi.org/10.1212/01.wnl.0000271090.28148.24

Schoenfeld, D. (1982). Partial Residuals for the Proportional Hazards Regression Model. *Biometrika*, *69*(1), 239–241. https://doi.org/10.1093/biomet/69.1.239

Schwarzinger, M., & Dufouil, C. (2022). Forecasting the prevalence of dementia. *The Lancet Public Health*, *7*(2), e94–e95. https://doi.org/10.1016/S2468-2667(21)00277-2

Seblova, D., Fischer, M., Fors, S., Johnell, K., Karlsson, M., Nilsson, T., Svensson, A. C., Lövdén, M., & Lager, A. (2021). Does Prolonged Education Causally Affect Dementia Risk When Adult Socioeconomic Status Is Not Altered? A Swedish Natural Experiment in 1.3 Million Individuals. *American Journal of Epidemiology*, *190*(5), 817–826. https://doi.org/10.1093/aje/kwaa255

Sevi, B., Gutiérrez, Á., & Muniz-Terrera, G. (2024). Underlining neighbourhood perception: A possible risk factor for dementia that deserves more attention. *Brain Communications*, *6*(2), fcae037. https://doi.org/10.1093/braincomms/fcae037

Shi, B., Choirat, C., Coull, B. A., VanderWeele, T. J., & Valeri, L. (2021). CMAverse: A Suite of Functions for Reproducible Causal Mediation Analyses. *Epidemiology*, *32*(5), e20. https://doi.org/10.1097/EDE.0000000000001378

Smith, J. R., Sharrett, A. R., Pike, J. R., Gottesman, R. F., Knopman, D. S., Lee, M., Lutsey, P. L., Palta, P., Windham, B. G., Coresh, J., & Deal, J. A. (2023). Dementia occurring over a 32-year follow-up attributable to hypertension observed at different ages: Implications for dementia prevention. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *19*(8), 3435–3447. https://doi.org/10.1002/alz.12984

Smith, S., M., Alfaro-Almagro, F., & Miller, K. L. (2020, December 1). *UK Biobank Brain Imaging Documentation*. https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf

Snyder, H. M., Corriveau, R. A., Craft, S., Faber, J. E., Greenberg, S. M., Knopman, D., Lamb, B. T., Montine, T. J., Nedergaard, M., Schaffer, C. B., Schneider, J. A., Wellington, C., Wilcock, D. M., Zipfel, G. J., Zlokovic, B., Bain, L. J., Bosetti, F., Galis, Z. S., Koroshetz, W., & Carrillo, M. C. (2015). Vascular contributions to cognitive impairment and dementia including Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *11*(6), 710–717. https://doi.org/10.1016/j.jalz.2014.10.008

Solomon, A., Turunen, H., Ngandu, T., Peltonen, M., Levälahti, E., Helisalmi, S., Antikainen, R., Bäckman, L., Hänninen, T., Jula, A., Laatikainen, T., Lehtisalo, J., Lindström, J., Paajanen, T., Pajala, S., Stigsdotter-Neely, A., Strandberg, T., Tuomilehto, J., Soininen, H., & Kivipelto, M. (2018). Effect of the Apolipoprotein E Genotype on Cognitive Change During a Multidomain Lifestyle Intervention: A Subgroup Analysis of a Randomized Clinical Trial. *JAMA Neurology*, *75*(4), 462–470. https://doi.org/10.1001/jamaneurol.2017.4365

Sommerlad, A., Perera, G., Mueller, C., Singh-Manoux, A., Lewis, G., Stewart, R., & Livingston, G. (2019). Hospitalisation of people with dementia: Evidence from

English electronic health records from 2008 to 2016. *European Journal of Epidemiology*, *34*(6), 567–577. https://doi.org/10.1007/s10654-019-00481-x

Stakos, D. A., Stamatelopoulos, K., Bampatsias, D., Sachse, M., Zormpas, E., Vlachogiannis, N. I., Tual-Chalot, S., & Stellos, K. (2020). The Alzheimer's Disease Amyloid-Beta Hypothesis in Cardiovascular Aging and Disease: JACC Focus Seminar. *Journal of the American College of Cardiology*, *75*(8), 952–967. https://doi.org/10.1016/j.jacc.2019.12.033

Stamatakis, E., Owen, K. B., Shepherd, L., Drayton, B., Hamer, M., & Bauman, A. E. (2021). Is Cohort Representativeness Passé? Poststratified Associations of Lifestyle Risk Factors with Mortality in the UK Biobank. *Epidemiology (Cambridge, Mass.)*, *32*(2), 179–188. https://doi.org/10.1097/EDE.0000000000001316

Stephen, R., Barbera, M., Peters, R., Ee, N., Zheng, L., Lehtisalo, J., Kulmala, J., Håkansson, K., Chowdhary, N., Dua, T., Solomon, A., Anstey, K. J., & Kivipelto, M. (2021). Development of the First WHO Guidelines for Risk Reduction of Cognitive Decline and Dementia: Lessons Learned and Future Directions. *Frontiers in Neurology*, *12*, 763573. https://doi.org/10.3389/fneur.2021.763573

Stern, Y., Arenaza-Urquijo, E. M., Bartrés-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., Ewers, M., Franzmeier, N., Kempermann, G., Kremen, W. S., Okonkwo, O., Scarmeas, N., Soldan, A., Udeh-Momoh, C., Valenzuela, M., Vemuri, P., Vuoksimaa, E., & and the Reserve, R. and P. F. P. E. D. and C. F. W. (2020). Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimer's & Dementia*, *16*(9), 1305–1311. https://doi.org/10.1016/j.jalz.2018.07.219

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young,

A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Sutin, A. R., Stephan, Y., & Terracciano, A. (2019). Verbal fluency and risk of dementia. *International Journal of Geriatric Psychiatry*, *34*(6), 863–867. https://doi.org/10.1002/gps.5081

Szanton, S. L., Gill, J. M., & Allen, J. K. (2005). Allostatic Load: A Mechanism of Socioeconomic Health Disparities? *Biological Research for Nursing*, *7*(1), 7–15. https://doi.org/10.1177/1099800405278216

Tai, X. Y., Veldsman, M., Lyall, D. M., Littlejohns, T. J., Langa, K. M., Husain, M., Ranson, J., & Llewellyn, D. J. (2022). Cardiometabolic multimorbidity, genetic risk, and dementia: A prospective cohort study. *The Lancet Healthy Longevity*, *3*(6), e428–e436. https://doi.org/10.1016/S2666-7568(22)00117-9

Tchetgen, E. J. T., & VanderWeele, T. J. (2014). On identification of natural direct effects when a confounder of the mediator is directly affected by exposure. *Epidemiology (Cambridge, Mass.)*, *25*(2), 282–291. https://doi.org/10.1097/EDE.0000000000000054

Tennant, P. W. G., Murray, E. J., Arnold, K. F., Berrie, L., Fox, M. P., Gadd, S. C., Harrison, W. J., Keeble, C., Ranker, L. R., Textor, J., Tomova, G. D., Gilthorpe, M. S., & Ellison, G. T. H. (2020). Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: Review and recommendations. *International Journal of Epidemiology*, *50*(2), 620–632. https://doi.org/10.1093/ije/dyaa213

The Lancet Neurology. (2024). Dementia diagnosis in the anti-amyloid era. *The Lancet Neurology*, *23*(3), 219. https://doi.org/10.1016/S1474-4422(24)00041-3

Therneau, M. (2021). *A Package for Survival Analysis in R* (R package version 3.2-11) [Computer software]. https://CRAN.R-project.org

Thunell, J., Chen, Y., Joyce, G., Barthold, D., Shekelle, P. G., Brinton, R. D., & Zissimopoulos, J. (2021). Drug Therapies for Chronic Conditions and Risk of Alzheimer's Disease and Related Dementias: A Scoping Review. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *17*(1), 41–48. https://doi.org/10.1002/alz.12175

Tosto, G., Zimmerman, M. E., Carmichael, O. T., Brickman, A. M., & for the Alzheimer's Disease Neuroimaging Initiative. (2014). Predicting Aggressive Decline in Mild Cognitive Impairment: The Importance of White Matter Hyperintensities. *JAMA Neurology*, *71*(7), 872–877. https://doi.org/10.1001/jamaneurol.2014.667

Townsend, P. (1987). Deprivation. *Journal of Social Policy*, *16*(2), 125–146. https://doi.org/10.1017/S0047279400020341

Twait, E. L., Andaur Navarro, C. L., Gudnason, V., Hu, Y.-H., Launer, L. J., & Geerlings, M. I. (2023). Dementia prediction in the general population using clinically accessible variables: A proof-of-concept study using machine learning. The AGES-Reykjavik study. *BMC Medical Informatics and Decision Making*, *23*(1), 168. https://doi.org/10.1186/s12911-023-02244-x

UNESCO Institute for Statistics. (2012). *International Standard Classification of Education 2011*. http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf

United Nations, Department of Economic and Social Affairs, Population Division. (2017). *World Population Prospects: The 2017 Revision, Volume I: Comprehensive Tables (ST/ESA/SER.A/399)*. United Nations. https://doi.org/10.18356/9789210001014

Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, *18*(2), 137–150. https://doi.org/10.1037/a0031034

van Alten, S., Domingue, B. W., Galama, T., & Marees, A. T. (2022). *Reweighting the UK Biobank to reflect its underlying sampling population substantially reduces pervasive selection bias due to volunteering*. medRxiv. https://doi.org/10.1101/2022.05.16.22275048

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

van de Veen, D., Bakker, C., Peetoom, K., Pijnenburg, Y., Papma, J. M., Group, T. P. S., de Vugt, M., & Koopmans, R. (2021). An Integrative Literature Review on the Nomenclature and Definition of Dementia at a Young Age. *Journal of Alzheimer's Disease*, *83*(4), 1891–1916. https://doi.org/10.3233/JAD-210458

van den Goorbergh, R., van Smeden, M., Timmerman, D., & Van Calster, B. (2022). The harm of class imbalance corrections for risk prediction models: Illustration and simulation using logistic regression. *Journal of the American Medical Informatics Association: JAMIA*, *29*(9), 1525–1534. https://doi.org/10.1093/jamia/ocac093

van Dyck, C. H., Swanson, C. J., Aisen, P., Bateman, R. J., Chen, C., Gee, M., Kanekiyo, M., Li, D., Reyderman, L., Cohen, S., Froelich, L., Katayama, S., Sabbagh, M., Vellas, B., Watson, D., Dhadda, S., Irizarry, M., Kramer, L. D., & Iwatsubo, T. (2023). Lecanemab in Early Alzheimer's Disease. *New England Journal of Medicine*, *388*(1), 9–21. https://doi.org/10.1056/NEJMoa2212948

van Gennip, A. C. E., van Sloten, T. T., Fayosse, A., Sabia, S., & Singh-Manoux, A. (2024). Age at cardiovascular disease onset, dementia risk, and the role of lifestyle factors. *Alzheimer's & Dementia*, *20*(3), 1693–1702. https://doi.org/10.1002/alz.13562

VanderWeele, T. J. (2011). Causal mediation analysis with survival data. *Epidemiology (Cambridge, Mass.)*, *22*(4), 582–585. https://doi.org/10.1097/EDE.0b013e31821db37e

VanderWeele, T. J. (2014). A unification of mediation and interaction: A four-way decomposition. *Epidemiology (Cambridge, Mass.)*, *25*(5), 749. https://doi.org/10.1097/EDE.0000000000000121

VanderWeele, T. J., & Vansteelandt, S. (2014). Mediation Analysis with Multiple Mediators. *Epidemiologic Methods*, *2*(1), 95–115. https://doi.org/10.1515/em-2012-0010

Vernooij-Dassen, M. J. F. J., Moniz-Cook, E. D., Woods, R. T., Lepeleire, J. D., Leuschner, A., Zanetti, O., Rotrou, J. de, Kenny, G., Franco, M., Peters, V., & Iliffe, S. (2005). Factors affecting timely recognition and diagnosis of dementia across Europe: From awareness to stigma. *International Journal of Geriatric Psychiatry*, *20*(4), 377–386. https://doi.org/10.1002/gps.1302

Villain, N., Planche, V., & Levy, R. (2022). High-clearance anti-amyloid immunotherapies in Alzheimer's disease. Part 1: Meta-analysis and review of efficacy and safety data, and medico-economical aspects. *Revue Neurologique*, *178*(10), 1011–1030. https://doi.org/10.1016/j.neurol.2022.06.012

Vogt, N. M., Kerby, R. L., Dill-McFarland, K. A., Harding, S. J., Merluzzi, A. P., Johnson, S. C., Carlsson, C. M., Asthana, S., Zetterberg, H., Blennow, K., Bendlin, B. B., & Rey, F. E. (2017). Gut microbiome alterations in Alzheimer's disease. *Scientific Reports*, *7*(1), 13537. https://doi.org/10.1038/s41598-017-13601-y

Walsh, S., Wallace, L., Mukadam, N., Mytton, O., Lafortune, L., Wills, W., & Brayne, C. (2023). What is a population-level approach to prevention, and how could we apply it to dementia risk reduction? *Public Health*, *225*, 22–27. https://doi.org/10.1016/j.puhe.2023.09.019

Wampach, L., Heintz-Buschart, A., Fritz, J. V., Ramiro-Garcia, J., Habier, J., Herold, M., Narayanasamy, S., Kaysen, A., Hogan, A. H., Bindl, L., Bottu, J., Halder, R., Sjöqvist, C., May, P., Andersson, A. F., de Beaufort, C., & Wilmes, P. (2018). Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential. *Nature Communications*, *9*(1), 5091. https://doi.org/10.1038/s41467-018-07631-x

Wang, C., Zheng, D., Weng, F., Jin, Y., & He, L. (2022). Sodium butyrate ameliorates the cognitive impairment of Alzheimer's disease by regulating the metabolism of astrocytes. *Psychopharmacology*, *239*(1), 215–227. https://doi.org/10.1007/s00213-021-06025-0

Wang, X., Wang, Z., Cao, J., Dong, Y., & Chen, Y. (2023). Gut microbiota-derived metabolites mediate the neuroprotective effect of melatonin in cognitive impairment induced by sleep deprivation. *Microbiome*, *11*(1), 17. https://doi.org/10.1186/s40168-022-01452-3

Wardlaw, J. M., Valdés Hernández, M. C., & Muñoz-Maniega, S. (2015). What are White Matter Hyperintensities Made of? *Journal of the American Heart Association: Cardiovascular and Cerebrovascular Disease*, *4*(6), e001140. https://doi.org/10.1161/JAHA.114.001140

Weiss, J., Puterman, E., Prather, A. A., Ware, E. B., & Rehkopf, D. H. (2020). A data-driven prospective study of dementia among older adults in the United States. *PLOS ONE*, *15*(10), e0239994. https://doi.org/10.1371/journal.pone.0239994

Weißbecker, C., Schnabel, B., & Heintz-Buschart, A. (2020). Dadasnake, a Snakemake implementation of DADA2 to process amplicon sequencing data for microbial ecology. *GigaScience*, *9*(12), giaa135. https://doi.org/10.1093/gigascience/giaa135

Welberry, H. J., Tisdell, C. C., Huque, M. H., & Jorm, L. R. (2023). Have We Been Underestimating Modifiable Dementia Risk? An Alternative Approach for Calculating the Combined Population Attributable Fraction for Modifiable Dementia Risk Factors. *American Journal of Epidemiology*, *192*(10), 1763–1771. https://doi.org/10.1093/aje/kwad138

Weston, P. S. J., Poole, T., O'Connor, A., Heslegrave, A., Ryan, N. S., Liang, Y., Druyeh, R., Mead, S., Blennow, K., Schott, J. M., Frost, C., Zetterberg, H., & Fox, N. C. (2019). Longitudinal measurement of serum neurofilament light in presymptomatic familial Alzheimer's disease. *Alzheimer's Research & Therapy*, *11*, 19. https://doi.org/10.1186/s13195-019-0472-5

Weuve, J., Proust-Lima, C., Power, M. C., Gross, A. L., Hofer, S. M., Thiébaut, R., Chêne, G., Glymour, M. M., & Dufouil, C. (2015). Guidelines for reporting methodological challenges and evaluating potential bias in dementia research. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *11*(9), 1098–1109. https://doi.org/10.1016/j.jalz.2015.06.1885

Wild, C. P. (2012). The exposome: From concept to utility. *International Journal of Epidemiology*, *41*(1), 24–32. https://doi.org/10.1093/ije/dyr236

Wilkinson, T., Schnier, C., Bush, K., Rannikmäe, K., Henshall, D. E., Lerpiniere, C., Allen, N. E., Flaig, R., Russ, T. C., Bathgate, D., Pal, S., O'Brien, J. T., Sudlow, C. L. M., & on behalf of Dementias Platform UK and UK Biobank. (2019). Identifying dementia outcomes in UK Biobank: A validation study of primary care, hospital admissions and

mortality data. *European Journal of Epidemiology*, *34*(6), 557–565. https://doi.org/10.1007/s10654-019-00499-1

Wilmes, P., Trezzi, J.-P., Aho, V., Jäger, C., Schade, S., Janzen, A., Hickl, O., Kunath, B., Thomas, M., Schmit, K., Garcia, P., Sciortino, A., Martin-Gallausiaux, C., Halder, R., Huarte, O. U., Heurtaux, T., Heins-Marroquin, U., Gomez-Giro, G., Weidenbach, K., … Mollenhauer, B. (2022). *An archaeal compound as a driver of Parkinson's disease pathogenesis*. Research Square. https://doi.org/10.21203/rs.3.rs-1827631/v1

Wimo, A., Jönsson, L., Johansson, G., & Winblad, B. (2023). Lecanemab: The Price of a Breakthrough. *touchREVIEWS in Neurology*, *19*(1), 14–15. https://doi.org/10.17925/USN.2023.19.1.14

Witte, A. V., Fobker, M., Gellner, R., Knecht, S., & Flöel, A. (2009). Caloric restriction improves memory in elderly humans. *Proceedings of the National Academy of Sciences*, *106*(4), 1255–1260. https://doi.org/10.1073/pnas.0808587106

Wittmann, F. G., Pabst, A., Zülke, A., Luppa, M., Blotenberg, I., Cardona, M. I., Bauer, A., Fuchs, S., Zöllinger, I., Sanftenberg, L., Brettschneider, C., Döhring, J., Lunden, L., Czock, D., Wiese, B., Thyrian, J. R., Hoffmann, W., Frese, T., Gensichen, J., … Riedel-Heller, S. G. (2024). Who Benefited the Most? Effectiveness of a Lifestyle Intervention Against Cognitive Decline in Older Women and Men—Secondary Analysis of the AgeWell.de-trial. *The Journal of Prevention of Alzheimer's Disease*, *11*(2), 348–355. https://doi.org/10.14283/jpad.2024.13

World Health Organization. (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*.

World Health Organization. (2017). *Global action plan on the public health response to dementia 2017–2025*. https://iris.who.int/handle/10665/259615

World Health Organization. (2019). *Risk reduction of cognitive decline and dementia: WHO guidelines*. https://iris.who.int/handle/10665/312180

World Health Organization. (2020). *Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019*. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death

World Health Organization. (2021). *Global status report on the public health response to dementia*. https://www.who.int/publications/i/item/9789240033245

World Health Organization. (2022a). *A blueprint for dementia research*. https://www.who.int/publications/i/item/9789240058248

World Health Organization. (2022b). *International Classification of Diseases, Eleventh Revision (ICD-11)*.

Wu, Q., Tchetgen Tchetgen, E. J., Osypuk, T. L., White, K., Mujahid, M., & Maria Glymour, M. (2013). Combining direct and proxy assessments to reduce attrition bias in a longitudinal study. *Alzheimer Disease and Associated Disorders*, *27*(3), 207–212. https://doi.org/10.1097/WAD.0b013e31826cfe90

Wu, Y.-T., Fratiglioni, L., Matthews, F. E., Lobo, A., Breteler, M. M. B., Skoog, I., & Brayne, C. (2016). Dementia in western Europe: Epidemiological evidence and implications for policy making. *The Lancet. Neurology*, *15*(1), 116–124. https://doi.org/10.1016/S1474-4422(15)00092-7

Yaffe, K., Vittinghoff, E., Dublin, S., Peltz, C. B., Fleckenstein, L. E., Rosenberg, D. E., Barnes, D. E., Balderson, B. H., & Larson, E. B. (2024). Effect of Personalized Risk-Reduction Strategies on Cognition and Dementia Risk Profile Among Older Adults: The SMARRT Randomized Clinical Trial. *JAMA Internal Medicine*, *184*(1), 54–62. https://doi.org/10.1001/jamainternmed.2023.6279

Yeo, B. S. Y., Song, H. J. J. M. D., Toh, E. M. S., Ng, L. S., Ho, C. S. H., Ho, R., Merchant, R. A., Tan, B. K. J., & Loh, W. S. (2023). Association of Hearing Aids and Cochlear Implants With Cognitive Decline and Dementia: A Systematic Review and Meta-analysis. *JAMA Neurology*, *80*(2), 134–141. https://doi.org/10.1001/jamaneurol.2022.4427

Yu, W., Esposito, M., Li, M., Clarke, P., Judd, S., & Finlay, J. (2023). Neighborhood 'Disamenities': Local barriers and cognitive function among Black and white aging adults. *BMC Public Health*, *23*(1), 197. https://doi.org/10.1186/s12889-023-15026-x

Yue, Y., & Hu, Y.-J. (2022a). A new approach to testing mediation of the microbiome at both the community and individual taxon levels. *Bioinformatics (Oxford, England)*, *38*(12), 3173–3180. https://doi.org/10.1093/bioinformatics/btac310

Yue, Y., & Hu, Y.-J. (2022b). Extension of PERMANOVA to Testing the Mediation Effect of the Microbiome. *Genes*, *13*(6), 940. https://doi.org/10.3390/genes13060940

Zekry, D., Duyckaerts, C., Moulias, R., Belmin, J., Geoffre, C., Herrmann, F., & Hauw, J.-J. (2002). Degenerative and vascular lesions of the brain have synergistic effects in dementia of the elderly. *Acta Neuropathologica*, *103*(5), 481–487. https://doi.org/10.1007/s00401-001-0493-5

Zetterberg, H., & Bendlin, B. B. (2021). Biomarkers for Alzheimer's disease—Preparing for a new era of disease-modifying therapies. *Molecular Psychiatry*, *26*(1), 296–308. https://doi.org/10.1038/s41380-020-0721-9

Zhang, N., Ranson, J. M., Zheng, Z.-J., Hannon, E., Zhou, Z., Kong, X., Llewellyn, D. J., King, D. A., & Huang, J. (2021). Interaction between genetic predisposition, smoking, and dementia risk: A population-based cohort study. *Scientific Reports*, *11*(1), 12953. https://doi.org/10.1038/s41598-021-92304-x

Zhang, Y., Jin, X., Lutz, M. W., Ju, S.-Y., Liu, K., Guo, G., Zeng, Y., & Yao, Y. (2021). Interaction between APOE ε4 and dietary protein intake on cognitive decline: A longitudinal cohort study. *Clinical Nutrition*, *40*(5), 2716–2725. https://doi.org/10.1016/j.clnu.2021.03.004

Zhang, Y., Qiu, C., Lindberg, O., Bronge, L., Aspelin, P., Bäckman, L., Fratiglioni, L., & Wahlund, L.-O. (2010). Acceleration of hippocampal atrophy in a non-demented elderly population: The SNAC-K study. *International Psychogeriatrics*, *22*(1), 14–25. https://doi.org/10.1017/S1041610209991396

Zhu, Z., Satten, G. A., & Hu, Y.-J. (2022). Integrative analysis of relative abundance data and presence–absence data of the microbiome using the LDM. *Bioinformatics*, *38*(10), 2915–2917. https://doi.org/10.1093/bioinformatics/btac181

Zuber, S., Bechtiger, L., Bodelet, J. S., Golin, M., Heumann, J., Kim, J. H., Klee, M., Mur, J., Noll, J., Voll, S., O'Keefe, P., Steinhoff, A., Zölitz, U., Muniz-Terrera, G., Shanahan, L., Shanahan, M. J., & Hofer, S. M. (2023). An integrative approach for the analysis of risk and health across the life course: Challenges, innovations, and opportunities for life course research. *Discover Social Science and Health*, *3*(1), 14. https://doi.org/10.1007/s44155-023-00044-2

Appendices

Appendix I Contributions

This section lists references to papers constituting the present thesis, further related studies in which I was first or contributing author, a selection of presentations at scientific conferences and research institutions as well as media coverage.

In addition to conducting the three studies presented in Chapters II to IV, I led the formal statistical analysis and visual data presentation for a rapid report on socioeconomic factors associated with vaccination willingness in Luxembourg (Leist et al., 2021). I further engaged in the Deep Dementia Phenotyping (DEMON) Network and drafted a section for a consensus paper in preparation on causal research questions regarding dementia risk factors (Foote et al., Manuscript in preparation, "Can longitudinal studies effectively ascertain temporality?"). Additionally, I participated in drafting and revising a publication discussing the application of ML approaches for research questions in the domain of social and health sciences, also involving the creation of figures (Leist et al., 2022). Moreover, I partook in a workshop of an international, interdisciplinary research group with expertise in aging, developmental psychology, youth research and statistics, and contributed to drafting and revising a publication supporting and guiding analysis of health outcomes with a life course model of risk (Zuber et al., 2023). Participation in this workshop further led to collaboration and my lead on a manuscript in preparation, investigating the accumulation of depressive symptoms in later life regarding dementia risk (Klee, Bodelet, et al., Manuscript in preparation). Lastly, I was invited by Prof. Heather Allore, to visit the Yale School of Medicine. During a four-month long research stay I collaborated with Prof. Allore and further visited Oregon Health & Science University to lead on a manuscript in preparation, applying a longitudinal extension of the PAF to examine the population-level absolute contribution of eight somatic conditions to dementia incidence in the US (Klee, Markwardt, et al., Manuscript in preparation).

Publications

*References constituting chapters II to IV are printed in bold font

Foote, I. F., Anderson, E. L., Armstrong, J. J., Bagshaw, P., Bocobo, G. A., Booi, L., Bothongo, P. L. K., Byford, M., Chandra, A., Das, S., Farina, F., Ferro, O., Geraets, A., Guo, H., Harshfield, E., Hassan, E., Jensen, D. E. A., Klee, M., Lophatananon, A., … Marshall, C. R. (Manuscript in preparation). *Assessing evidence for the causality of dementia risk factors: A DEMON Network consensus paper.*

Klee, M., Aho, V. T. E., May, P., Heintz-Buschart, A., Landoulsi, Z., Jónsdóttir, S. R., Pauly, C., Pavelka, L., Delacour, L., Kaysen, A., Krüger, R., Wilmes, P., Leist, A. K., NCER-PD Consortium. (2024). Education as Risk Factor of Mild Cognitive Impairment: The Link to the Gut Microbiome. *The Journal of Prevention of Alzheimer's Disease, 11, 759-768*. https://doi.org/10.14283/jpad.2024.19

Klee, M., Bodelet, J. S., Leist, A. K., & Muniz-Terrera, G. (Manuscript in preparation). *Investigating the Association of the Accumulation of Depressive Symptoms with Dementia Risk*.

Klee, M., Langa, K. M., & Leist, A. K. (2024). Performance of probable dementia classification in a European multi-country survey. *Scientific Reports*, *14*(1), 6657. https://doi.org/10.1038/s41598-024-56734-7

Klee, M., Leist, A. K., Veldsman, M., Ranson, J. M., & Llewellyn, D. J. (2023). Socioeconomic Deprivation, Genetic Risk, and Incident Dementia. *American Journal of Preventive Medicine*, *64*(5), 621–630. https://doi.org/10.1016/j.amepre.2023.01.012

Klee, M., Markwardt, S., Elman, M., Han, L., Leist, A. K., Allore, H. G., & Quiñones, A. R. (Manuscript in preparation). *Examining dementia fraction attributable to cardiometabolic multimorbidity over time*.

Leist, A. K., Klee, M., Kim, J. H., Rehkopf, D. H., Bordas, S. P. A., Muniz-Terrera, G., & Wade, S. (2022). Mapping of machine learning approaches for description, prediction, and causal inference in the social and health sciences. *Science Advances*, *8*(42), eabk1942. https://doi.org/10.1126/sciadv.abk1942

Leist, A. K., Klee, M., Paccoud, I., Pauly, L., Ghosh, S., Fritz, J., O Sullivan, M. P., Rommes, B., Wilmes, P., Krüger, R., & Consortium, C.-V. (2021). *Which demographic and socio-economic factors are associated with vaccination willingness and beliefs towards vaccination? Rapid report with first results*. https://orbilu.uni.lu/handle/10993/48567

Zuber, S., Bechtiger, L., Bodelet, J. S., Golin, M., Heumann, J., Kim, J. H., Klee, M., Mur, J., Noll, J., Voll, S., O'Keefe, P., Steinhoff, A., Zölitz, U., Muniz-Terrera, G., Shanahan, L., Shanahan, M. J., & Hofer, S. M. (2023). An integrative approach for the analysis of risk and health across the life course: Challenges, innovations, and opportunities for life course research. *Discover Social Science and Health*, *3*(1), 14. https://doi.org/10.1007/s44155-023-00044-2

Presentations at Scientific Conferences

Klee, M., Aho, V. T. E., May, P., Krüger, R., Wilmes, P., & Leist, A. K. (2023, November 8). *Education as a Risk Factor of Mild Cognitive Impairment – The Role of the Gut Microbiome. In A. K. Leist (Chair), The Social Determinants of Brain Health: From the Search for Mechanisms to Recommendations to Increase Equity* [Symposium presentation]. Annual Meeting of the Gerontological Society of America, Tampa, USA. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10737051/

Klee, M., Bodelet, J., & Muniz-Terrera, G. (2023, July 19). *Investigating the Role of the Accumulation of Depressive Symptoms for the Risk of Dementia* [Poster presentation].

Alzheimer's Association International Conference, Amsterdam, Netherlands. https://alz-journals.onlinelibrary.wiley.com/doi/full/10.1002/alz.082430

Klee, M., Langa, K. M., & Leist, A. K. (2022, June 8). *Country-level Variation in Dementia Prevalence in Europe: A Comparison of World Alzheimer Report and SHARE data. In A. K. Leist (Chair), Contextual and social determinants of cognitive ageing and dementia* [Symposium presentation]. 26th Nordic Congress of Gerontology, Odense, Denmark. https://www.26nkg.dk/files/26NKG%20abstract%20book_final_online_june6.pdf

Klee, M., & Leist, A. K. (2021, November 10). *Investigating Sequential and Simultaneous Changes in Trajectories of Cognitive Decline and Depressive Symptoms* [Oral presentation]. Annual Meeting of the Gerontological Society of America, online. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8680963/

Klee, M., Leist, A. K., Veldsman, M., Ranson, J. M., & Llewellyn, D. J. (2022, August 2). *Socioeconomic Deprivation, Genetics and Risk of Dementia* [Oral presentation]. Alzheimer's Association International Conference, San Diego, USA. https://alz-journals.onlinelibrary.wiley.com/doi/full/10.1002/alz.066803

Media Coverage

Hill, P. (2023). *The 'D' Word* (24) [Audio podcast episode]. Retrieved 9 January 2024, from https://audioboom.com/posts/8312538-pete-hill-the-d-word-episode-24

Klee, M. (2023, March). Meet the researcher. *Forum*, *430*, 55.

University of Luxembourg: New study finds link between socioeconomic deprivation and dementia risk. (2022, August). *RTL Today*. https://today.rtl.lu/news/luxembourg/a/1954764.html

Appendix II – Supplementary Material Chapter II

Appendix II Table S1 *Coefficients Used for Weighting of Individual-Level Socioeconomic Deprivation Scores*

Characteristic	Including not Disclosed[a]			Excluding not Disclosed[b]		
	Total No.[d]	Coefficient (95% CI)	*p*	Total No.[c]	Coefficient (95% CI)	*p*
Income[d]						
Greater 31 000	53 346	0 [Reference]		53 146	0 [Reference]	
From 18 000 to 31 000	52 834	0.34 (0.18, 0.50)	<.001	52 572	0.38 (0.21, 0.54)	<.001
Smaller 18 000	54 866	0.48 (0.31, 0.65)	<.001	54 339	0.54 (0.36, 0.71)	<.001
Not disclosed	35 322	0.67 (0.50, 0.85)	<.001			
Housing Type						
House or Flat	194 542	0 [Reference]		159 059	0 [Reference]	
Other	1,826	0.45 (0.11, 0.78)	.01	998	0.74 (0.33, 1.14)	<.001
Home Ownership						
Own Outright	151 971	0 [Reference]		123 790	0 [Reference]	
Other	44 397	0.35 (0.24, 0.46)	<.001	36 267	0.39 (0.26, 0.52)	<.001
Car Ownership						
One or more	179 091	0 [Reference]		146 778	0 [Reference]	
Other	17 277	0.32 (0.18, 0.46)	<.001	13 279	0.29 (0.12, 0.45)	<.001

Note. All Cox proportional-hazards regressions were adjusted for the 20 first principal components, third degree relatedness, age, sex, education, retirement status and number of people in the household. [a]Coefficients used to compute individual-level socioeconomic deprivation with not disclosed information included in other category for housing type, home ownership and car ownership. [b]Coefficients used to compute individual-level socioeconomic deprivation when excluding not disclosed information relating to income, housing type, home ownership and car ownership. [c]Reported results are based on the first imputed data set. [d]Income assessed in Pound sterling (£) based on average total household income before tax.

Appendix II Table S2 *Incident Dementia Cases in Area-Level Socioeconomic Deprivation Groups*

Area-Level Socioeconomic Deprivation	Low-to- Moderate	High
No. of Dementia Cases[a]	1,266	503
Absolute Risk, % (95% CI)[a]	0.81 (0.76, 0.85)	1.28 (1.17, 1.40)
Incidence Rates per 1,000 Person-Years (95% CI)[a]	1.02 (0.97, 1.08)	1.65 (1.51, 1.80)
Total No.[a]	157 095	39 273

[a]Reported results are based on the first imputed data set.

Appendix II Table S3 *Incident Dementia Cases in Individual-Level Socioeconomic Deprivation Groups*

Area-Level Socioeconomic Deprivation	Low	Intermediate	High
No. of Dementia Cases[a]	174	1,042	553
Absolute Risk, % (95% CI)[a]	0.44 (0.38, 0.51)	0.88 (0.83, 0.94)	1.41 (1.29, 1.53)
Incidence Rates per 1,000 Person-Years (95% CI)[a]	0.56 (0.48, 0.65)	1.12 (1.05, 1.19)	1.80 (1.66, 1.96)
Total No.[a]	39 274	117 821	39 273

[a]Reported results are based on the first imputed data set.

Appendix II Table S4 *Total Participants and Incident Dementia Cases According to Area-Level Socioeconomic Deprivation Within Each Genetic Risk Category*

Genetic risk	Low		Intermediate		High	
Area-Level Socioeconomic Deprivation	Low-to- Moderate	High	Low-to- Moderate	High	Low-to- Moderate	High
No. of Dementia Cases[a]	177	70	744	294	345	139
Absolute Risk, % (95% CI)[a]	0.56 (0.48, 0.65)	0.92 (0.72, 1.16)	0.79 (0.73, 0.85)	1.25 (1.11, 1.40)	1.11 (0.99, 1.23)	1.71 (1.44, 2.01)
Incidence Rates per 1,000 Person-Years (95% CI)[a]	0.71 (0.61, 0.82)	1.18 (0.92, 1.50)	1.00 (0.93, 1.07)	1.61 (1.43, 1.81)	1.40 (1.26, 1.56)	2.20 (1.85, 2.59)
Total No.[a]	31 648	7,626	94 316	23 505	31 131	8,142

[a]Reported results are based on the first imputed data set.

Appendix II Table S5 *Total Participants and Incident Dementia Cases According to Individual-Level Socioeconomic Deprivation Within Each Genetic Risk Category*

Genetic risk	Low			Intermediate			High		
Individual-Level Socioeconomic Deprivation	Low	Intermediate	High	Low	Intermediate	High	Low	Intermediate	High
No. of Dementia Cases[a]	25	134	88	103	614	321	46	294	144
Absolute Risk, % (95% CI)[a]	0.31 (0.20, 0.45)	0.57 (0.48, 0.67)	1.17 (0.94, 1.44)	0.44 (0.36, 0.53)	0.87 (0.80, 0.94)	1.36 (1.21, 1.51)	0.59 (0.44, 0.79)	1.26 (1.12, 1.41)	1.78 (1.50, 2.09)
Incidence Rates per 1,000 Person-Years (95% CI)[a]	0.39 (0.25, 0.58)	0.72 (0.60, 0.85)	1.49 (1.20, 1.84)	0.56 (0.46, 0.68)	1.10 (1.01, 1.19)	1.74 (1.56, 1.94)	0.75 (0.55, 1.00)	1.59 (1.41, 1.78)	2.27 (1.92, 2.67)
Total No.[a]	8,110	23 624	7,540	23 417	70 774	23 630	7,747	23 423	8,103

[a]Reported results are based on the first imputed data set.

Appendix II Table S6 Coefficients for Multivariable Linear Regressions of White Matter Hyperintensities With Full and Reduced Deconfounding Set

Characteristic	Imputed Data (*n*=11 035)				Complete-Case Data (*n*=8,131)			
	Full Set		Reduced Set		Full Set		Reduced Set	
	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*
Individual-Level Socioeconomic Deprivation								
Low	0 [Reference]		0 [Reference]		0 [Reference]		0 [Reference]	
Intermediate	0.05 (0.00, 0.10)	.04	0.04 (-0.01, 0.09)	.09	0.06 (0.00, 0.12)	.04	0.05 (-0.01, 0.10)	.08
High	0.10 (0.01, 0.19)	.03	0.10 (0.01, 0.19)	.03	0.06 (-0.06, 0.17)	.33	0.06 (-0.06, 0.17)	.34
Area-Level Socioeconomic Deprivation								
Low-to-Moderate	0 [Reference]		0 [Reference]		0 [Reference]		0 [Reference]	
High	0.08 (0.01, 0.15)	.03	0.07 (0.01, 0.14)	.03	0.07 (-0.03, 0.17)	.16	0.07 (-0.03, 0.16)	.17

Note. All imaging derived phenotypes were deconfounded in multivariable linear regressions, either adjusting for the full set including site-specific derivatives capturing indicators of age, age squared, sex, age-sex interactions, head size, days since the scanner start-up, days since the scanner start-up squared and two dummy variables coding site or the reduced set including age, sex, age-sex interactions, head size and two dummy variables coding site. Residuals were then entered in secondary multivariable linear regressions including 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, education, marital status, healthy lifestyle, depressive symptoms in last two weeks, individual-level and area-level socioeconomic deprivation as well as genetic risk.

Appendix II Table S7 *Coefficients for Multivariable Linear Regressions of Hippocampal Volume (right) With Full and Reduced Deconfounding Set*

Characteristic	Imputed Data (*n*=10 838)				Complete-Case Data (*n*=7,999)			
	Full Set		Reduced Set		Full Set		Reduced Set	
	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*
Individual-Level Socioeconomic Deprivation								
Low	0 [Reference]		0 [Reference]		0 [Reference]		0 [Reference]	
Intermediate	-0.03 (-0.09, 0.02)	.18	-0.04 (-0.09, 0.01)	.15	-0.04 (-0.09, 0.02)	.20	-0.04 (-0.10, 0.02)	.16
High	-0.00 (-0.09, 0.09)	.94	-0.01 (-0.10, 0.08)	.88	0.04 (-0.07, 0.15)	.50	0.03 (-0.08, 0.15)	.55
Area-Level Socioeconomic Deprivation								
Low-to-Moderate	0 [Reference]		0 [Reference]		0 [Reference]		0 [Reference]	
High	-0.04 (-0.11, 0.03)	.26	-0.04 (-0.11, 0.03)	.27	0.00 (-0.09, 0.09)	.98	-0.00 (-0.09, 0.09)	1.00

Note. All imaging derived phenotypes were deconfounded in multivariable linear regressions, either adjusting for the full set including site-specific derivatives capturing indicators of age, age squared, sex, age-sex interactions, head size, days since the scanner start-up, days since the scanner start-up squared and two dummy variables coding site or the reduced set including age, sex, age-sex interactions, head size and two dummy variables coding site. Residuals were then entered in secondary multivariable linear regressions including 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, education, marital status, healthy lifestyle, depressive symptoms in last two weeks, individual-level and area-level socioeconomic deprivation as well as genetic risk.

Appendix II Table S8 *Coefficients for Multivariable Linear Regressions of Hippocampal Volume (left) With Full and Reduced Deconfounding Set*

Characteristic	Imputed Data (*n*=10 920)				Complete-Case Data (*n*=8,056)			
	Full Set		Reduced Set		Full Set		Reduced Set	
	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*
Individual-Level Socioeconomic Deprivation								
Low	0 [Reference]		0 [Reference]		0 [Reference]		0 [Reference]	
Intermediate	-0.05 (-0.10, 0.00)	.06	-0.05 (-0.10, 0.00)	.06	-0.03 (-0.08, 0.03)	.37	-0.03 (-0.08, 0.03)	.32
High	-0.01 (-0.09, 0.08)	.91	-0.01 (-0.10, 0.09)	.91	0.07 (-0.04, 0.18)	.22	0.07 (-0.04, 0.18)	.22
Area-Level Socioeconomic Deprivation								
Low-to-Moderate	0 [Reference]		0 [Reference]		0 [Reference]		0 [Reference]	
High	-0.06 (-0.13, 0.01)	.07	-0.06 (-0.13, 0.01)	.08	-0.05 (-0.14, 0.04)	.26	-0.05 (-0.14, 0.04)	.29

Note. All imaging derived phenotypes were deconfounded in multivariable linear regressions, either adjusting for the full set including site-specific derivatives capturing indicators of age, age squared, sex, age-sex interactions, head size, days since the scanner start-up, days since the scanner start-up squared and two dummy variables coding site or the reduced set including age, sex, age-sex interactions, head size and two dummy variables coding site. Residuals were then entered in secondary multivariable linear regressions including 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, education, marital status, healthy lifestyle, depressive symptoms in last two weeks, individual-level and area-level socioeconomic deprivation as well as genetic risk.

Appendix II Table S9 *Coefficients for Multivariable Linear Regressions of Whole Brain Volume With Full and Reduced Deconfounding Set*

Characteristic	Imputed Data (*n*=11 035)				Complete-Case Data (*n*=8,139)			
	Full Set		Reduced Set		Full Set		Reduced Set	
	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*
Individual-Level Socioeconomic Deprivation								
Low	0 [Reference]		0 [Reference]		0 [Reference]		0 [Reference]	
Intermediate	-0.03 (-0.08, 0.02)	.25	-0.03 (-0.08, 0.02)	.30	-0.03 (-0.08, 0.03)	.31	-0.02 (-0.08, 0.03)	.39
High	-0.03 (-0.12, 0.06)	.46	-0.03 (-0.12, 0.06)	.49	0.07 (-0.03, 0.18)	.17	0.08 (-0.03, 0.18)	.15
Area-Level Socioeconomic Deprivation								
Low-to-Moderate	0 [Reference]		0 [Reference]		0 [Reference]		0 [Reference]	
High	-0.05 (-0.12, 0.02)	.17	-0.05 (-0.12, 0.02)	.19	0.00 (-0.09, 0.09)	.99	-0.00 (-0.09, 0.09)	.94

Note. All imaging derived phenotypes were deconfounded in multivariable linear regressions, either adjusting for the full set including site-specific derivatives capturing indicators of age, age squared, sex, age-sex interactions, head size, days since the scanner start-up, days since the scanner start-up squared and two dummy variables coding site or the reduced set including age, sex, age-sex interactions, head size and two dummy variables coding site. Residuals were then entered in secondary multivariable linear regressions including 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, education, marital status, healthy lifestyle, depressive symptoms in last two weeks, individual-level and area-level socioeconomic deprivation as well as genetic risk.

Appendix II Table S10 *Coefficients for Multivariable Linear Regressions of White Matter Volume With Full and Reduced Deconfounding Set*

Characteristic	Imputed Data (*n*=11 039)				Complete-Case Data (*n*=8,140)			
	Full Set		Reduced Set		Full Set		Reduced Set	
	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*
Individual-Level Socioeconomic Deprivation								
Low	0 [Reference]		0 [Reference]		0 [Reference]		0 [Reference]	
Intermediate	-0.00 (-0.05, 0.05)	.91	0.01 (-0.04, 0.06)	.82	0.01 (-0.04, 0.06)	.75	0.02 (-0.04, 0.07)	.50
High	-0.02 (-0.10, 0.06)	.60	-0.02 (-0.10, 0.07)	.67	0.06 (-0.05, 0.16)	.30	0.06 (-0.04, 0.17)	.26
Area-Level Socioeconomic Deprivation								
Low-to-Moderate	0 [Reference]		0 [Reference]		0 [Reference]		0 [Reference]	
High	0.02 (-0.05, 0.09)	.63	0.02 (-0.05, 0.09)	.59	0.02 (-0.07, 0.12)	.65	0.02 (-0.08, 0.11)	.74

Note. All imaging derived phenotypes were deconfounded in multivariable linear regressions, either adjusting for the full set including site-specific derivatives capturing indicators of age, age squared, sex, age-sex interactions, head size, days since the scanner start-up, days since the scanner start-up squared and two dummy variables coding site or the reduced set including age, sex, age-sex interactions, head size and two dummy variables coding site. Residuals were then entered in secondary multivariable linear regressions including 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, education, marital status, healthy lifestyle, depressive symptoms in last two weeks, individual-level and area-level socioeconomic deprivation as well as genetic risk.

Appendix II Table S11 *Coefficients for Multivariable Linear Regressions of Grey Matter Volume with Full and Reduced Deconfounding Set*

Characteristic	Imputed Data (*n*=11 018)				Complete-Case Data (*n*=8,128)			
	Full Set		Reduced Set		Full Set		Reduced Set	
	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*	Coefficient (95% CI)	*p*
Individual-Level Socioeconomic Deprivation								
Low	0 [Reference]		0 [Reference]		0 [Reference]		0 [Reference]	
Intermediate	-0.05 (-0.10, 0.00)	.07	-0.05 (-0.10, 0.00)	.04	-0.06 (-0.12, -0.00)	.046	-0.06 (-0.12, -0.01)	.03
High	-0.04 (-0.14, 0.05)	.39	-0.04 (-0.14, 0.05)	.35	0.04 (-0.07, 0.14)	.50	0.03 (-0.07, 0.14)	.55
Area-Level Socioeconomic Deprivation								
Low-to-Moderate	0 [Reference]		0 [Reference]		0 [Reference]		0 [Reference]	
High	-0.11 (-0.18, -0.04)	.004	-0.11 (-0.18, -0.03)	.004	-0.04 (-0.13, 0.06)	.43	-0.04 (-0.13, 0.06)	.43

Note. All imaging derived phenotypes were deconfounded in multivariable linear regressions, either adjusting for the full set including site-specific derivatives capturing indicators of age, age squared, sex, age-sex interactions, head size, days since the scanner start-up, days since the scanner start-up squared and two dummy variables coding site or the reduced set including age, sex, age-sex interactions, head size and two dummy variables coding site. Residuals were then entered in secondary multivariable linear regressions including 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, education, marital status, healthy lifestyle, depressive symptoms in last two weeks, individual-level and area-level socioeconomic deprivation as well as genetic risk.

Appendix II Table S12 *Risk of Dementia with Area-Level Socioeconomic Deprivation and Genetic Risk in Complete-Case Data*

Genetic risk	Low		Intermediate		High	
Area-Level Socioeconomic Deprivation	Low-to- Moderate	High	Low-to- Moderate	High	Low-to- Moderate	High
Total No.	22 154	3,875	66 516	11 927	21 778	4,055
No. of Dementia Cases / Person-Years	120 / 173 636	31 / 29 809	488 / 521 545	138 / 92 019	215 / 171 055	50 / 31 449
HR (95% CI)	1 [Reference]	1.26 (0.85, 1.88)	1.34 (1.10, 1.64)	1.76 (1.37, 2.25)	1.81 (1.45, 2.27)	1.80 (1.29, 2.52)
p		.25	.004	<.001	<.001	<.001

Note. The cox proportional-hazards regression model was adjusted for the 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, age, sex, education, marital status, healthy lifestyle, depressive symptoms in last two weeks and individual-level socioeconomic deprivation.

HR=Hazard ratio.

Appendix II Table S13 *Risk of Dementia with Individual-Level Socioeconomic Deprivation and Genetic Risk in Complete-Case Data*

Genetic risk	Low			Intermediate			High		
Individual-Level Socioeconomic Deprivation	Low	Intermediate	High	Low	Intermediate	High	Low	Intermediate	High
Total No.	6,849	15 696	3,484	19 790	47 819	10 834	6,554	15 593	3,686
No. of Dementia Cases / Person-Years	23 / 53 656	89 / 122 655	39 / 27 133	84 / 155 026	399 / 374 404	143 / 84 132	39 / 51 460	174 / 122 193	52 / 28 852
HR (95% CI)	1 [Reference]	1.37 (0.86, 2.17)	2.29 (1.36, 3.86)	1.26 (0.79, 2.00)	1.97 (1.29, 3.01)	2.75 (1.75, 4.31)	1.80 (1.07, 3.01)	2.63 (1.69, 4.08)	2.89 (1.75, 4.76)
p		.18	.002	.33	.002	<.001	.03	<.001	<.001

Note. The cox proportional-hazards regression model was adjusted for the 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, age, sex, education, marital status, healthy lifestyle, depressive symptoms in last two weeks and area-level socioeconomic deprivation. HR=Hazard ratio.

Appendix II Table S14 *Risk of Dementia According to Area-Level Socioeconomic Deprivation in Subgroups Stratified by Genetic Risk*

Genetic risk	Low		Intermediate		High	
Area-Level Socioeconomic Deprivation[a]	Low-to-Moderate (*n*=31 648)	High (*n*=7,626)	Low-to-Moderate (*n*=94 316)	High (*n*=23 505)	Low-to-Moderate (*n*=31 131)	High (*n*=8,142)
No. of Dementia Cases / Person-Years[a]	177 / 249 647	70 / 59 124	744 / 744 724	294 / 182 389	345 / 246 144	139 / 63 285
HR (95% CI)	1 [Reference]	1.18 (0.87, 1.61)	1 [Reference]	1.29 (1.11, 1.50)	1 [Reference]	1.32 (1.06, 1.64)
p		.29		<.001		.01

Note. All Cox proportional-hazards regressions were adjusted for the 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, age, sex, education, marital status, and individual-level socioeconomic deprivation. [a]Reported results are based on the first imputed data set. HR=Hazard ratio.

Appendix II Table S15 *Risk of Dementia According to Individual-Level Socioeconomic Deprivation in Subgroups Stratified by Genetic Risk*

Genetic risk	Low			Intermediate			High		
Individual-Level Socioeconomic Deprivation[a]	Low (*n*=8,110)	Intermediate (*n*=23 624)	High (*n*=7,540)	Low (*n*=23 417)	Intermediate (*n*=70 774)	High (*n*=23 630)	Low (*n*=7,747)	Intermediate (*n*=23 423)	High (*n*=8,103)
No. of Dementia Cases / Person-Years[a]	25 / 63 790	134 / 186 093	88 / 58 887	103 / 184 307	614 / 558 529	321 / 184 276	46 / 61 124	294 / 184 928	144 / 63 377
HR (95% CI)	1 [Reference]	1.50 (0.96, 2.35)	2.73 (1.66, 4.50)	1 [Reference]	1.61 (1.29, 2.00)	2.34 (1.83, 2.99)	1 [Reference]	1.72 (1.25, 2.38)	2.31 (1.61, 3.32)
p		.07	<.001		<.001	<.001		<.001	<.001
p for Trend		<.001			<.001			<.001	

Note. All Cox proportional-hazards regressions were adjusted for the 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, age, sex, education, marital status, and area-level socioeconomic deprivation. [a]Reported results are based on the first imputed data set. HR=Hazard ratio.

Appendix II Table S16 *Risk of Dementia According to Area-Level Socioeconomic Deprivation in Subgroups Stratified by Sex*

Sex	Female		Male	
Area-Level Socioeconomic Deprivation[a]	Low-to-Moderate (*n*=82 938)	High (*n*=20 496)	Low-to-Moderate (*n*=74 157)	High (*n*=18 777)
No. of Dementia Cases / Person-Years[a]	569 / 659 247	221 / 161 207	697 / 581 268	282 / 143 591
HR (95% CI)	1 [Reference]	1.25 (1.05, 1.48)	1 [Reference]	1.31 (1.12, 1.53)
p		.01		<.001

Note. All Cox proportional-hazards regressions were adjusted for the 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, genetic risk, age, sex, education, marital status, and individual-level socioeconomic deprivation. [a]Reported results are based on the first imputed data set.

HR=Hazard ratio.

Appendix II Table S17 *Risk of Dementia According to Individual-Level Socioeconomic Deprivation in Subgroups Stratified by Sex*

Sex	Female			Male		
Individual-Level Socioeconomic Deprivation[a]	Low (*n*=16 617)	Intermediate (*n*=65 137)	High (*n*=21 680)	Low (*n*=22 657)	Intermediate (*n*=52 684)	High (*n*=17 593)
No. of Dementia Cases / Person- Years[a]	57 / 131 217	467 / 517 550	266 / 171 687	117 / 178 004	575 / 412 001	287 / 134 854
HR (95% CI)	1 [Reference]	1.51 (1.14, 2.00)	2.29 (1.68, 3.13)	1 [Reference]	1.71 (1.39, 2.10)	2.44 (1.92, 3.11)
p		.004	<.001		<.001	<.001
p for Trend		<.001			<.001	

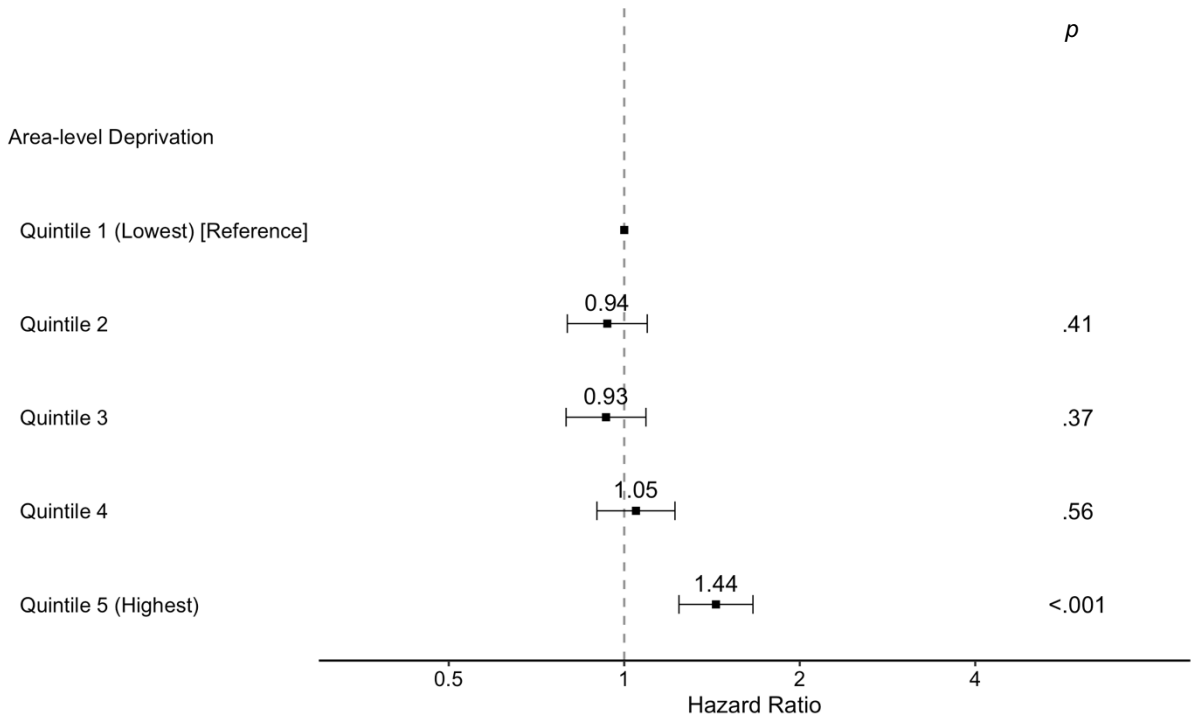
Note. All Cox proportional-hazards regressions were adjusted for the 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, genetic risk, age, sex, education, marital status, and area-level socioeconomic deprivation. [a]Reported results are based on the first imputed data set. HR=Hazard ratio.

Appendix II Table S18 *Proportion of Individual-Level Socioeconomic Deprivation Across Lifestyle in Complete-Case and Imputed Data for the Full and Imaging Subsample*

Lifestyle	Individual-Level Socioeconomic Deprivation – Full Sample					Individual-Level Socioeconomic Deprivation – Imaging Subsample				
	Low	Intermediate	High	Missing No.	Total No.		Intermediate	High	Missing No.	Total No.
					Complete-Case Data					
Favourable	19.79%	60.13%	20.08%	79	32 761	29.27%	58.53%	12.20%	8	2,033
Intermediate	22.94%	61.06%	15.99%	419	98 104	33.68%	58.40%	7.92%	29	6,123
Unfavourable	20.94%	56.87%	22.19%	282	32 559	37.70%	54.00%	8.30%	8	1,626
Missing	10.21%	59.79%	30.00%	250	31 914	19.39%	65.46%	15.14%	8	1,248
					Imputed Data[a]					
Favourable	17.44%	64.87%	17.69%		39 273	26.52%	61.94%	11.54%		2,244
Intermediate	20.90%	60.26%	18.84%		117 821	32.53%	57.05%	10.42%		6,898
Unfavourable	19.88%	54.35%	25.78%		39 274	36.32%	52.24%	11.44%		1,941

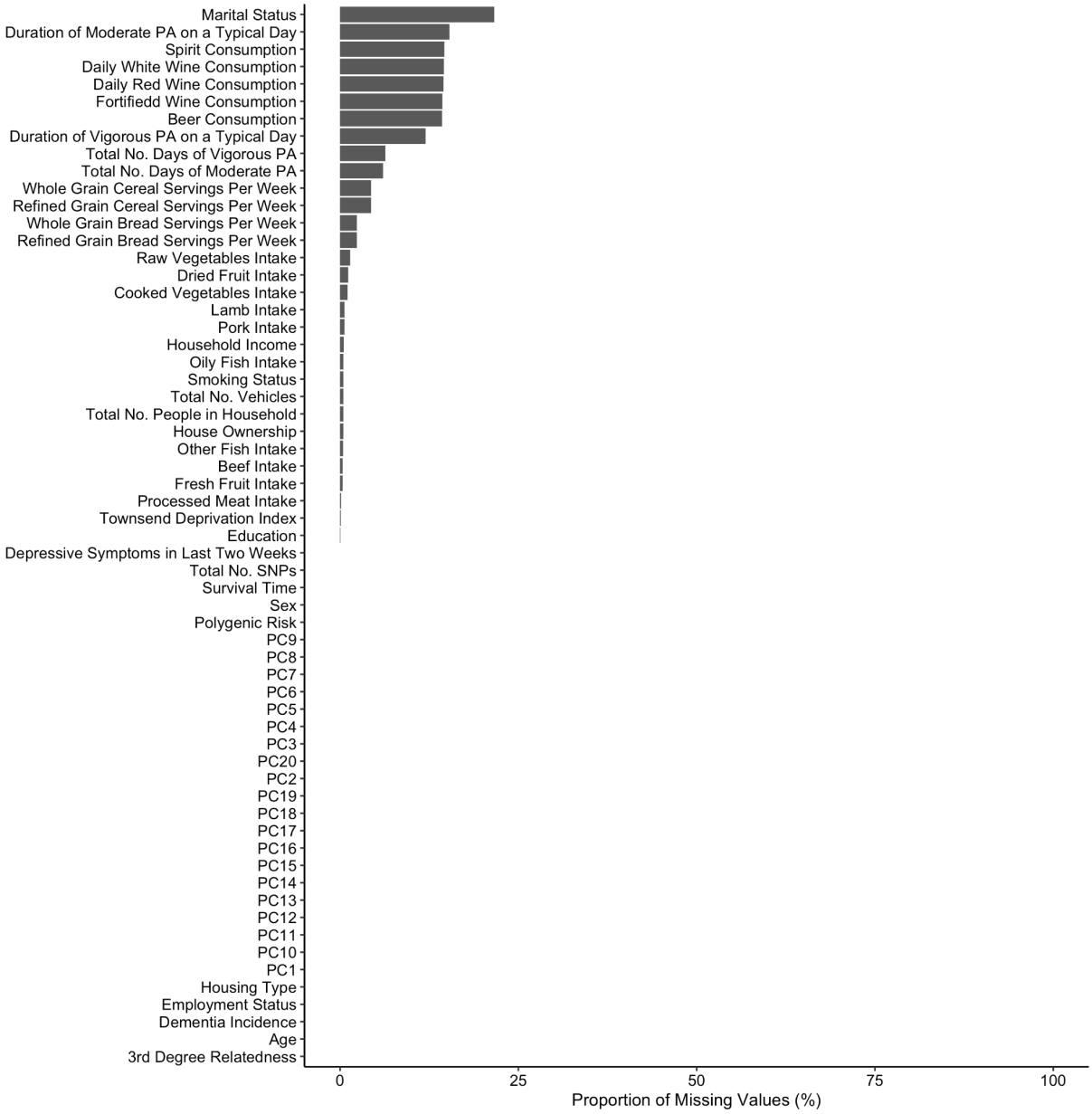
Note. Percentages are based on the total number of participants without missing data on individual-level deprivation and may not sum to 100 because of rounding. [a]Reported results are based on the first imputed data set.

Appendix II Figure S1 *Risk of Incident Dementia Across Area-Level Socioeconomic Deprivation Quintiles*



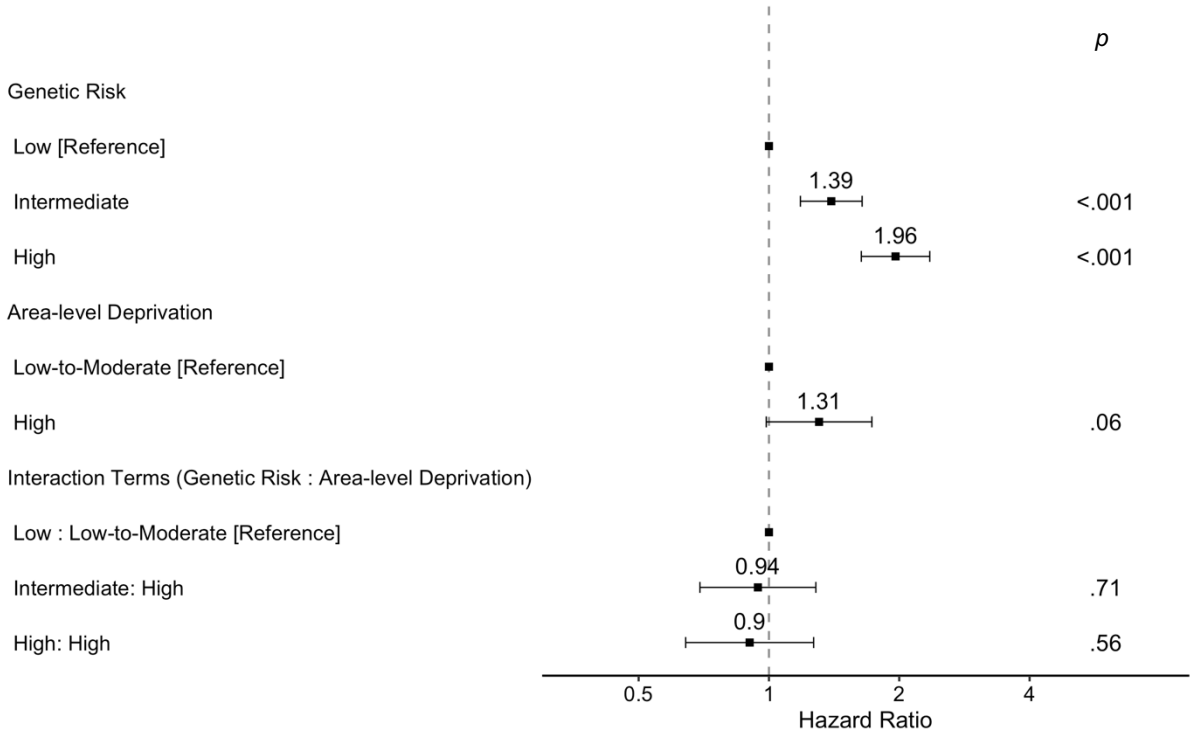
Note. Bars indicate 95% confidence intervals. Hazard ratios are depicted on a log-scale. Cox proportional-hazards regression was adjusted for the 20 first principal components, third degree relatedness, age, sex, education, and marital status.

Appendix II Figure S2 *Proportion of Missing Data Prior to Imputation*



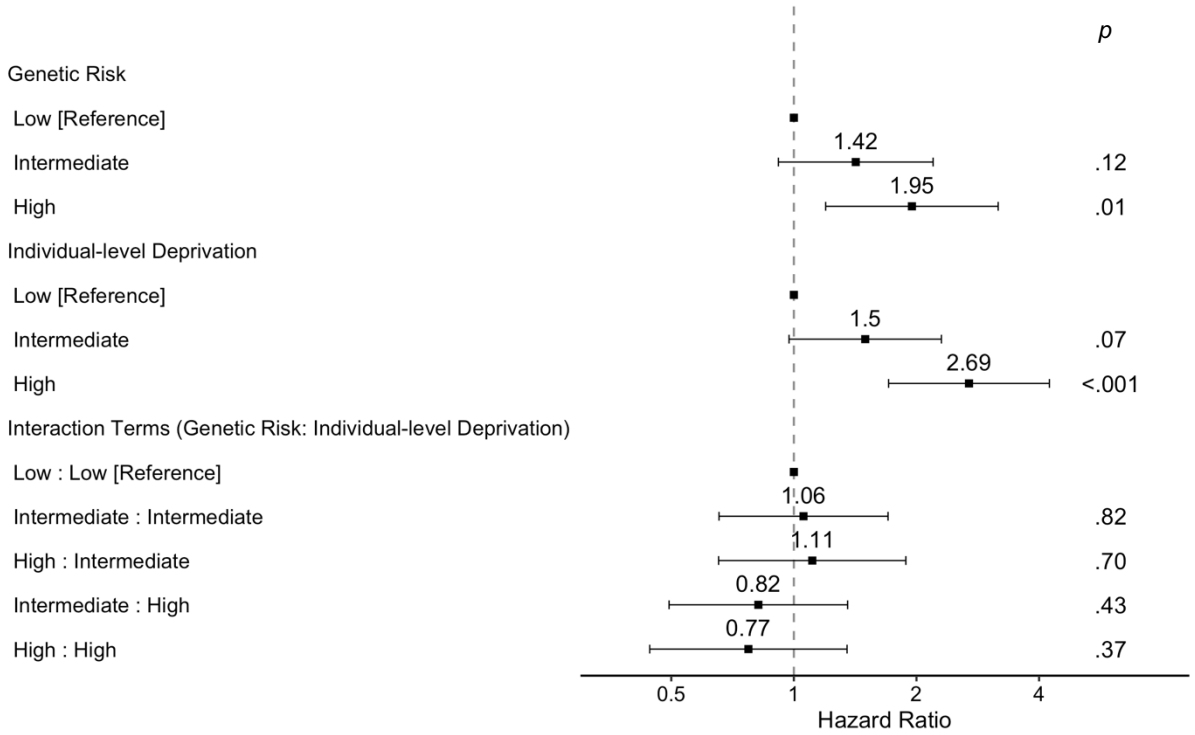
Note. All variable relevant to our analyses were used to impute missing values, including 29 variables were complete after application of eligibility criteria. Some variables were considered relevant to all analyses and thus used for imputation of all variables: 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, polygenic risk score, age, sex, education, dementia, follow-up time, household income, vehicle and home ownership, housing type, Townsend deprivation index, and number of people in the household. Further variables used for during imputation were retirement status, marital status, depressive symptoms in last two weeks and 24 variables indicating physical activity, diet, smoking behaviour, and alcohol intake, used to compute the healthy lifestyle index. PA=Physical Activity; SNP=Single Nucleotide Polymorphism; PC=principal component.

Appendix II Figure S3 *Dementia Risk With Interaction Terms for Area-Level Socioeconomic Deprivation*



Note. Bars indicate 95% confidence intervals. Hazard ratios are depicted on a log-scale. Colons indicate interaction terms. Cox proportional-hazards regression was adjusted for the 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, age, sex, education, marital status, healthy lifestyle, depressive symptoms in last two weeks and individual-level socioeconomic deprivation.

Appendix II Figure S4 *Dementia Risk With Interaction Terms for Individual-Level Socioeconomic Deprivation*



Note. Bars indicate 95% confidence intervals. Hazard ratios are depicted on a log-scale. Colons indicate interaction terms. Cox proportional-hazards regression was adjusted for the 20 first principal components, third degree relatedness, number of alleles used to compute the polygenic risk score, age, sex, education, marital status, healthy lifestyle, depressive symptoms in last two weeks and area-level socioeconomic deprivation.

Appendix III – Supplementary Material Chapter III

Appendix III *Diversity Measures*

Alpha Diversity:

Chao1 was computed as an indicator of richness, which ignores abundance and increases with the number of prevalent species (Chao, 1984). The Shannon and inverse Simpson indices were computed as indicators of both richness and evenness. They increase with the number of prevalent species and penalise presence of microbes dominating in prevalence (Hill et al., 2003). Measures were computed after rarefication (*estimate_richness* [*phyloseq*]; *rarefy_even_depth* [*phyloseq*]) (McMurdie & Holmes, 2013).

Beta Diversity:

Percentage difference, referred to sometimes as Bray-Curtis dissimilarity, and Jaccard distance were computed, which reflect the fraction of the number of unique species between individuals and the number of shared, and unique species (*vegdist* [*vegan*]) (Faith et al., 1987; Legendre & De Cáceres, 2013; Oksanen J. et al., 2022).

Appendix III *R Session Info Excerpt*

```
─ Session info ─────────────────────────────────────────────────────────────────

version  R version 4.3.1 (2023-06-16)
 os       macOS Ventura 13.5
 system   x86_64, darwin20
 ui       RStudio
 language (EN)
 rstudio  2023.06.2+561 Mountain Hydrangea (desktop)

─ Packages ─────────────────────────────────────────────────────────────────────

 ANCOMBC              * 2.2.1     2023-07-06 [1] Bioconductor
 CMAverse             * 0.1.0     2023-09-09 [1] Github (BS1125/CMAverse@fa8ccab)
 DESeq2               * 1.40.2    2023-06-25 [1] Bioconductor
 LDM                  * 6.0       2023-09-04 [1] CRAN (R 4.3.0)
 phyloseq             * 1.44.0    2023-05-11 [1] Bioconductor
 vegan                * 2.6-4     2022-10-11 [1] CRAN (R 4.3.0)
```

Appendix III *Description of Differential Abundance Analysis*

Since abundance reflects count data that is compositional (i.e., abundance of one species directly affects abundance of others), overdispersed and skewed (i.e., high variance and frequent zero counts for rare species), traditional approaches to statistical inference may lead to inflated false discovery rates and are thus not applicable (Aitchison, 1982; Mandal et al., 2015). Multiple methods have been developed to overcome these caveats, but they may provide discordant results in complex disease settings. Here applied approaches were the analysis of compositions of microbiomes with bias correction (*ancombc* [*ANCOMBC*]) and moderated estimation of fold change and dispersion (*DESeq* [*DESeq2*]). *DESeq* applies negative binomial generalised linear models and *ancombc* implements bias correction inherent to the underlying abundance quantification. For a more detailed description see function descriptions in packages *DESeq2* and *ANCOMBC* (Kaul et al., 2017; H. Lin & Peddada, 2020; Love et al., 2014; Mandal et al., 2015).

Appendix III *R Code Mediation Analysis With CMAverse*

```
do_cmest = function(data,
              mediator,
              a, astar = "0-10",
              mval,
              yreg = "logistic",
              int = FALSE,
              print = F){
  set.seed(123)
  med_analysis_1 <- cmest(data = data,
                          model = "rb",
                          outcome = "MCI_bin",
                          exposure = "Years_of_Education",
                          mediator = c(mediator),
                          basec = c("Age",
                                    "Gender",
                                    "ATB_in_last_6_months",
                                    #"BMI",
                                    "BDI_I_mild",
                                    "First_Language",
                                    "Living_With_Partner",
                                    "APOE4"),
                          mreg = list("linear"),
                          yreg = yreg,
                          EMint = int,
                          astar = astar,
                          a = a,
                          mval = mval,
                          yval = 1,
                          estimation = "imputation",
                          inference = "bootstrap",
                          nboot = 5000)
  if(print) {med_analysis_1 %>% summary() %>% print()}
  return(med_analysis_1)
}

# function call example
clin_df = ncer_adiv %>%
  mutate(Chao1 = as.vector(scale(Chao1)))

res16_Chao1_int = do_cmest(clin_df,
                           mediator = "Chao1",
                           a = "16+",
                           mval = list(0),
                           int = T)
```

Appendix III *R Code Mediation Analysis With Ldm and PermanovaFL*

```
# required formula
# otu.table | (set of confounders) ~ (set of exposures) + (set of outcomes)

sample_tab = ncer_phyloseqs$Genus %>% sample_data %>% as_tibble %>% as.data.frame
otu_tab = ncer_phyloseqs$Genus %>% otu_table() %>% t() %>% as.data.frame()

# ldm med
res.ldm.med <- ldm(formula = otu_tab |
                     (Age + Gender + ATB_in_last_6_months + BDI_I_mild +
                     First_Language + Living_With_Partner + APOE4) ~
                     (Years_of_Education) + # exposure
                     (MCI), # outcome
                   data=clin_df,
                   seed=67817,
                   n.cores=12,
                   test.mediation=TRUE,
                   test.omni3 = T)

# permanova med
res.perm.med = permanovaFL(otu_tab |
                                (Age + Gender + ATB_in_last_6_months + BDI_I_mild
                                + First_Language + Living_With_Partner + APOE4) ~
                                (Years_of_Education) + # exposure
                                (MCI), # outcome
                           data=clin_df, seed=82955,
                           test.mediation = T,
                           dist.method = c("jaccard", "bray"),
                           binary = c(TRUE, FALSE))
```

Note that *ldm* and *permanovaFL* preclude effect decomposition. While other approaches to mediation analysis with multiple mediators and effect decomposition exist, the assumption of multivariate normal distribution among mediators is not met for compositional abundance data (Aitchison, 1982). Further, given the large number of mediators, close to sample size, regularisation would be required (VanderWeele & Vansteelandt, 2014; Yue & Hu, 2022b).

Appendix III *Direct and Indirect Effects*

We acknowledge that temporality, necessary for causal interpretation, is not given with cross-sectional data. However, the cited literature refers to the term 'effects' hence this statistical terminology is used to avoid misinterpretation. Here, NDE refers to the difference in the counterfactual outcome, fixing the mediator to the level it would have taken with education 0-10 years, and intervening to change education from 0-10 years to >10 years, i.e., pure natural direct effect (VanderWeele, 2014). Hence, NDE describes the effect of education on MCI irrespective of alpha diversity (2). Note that CDE describes the effect of education on MCI for a specified level of alpha diversity, i.e., the sample mean. The NIE refers to the difference in the counterfactual outcome, fixing education to >10 years and intervening to change the mediator from the level it would have taken in 0-10 years to the level it would have taken in >10 years, i.e., total indirect effect (VanderWeele, 2014). Hence, NIE describes the effect of education on MCI only passing through alpha diversity (3) (Richiardi et al., 2013). Note that NDE and NIE reflect direct and indirect effects obtained using the approach of Baron and Kenny in absence of interaction between education and microbiome diversity (Baron & Kenny, 1986; Valeri & VanderWeele, 2013).

$$E\left[Y_{a,M(a^*)} - Y_{a^*,M(a^*)} | C = c\right] \quad (2)$$

$$E\left[Y_{a,M(a)} - Y_{a,M(a^*)} | C = c\right] \quad (3)$$

With

exposure a^* at reference level and exposure a at intervention level,

mediator M observed at a^* or a, conditional on covariates $C = c$

Appendix III Table S1 *Coefficients of Regression Models With Chao1*

Variable	Mediator Model			Outcome Model					
				With Interaction			Without Interaction		
	Estimate (95% CI)	*p*		Estimate (95% CI)	*p*		Estimate (95% CI)	*p*	
YEDU									
0-10	[Reference]			[Reference]			[Reference]		
11-16	0.42 (0.07, 0.77)	.018	*	-1.24 (-2.12, -0.35)	.005	**	-1.21 (-2.07, -0.35)	.006	**
16+	0.38 (0.00, 0.76)	.050		-1.26 (-2.22, -0.30)	.010	*	-1.23 (-2.16, -0.31)	.009	**
Chao1	-	-		-0.14 (-0.76, 0.44)	.640		-0.21 (-0.54, 0.11)	.198	
YEDU:Chao1	-	-					-	-	
11-16:Chao1	-	-		-0.02 (-0.80, 0.79)	.967		-	-	
16+:Chao1	-	-		-0.20 (-1.00, 0.62)	.630		-	-	

Note. Regression coefficients for mediator and outcome models, used for mediation analysis with alpha diversity metric Chao1. All analyses were adjusted for age, sex/gender, use of antibiotic medication in the last 6 months, mild depressive symptoms based on the Beck Depression Inventory I, first language, partnership status and Apolipoprotein E ε4 status. YEDU=Years of education. *p<.05. **p<.01. ***p<.001.

Appendix III Table S2 *Coefficients of Regression Models With Shannon*

Variable	Mediator Model			Outcome Model					
				With Interaction			Without Interaction		
	Estimate (95% CI)	*p*		Estimate (95% CI)	*p*		Estimate (95% CI)	*p*	
Years of Education									
0-10	[Reference]			[Reference]			[Reference]		
11-16	0.36 (0.01, 0.71)	.042	*	-1.31 (-2.18, -0.45)	.003	**	-1.26 (-2.11, -0.41)	.003	**
16+	0.37 (-0.01, 0.75)	.055		-1.28 (-2.22, -0.36)	.007	**	-1.25 (-2.18, -0.34)	.007	**
Shannon	-	-		-0.01 (-0.61, 0.59)	.985		-0.13 (-0.46, 0.19)	.423	
Years of Education:Shannon	-	-					-	-	
11-16:Shannon	-	-		0.04 (-0.77, 0.87)	.924		-	-	
16+:Shannon	-	-		-0.38 (-1.19, 0.43)	.354		-	-	

Note. Regression coefficients for mediator and outcome models, used for mediation analysis with alpha diversity metric Shannon. All analyses were adjusted for age, sex/gender, use of antibiotic medication in the last 6 months, mild depressive symptoms based on the Beck Depression Inventory I, first language, partnership status and Apolipoprotein E ε4 status. YEDU=Years of education. *p<.05. **p<.01. ***p<.001.

Appendix III Table S3 *Coefficients of Regression Models With Inverse Simpson*

Variable	Mediator Model		Outcome Model					
			With Interaction			Without Interaction		
	Estimate (95% CI)	*p*	Estimate (95% CI)	*p*		Estimate (95% CI)	*p*	
Years of Education								
0-10	[Reference]		[Reference]			[Reference]		
11-16	0.17 (-0.18, 0.52)	.342	-1.33 (-2.18, -0.48)	.002	**	-1.29 (-2.14, -0.46)	.002	**
16+	0.22 (-0.16, 0.61)	.260	-1.30 (-2.23, -0.39)	.005	**	-1.28 (-2.21, -0.38)	.006	**
Inverse Simpson	-	-	-0.01 (-0.68, 0.63)	.975		-0.08 (-0.42, 0.25)	.625	
Years of Education:Inverse Simpson	-	-				-	-	
11-16:Inverse Simpson	-	-	0.19 (-0.62, 1.02)	.649		-	-	
16+:Inverse Simpson	-	-	-0.50 (-1.41, 0.41)	.282		-	-	

Note. Regression coefficients for mediator and outcome models, used for mediation analysis with alpha diversity metric Inverse Simpson. All analyses were adjusted for age, sex/gender, use of antibiotic medication in the last 6 months, mild depressive symptoms based on the Beck Depression Inventory I, first language, partnership status and Apolipoprotein E ε4 status. YEDU=Years of education. *p<.05. **p<.01. ***p<.001.

Appendix III Table S4 *Mediation Analysis With Shannon Index as Mediator*

Estimand	Comparing 0-10 to 11-16 Years of Education						Comparing 0-10 to 16+ Years of Education					
	With Interaction			Without Interaction			With Interaction			Without Interaction		
	Estimate (95% CI)	*p*		Estimate (95% CI)	*p*		Estimate (95% CI)	*p*		Estimate (95% CI)	*p*	
RCDE	0.31 (0.14, 0.69)	.007	**	0.32 (0.15, 0.69)	.007	**	0.32 (0.13, 0.74)	.012	*	0.33 (0.14, 0.75)	.014	*
RPNDE	0.31 (0.14, 0.69)	.007	**	0.32 (0.15, 0.70)	.007	**	0.36 (0.15, 0.86)	.022	*	0.33 (0.14, 0.75)	.014	*
RTNDE	0.31 (0.14, 0.70)	.008	**	0.32 (0.15, 0.70)	.007	**	0.32 (0.14, 0.74)	.011	*	0.33 (0.14, 0.75)	.014	*
RPNIE	1.00 (0.80, 1.25)	.986		0.96 (0.83, 1.09)	.526		1.00 (0.79, 1.26)	.985		0.96 (0.81, 1.10)	.531	
RTNIE	1.01 (0.79, 1.40)	.897		0.96 (0.82, 1.10)	.526		0.88 (0.65, 1.09)	.260		0.96 (0.81, 1.10)	.531	
RTE	0.31 (0.15, 0.67)	.005	**	0.31 (0.15, 0.67)	.004	**	0.32 (0.15, 0.72)	.009	**	0.31 (0.14, 0.71)	.008	**
ERCDE	-0.56 (-0.77, -0.22)	.007	**	-		-	-0.55 (-0.78, -0.18)	.012	*	-	-	
ERINTREF	-0.13 (-0.26, 0.03)	.093		-	-		-0.09 (-0.21, 0.17)	.344		-	-	
ERINTMED	0.01 (-0.26, 0.23)	.961		-	-		-0.04 (-0.37, 0.17)	.677		-	-	
ERPNIE	0.00 (-0.20, 0.25)	.986		-	-		0.00 (-0.21, 0.26)	.985		-	-	
ERCDE(P)	0.81 (0.58, 1.06)	.003	**	-	-		0.80 (0.56, 1.05)	.005	**	-	-	
ERINTREF(P)	0.19 (-0.07, 0.46)	.094		-	-		0.13 (-0.34, 0.34)	.336		-	-	
ERINTMED(P)	-0.01 (-0.43, 0.44)	.963		-	-		0.06 (-0.28, 0.71)	.682		-	-	
ERPNIE(P)	0.00 (-0.42, 0.35)	.988		-	-		0.00 (-0.43, 0.37)	.988		-	-	
PM	-0.01 (-0.21, 0.17)	.898		0.02 (-0.06, 0.17)	.527		0.06 (-0.05, 0.49)	.266		0.02 (-0.05, 0.20)	.535	
INT	0.19 (-0.03, 0.45)	.070		-	-		0.19 (0.00, 0.49)	.053		-	-	
PE	0.19 (-0.06, 0.42)	.107		-	-		0.20 (-0.05, 0.44)	.089		-	-	

Note. Results of mediation analysis with or without interaction terms of education and Shannon in the outcome model. Standard errors were estimated with 5,000 bootstraps. RCDE=controlled direct effect odds ratio (referring to CDE); RPNDE=pure natural direct effect odds ratio (referring to NDE); RTNDE=total natural direct effect odds ratio; RPNIE=pure natural indirect effect odds ratio; RTNIE=total natural indirect effect odds ratio (referring to NIE); RTE=total effect odds ratio; ERCDE=excess relative risk due to controlled direct effect; ERINTREF=excess relative risk due to reference interaction; ERINTMED=excess relative risk due to mediated interaction; ERPNIE=excess

relative risk due to pure natural indirect effect; ERCDE(P)=proportion ERCDE; ERINTREF(P)=proportion ERINTREF; ERINTMED(P)=proportion ERINTMED; ERPNIE(P)=proportion ERPNIE; PM=overall proportion mediated; INT=overall proportion attributable to interaction; PE=overall proportion eliminated. Cells with – indicate n/a. *p<.05. **p<.01. ***p<.001.

Appendix III Table S5 *Mediation Analysis With Inverse Simpson Index as Mediator*

Estimand	Comparing 0-10 to 11-16 Years of Education						Comparing 0-10 to 16+ Years of Education					
	With Interaction			Without Interaction			With Interaction			Without Interaction		
	Estimate (95% CI)	*p*		Estimate (95% CI)	*p*		Estimate (95% CI)	*p*		Estimate (95% CI)	*p*	
RCDE	0.31 (0.14, 0.68)	.005	**	0.31 (0.15, 0.67)	.005	**	0.32 (0.13, 0.73)	.010	*	0.32 (0.14, 0.73)	.009	**
RPNDE	0.30 (0.15, 0.66)	.004	**	0.31 (0.15, 0.67)	.005	**	0.36 (0.16, 0.83)	.020	*	0.32 (0.14, 0.73)	.009	**
RTNDE	0.31 (0.15, 0.69)	.006	**	0.31 (0.15, 0.67)	.005	**	0.32 (0.14, 0.75)	.010	*	0.32 (0.14, 0.73)	.009	**
RPNIE	1.00 (0.85, 1.20)	.964		0.99 (0.89, 1.08)	.832		1.00 (0.83, 1.25)	.991		0.98 (0.87, 1.09)	.762	
RTNIE	1.03 (0.89, 1.24)	.705		0.99 (0.88, 1.08)	.832		0.91 (0.67, 1.10)	.370		0.98 (0.86, 1.10)	.762	
RTE	0.31 (0.15, 0.67)	.005	**	0.31 (0.15, 0.67)	.005	**	0.32 (0.15, 0.73)	.009	**	0.31 (0.14, 0.71)	.008	**
ERCDE	-0.56 (-0.77, -0.24)	.005	**	-	-		-0.55 (-0.78, -0.20)	.010	*	-	-	
ERINTREF	-0.13 (-0.25, 0.00)	.055		-	-		-0.09 (-0.19, 0.17)	.326		-	-	
ERINTMED	0.01 (-0.19, 0.18)	.966		-	-		-0.03 (-0.36, 0.12)	.701		-	-	
ERPNIE	0.00 (-0.15, 0.20)	.964		-	-		0.00 (-0.17, 0.25)	.991		-	-	
ERCDE(P)	0.82 (0.61, 1.02)	.004	**	-	-		0.81 (0.58, 1.10)	.003	**	-	-	
ERINTREF(P)	0.20 (0.00, 0.44)	.054		-	-		0.14 (-0.35, 0.33)	.318		-	-	
ERINTMED(P)	-0.01 (-0.33, 0.33)	.965		-	-		0.05 (-0.21, 0.67)	.705		-	-	
ERPNIE(P)	0.00 (-0.34, 0.27)	.965		-	-		0.00 (-0.40, 0.31)	.990		-	-	
PM	-0.01 (-0.14, 0.07)	.705		0.01 (-0.05, 0.09)	.833		0.05 (-0.07, 0.41)	.375		0.01 (-0.05, 0.12)	.764	
INT	0.18 (0.02, 0.40)	.039	*	-	-		0.18 (-0.04, 0.48)	.084		-	-	
PE	0.18 (-0.02, 0.39)	.066		-	-		0.19 (-0.10, 0.42)	.114		-	-	

Note. Results of mediation analysis with or without interaction terms of education and Inverse Simpson in the outcome model. Standard errors were estimated with 5,000 bootstraps. RCDE=controlled direct effect odds ratio (referring to CDE); RPNDE=pure natural direct effect odds ratio (referring to NDE); RTNDE=total natural direct effect odds ratio; RPNIE=pure natural indirect effect odds ratio; RTNIE=total natural indirect effect odds ratio (referring to NIE); RTE=total effect odds ratio; ERCDE=excess relative risk due to controlled direct effect; ERINTREF=excess relative risk due to reference interaction; ERINTMED=excess relative risk due to mediated interaction;

ERPNIE=excess relative risk due to pure natural indirect effect; ERCDE(P)=proportion ERCDE; ERINTREF(P)=proportion ERINTREF; ERINTMED(P)=proportion ERINTMED; ERPNIE(P)=proportion ERPNIE; PM=overall proportion mediated; INT=overall proportion attributable to interaction; PE=overall proportion eliminated. Cells with – indicate n/a. *p<.05. **p<.01. ***p<.001.

Appendix III Table S6 *Taxonomic Classification of Identified Taxa*

Taxon	Domain	Phylum	Class	Order	Family	Genus
Bacilli	Bacteria	Firmicutes	Bacilli	NA	NA	NA
Actinobacteria	Bacteria	Actinobacteriota	Actinobacteria	NA	NA	NA
Lactobacillales	Bacteria	Firmicutes	Bacilli	Lactobacillales	NA	NA
Streptococcaceae	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	NA
Streptococcus	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
Lachnospiraceae UCG 001	Bacteria	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Lachnospiraceae UCG 001
ASV 000508	Bacteria	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Lachnospiraceae UCG 001
ASV 000053	Bacteria	Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	NK4A214_group

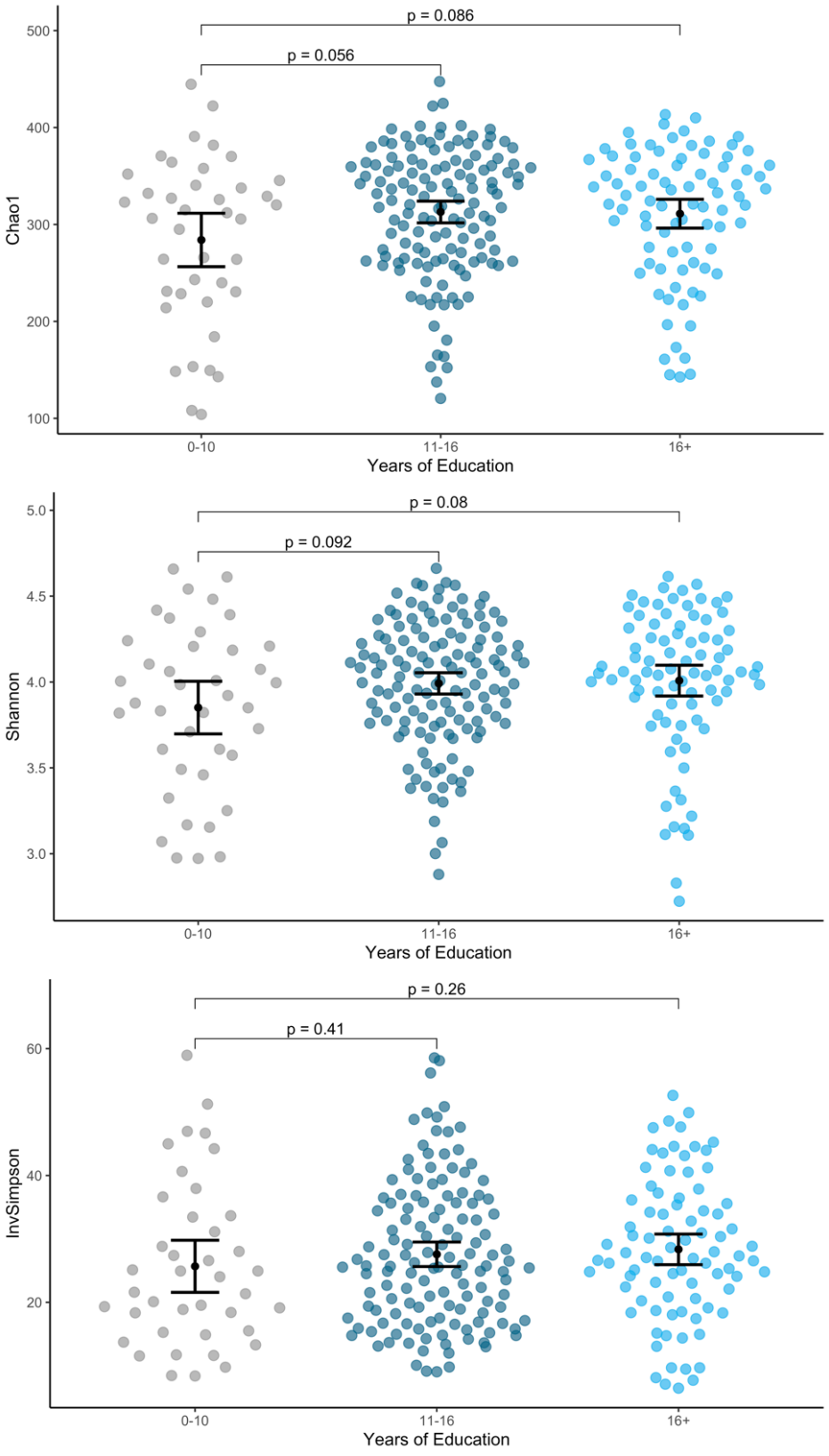
Note. Taxonomic classification as identified with *DESeq2* and *ancombc.* ASV=amplicon sequence variant.

Appendix III Table S7 *First Languages Spoken*

First Language	NC (*n*=200)	MCI (*n*=58)
Danish	3	0
Dutch	2	1
English	7	0
French	27	5
German	20	3
Hungarian	1	0
Italian	2	2
Luxembourgish	135	44
Portuguese	1	3
Slovene	1	0
Spanish	1	0

Note. NC=Normal Cognition; MCI=Mild cognitive impairment.

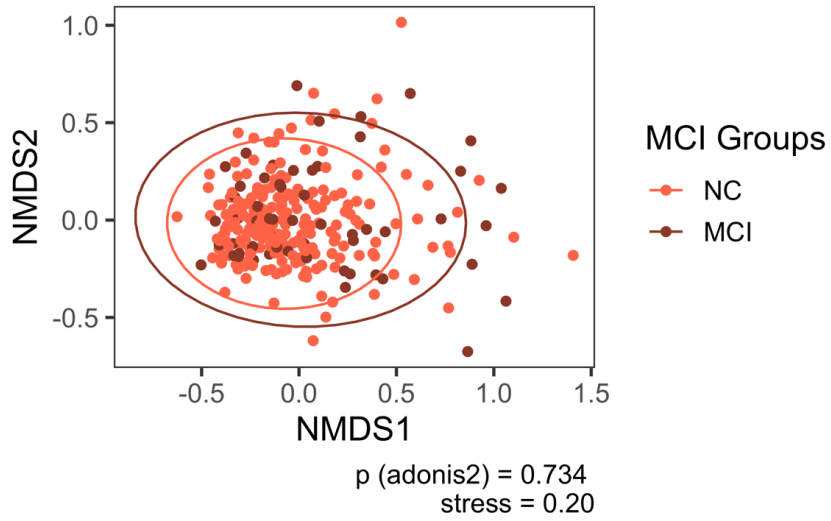
Appendix III Figure S1 *Alpha Diversity Across Education Groups*



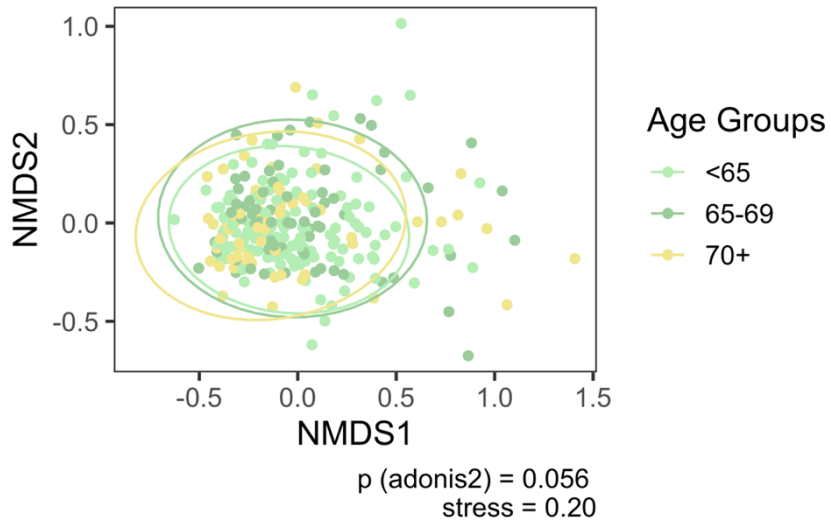
Note. Panels show results stratified by education groups with 0-10, 11-16 and 16+ years of education. Reported *p* values refer to Student's t-Tests. InvSimpson=Inverse Simpson.

Appendix III Figure S2 *Ordination Plots for MCI and Age Groups*

A MCI Groups

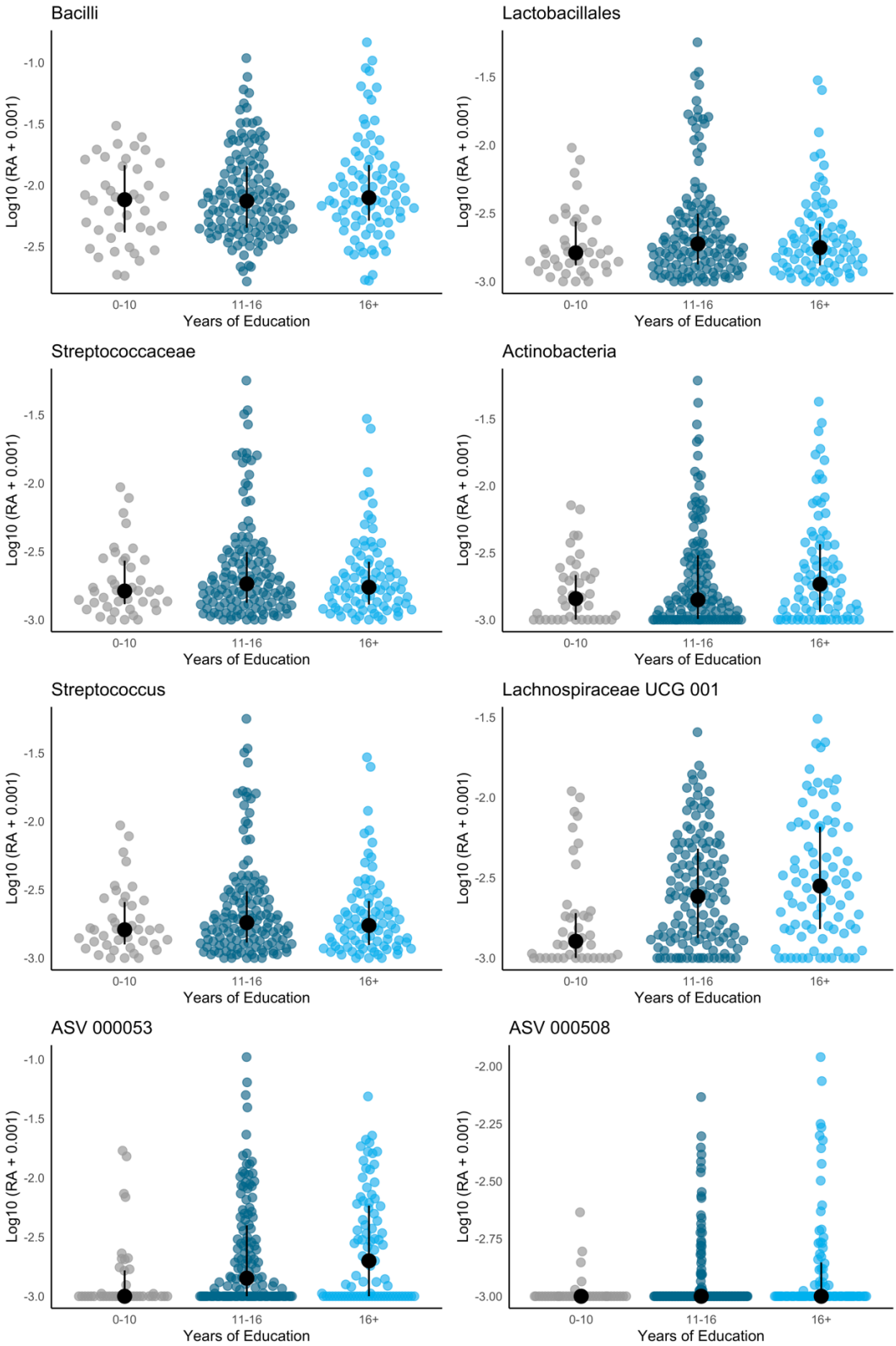


B Age Groups



Note. Ordination using NMDS based on Bray-Curtis dissimilarity for A MCI and B Age groups. Analysis with *adonis2* was adjusted for sex/gender, education, use of antibiotic medication in the last 6 months, mild depressive symptoms based on the Beck Depression Inventory I, first language, partnership status, and apolipoprotein ε4 status. A additionally adjusted for age and MCI, and B additionally adjusted for age categories. Authors MK and VTEA. NMDS=Non-metric Multidimensional Scaling.

Appendix III Figure S3 *Relative Abundance Across Groups of Education*



Note. Y axis shows log10-transformed relative abundance plus 10^{-3}. Bars indicate median and interquartile range. Taxa identified with *DESeq2* and *ancombc.* ASV=amplicon sequence variant. Authors MK and VTEA.

Appendix IV – Supplementary Material Chapter IV

Appendix IV Table S1 *Comparison of Items Constituting the Langa-Weir Algorithm*

Characteristics	HRS	Recall	Recall & IADL
Cognitive Function			
Immediate Recall	SR	SR	SR
Delayed Recall	SR	SR	SR
Serial 7's	SR	-	-
Backward Counting	SR	-	-
IADL			
Preparing Meals	Proxy	-	SR
Shopping Groceries	Proxy	-	SR
Making Phone Calls	Proxy	-	SR
Taking Medication	Proxy	-	SR
Managing Money	Proxy	-	SR
Using a Map	-	-	SR
Doing Housework	-	-	SR
Independent Mobility	-	-	SR
Doing Laundry	-	-	SR

Note. Proxy-rated memory and interviewer-perceived quality of cognition are not included. Recall=Langa-Weir algorithm based on Recall; Recall & IADL=Langa-Weir algorithm based on Recall and IADL; HRS=Health and Retirement Study; IADL=Instrumental Activities of Daily Living; Proxy=Reported by Proxy Respondent; SR=Self-Reported.

Appendix IV Table S2 *Performance of Classification Algorithms in the Test Set*

Classification	Accuracy	Balanced Accuracy	Sensitivity	Specificity	Precision	F1	AUC
LW (Recall)	0.96	0.64	0.31	0.97	0.18	0.23	0.64
LW (Recall)^{P}	0.92	0.73	0.53	0.93	0.14	0.22	0.73
LW (Recall & IADL)	0.97	0.63	0.27	0.98	0.27	0.27	0.63
LW (Recall & IADL)^{P}	0.96	0.70	0.43	0.97	0.23	0.30	0.70
GLM	0.98	0.55	0.11	1.00	0.51	0.18	0.88
GLM weighted	0.95	0.75	0.53	0.96	0.21	0.30	0.89
GLM DOWN	0.84	0.80	0.76	0.84	0.09	0.17	0.88
GLM SMOTE	0.93	0.77	0.60	0.94	0.16	0.26	0.88
RF	0.98	0.52	0.04	1.00	0.70	0.07	0.90
RF DOWN	0.84	0.81	0.77	0.84	0.10	0.17	0.89
RF SMOTE	0.93	0.77	0.60	0.94	0.17	0.26	0.88
XGB	0.98	0.50	0.00	1.00	1.00	0.01	0.89
XGB DOWN	0.83	0.80	0.77	0.84	0.09	0.16	0.88
XGB SMOTE	0.93	0.77	0.62	0.93	0.16	0.26	0.86

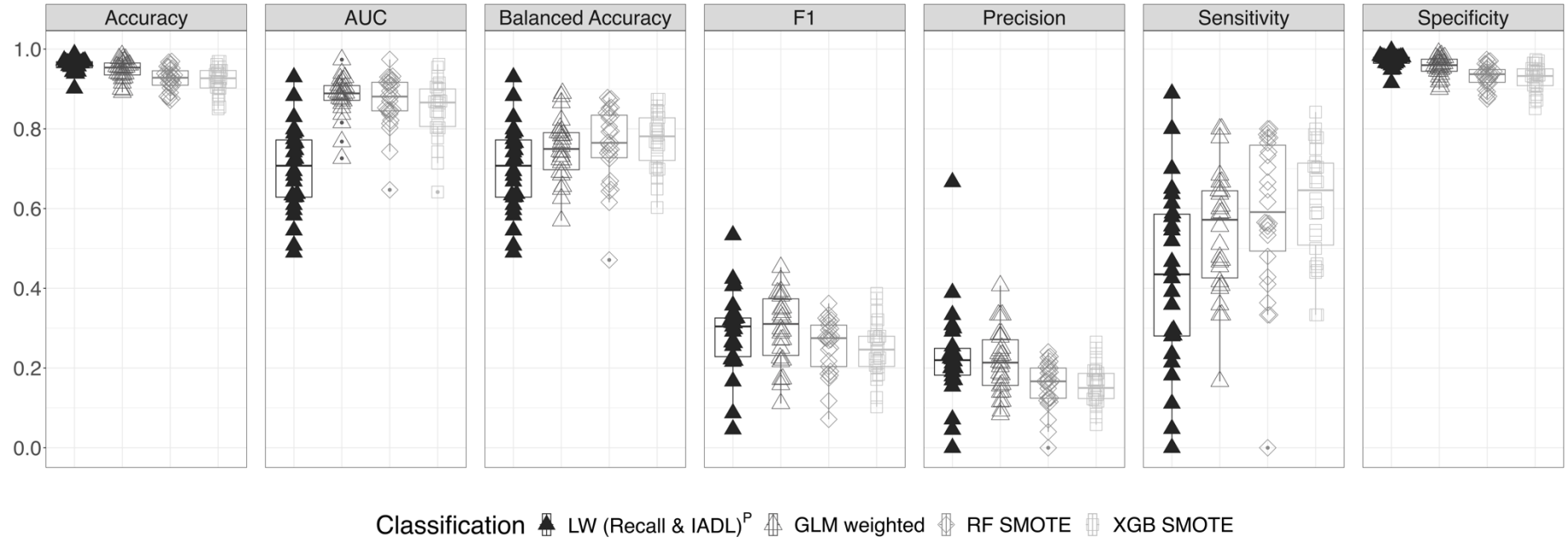
Note. AUC=Area under the receiver operating characteristic curve; LW (Recall)=Langa-Weir algorithm with a Recall-cutoff reflecting the 2.5th percentile; LW (Recall)^{P}=Langa-Weir algorithm with a Recall-cutoff reflecting country-level dementia prevalence; LW (Recall & IADL)=Langa-Weir algorithm based on LW (Recall) with an IADL cutoff reflecting 1.5 IQR above Q3; LW (Recall & IADL)^{P}=Langa-Weir algorithm based on LW (Recall)^{P} with an IADL cutoff reflecting 1.5 IQR above Q3; GLM=Logistic Regression, RF=Random Forest; XGB=XGBoost; DOWN=trained in data created with downsampling; SMOTE=trained in data created with the synthetic minority oversampling technique; IADL=Instrumental Activities of Daily Living.

Appendix IV Table S3 *Dementia Prevalence and Number of Expected Dementia Cases*

Country			OECD		SR-PD		LW (R&I)[P]		GLM (weighted)		RF SMOTE		XGB SMOTE	
	ISO	*n*	Prevalence	*n*	Prevalence	*n*	Prevalence	*n*	Prevalence	*n*	Prevalence	*n*	Prevalence	*n*
Austria	AT	1,267	7.22	91.52	3.75	47.47	5.10	64.63	4.89	61.99	10.11	128.05	14.37	182.04
Belgium	BE	1,791	7.24	129.73	1.96	35.08	5.02	89.85	5.57	99.73	7.74	138.69	9.32	166.96
Bulgaria	BG	700	4.10	28.68	1.67	11.69	2.11	14.77	7.07	49.49	8.61	60.24	7.02	49.13
Croatia	HR	860	2.83	24.37	2.70	23.25	2.70	23.24	8.33	71.61	9.71	83.48	8.35	71.81
Cyprus	CY	467	5.86	27.38	2.93	13.68	8.96	41.86	9.06	42.31	11.87	55.44	14.70	68.67
Czech Rep.	CZ	1,774	4.16	73.87	1.88	33.35	3.07	54.49	3.73	66.09	4.60	81.67	4.59	81.47
Denmark	DK	1,164	6.45	75.10	0.80	9.35	4.31	50.22	3.07	35.75	4.01	46.69	3.91	45.48
Estonia	EE	1,864	5.78	107.71	1.93	35.93	2.75	51.30	8.39	156.39	9.22	171.89	8.88	165.53
Finland	FI	680	6.67	45.35	2.79	18.97	3.12	21.25	2.86	19.44	7.75	52.73	7.93	53.93
France	FR	1,307	7.69	100.46	1.70	22.25	5.21	68.05	5.07	66.24	6.31	82.53	4.93	64.40
Germany	DE	1,406	7.27	102.21	2.16	30.33	3.35	47.08	3.47	48.82	4.89	68.74	5.03	70.72
Greece	GR	1,263	7.60	96.00	1.73	21.86	5.35	67.57	5.33	67.3	7.69	97.13	7.40	93.44
Hungary	HU	646	4.27	27.60	1.04	6.72	2.41	15.55	4.40	28.44	6.67	43.08	7.48	48.33
Israel	IL	821	6.00	49.26	3.20	26.25	4.69	38.52	9.83	80.67	11.22	92.10	11.47	94.16
Italy	IT	1,710	7.98	136.41	2.38	40.63	8.08	138.24	6.39	109.2	8.66	148.03	7.59	129.76
Latvia	LV	574	5.63	32.29	1.87	10.76	0.92	5.27	6.37	36.57	8.84	50.73	10.81	62.04
Lithuania	LT	668	5.80	38.72	3.12	20.85	3.40	22.73	9.52	63.58	11.85	79.17	14.01	93.59
Luxembourg	LU	404	6.77	27.34	1.50	6.07	2.87	11.61	3.96	16.01	8.23	33.23	9.79	39.56
Malta	MT	468	6.04	28.27	1.28	5.99	5.31	24.83	3.48	16.28	6.36	29.76	7.48	35.02
Poland	PL	1,599	4.16	66.55	2.41	38.59	3.82	61.08	6.84	109.35	9.58	153.14	9.16	146.47
Portugal	PT	489	7.26	35.50	4.02	19.68	4.30	21.03	9.04	44.23	14.61	71.46	12.34	60.34
Romania	RO	714	4.10	29.28	1.49	10.63	3.81	27.23	9.33	66.64	11.35	81.02	9.19	65.62
Slovenia	SI	1,435	4.50	64.62	3.22	46.19	3.76	54.01	6.46	92.75	9.18	131.78	11.88	170.42
Spain	ES	1,827	7.57	138.25	3.66	66.78	7.59	138.70	7.28	132.96	10.59	193.48	9.35	170.81
Sweden	SE	1,407	7.06	99.28	1.33	18.74	3.18	44.69	2.16	30.36	4.32	60.85	6.44	90.57
Switzerland	CH	1,004	7.15	71.76	1.11	11.18	2.64	26.51	1.70	17.09	3.32	33.31	3.28	32.92
Total		28 309		1,747.51		632.27		1,224.31		1,629.29		2,268.40		2,353.18

Note. Population-weighted prevalence and number of cases based on the test set. Numbers are rounded to the second decimal. Three participants excluded due to missing sampling weights. OECD=Data from the Organisation for Economic Co-operation and Development and a population-based study in Israel (Kodesh, 2019; OECD, 2018); SR-PD=self-reported physician-diagnosis of dementia; LW $(R\&I)^{P}$=Langa-Weir algorithm with a Recall-cutoff reflecting country-level dementia prevalence and a cutoff reflecting 1.5 IQR above Q3 for Instrumental Activities of Daily Living; GLM=Logistic Regression, RF=Random Forest; XGB=XGBoost; SMOTE=trained in data created with the synthetic minority oversampling technique.

Appendix IV Figure S1 *Performance Variation in the Test Set Across Countries*



Note. AUC=Area under the receiver operating characteristic curve; LW (Recall & IADL)^{P}=Langa-Weir algorithm with a Recall-cutoff reflecting country-level dementia prevalence and a cutoff reflecting 1.5 IQR above Q3 for Instrumental Activities of Daily Living; GLM=Logistic Regression, RF=Random Forest; XGB=XGBoost; SMOTE=trained in data created with the synthetic minority oversampling technique.