Title

Methodological Innovations to Assess Detection Bias and Infectious Disease Risk Factors for Dementia in Real-World Data

Author

Wang, Jingxuan

Publication Date

2025

Peer reviewed|Thesis/dissertation

Methodological Innovations to Assess Detection Bias and Infectious Disease Risk Factors for Dementia in Real-World Data

by
Jingxuan Wang

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Epidemiology and Translational Science

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

Signed by:
M. Maria Glymour
B96511A557AB4E4...
M. Maria Glymour
Chair

DocuSigned by:
Dan Kelly
Dan Kelly

DocuSigned by:
Renaud La Joie
Renaud La Joie

DocuSigned by:
John Kornak
B8EB0AA53B334CD...
John Kornak

Committee Members

Acknowledgments

There are so many people who have been a part of this journey, and I am deeply grateful to each of you. Just as a single tree cannot make a forest, I would not be where I am today without the collective strength, guidance, and support of so many.

First and foremost, I extend my deepest gratitude to my amazing mentor, Dr. M. Maria Glymour, thank you for taking a chance on me five years ago, despite my completely different background. Your belief in me, your inspiration, and your unwavering support have not only shaped my work in ways I could never have imagined, but also the way I approach challenges and growth. I am incredibly fortunate to have you as a mentor.

To my committee members, Drs. J. Dan Kelly, Renaud La Joie, and John Kornak—Dan, thank you for always encouraging me to grow, while offering your constant support every step of the way. Renaud, I am deeply grateful for how you lifted me during my struggles and for the inspiring example of dedication and resilience you've set. John, your generous support, kindness, and the thoughtful words you've shared—whether in moments of celebration or struggle—along with your invaluable expertise and guidance, have meant more to me than I can say and have truly kept me going.

Thank you to the faculty and staff at the UCSF Department of Epidemiology and Biostatistics, especially Eva Wong-Moy, Catie Oldenburg, Pam Murnane, and Dave Glidden. Thank you for fostering a supportive and welcoming environment that made this journey so much more fulfilling. To my fellow PhD students, especially my cohort—

Carolyn Smith Hughes, Eduardo Rodriguez Almaraz, Kate Chirikova, Leah Koenig, Priz Espinosa Tamez, and Sirena Gutierrez—thank you for your encouragement, warmth, humor, and endless support. Sharing this journey with you has made the weight of challenges lighter and the joy of milestones brighter.

I am also incredibly grateful to my fellows in the Glymour Group, especially Ruijia Chen, Sarah Ackley, Erin Ferguson, Yulin Yang, Peter Buto, Annie Pederson, Minhyuk Choi, Scott Zimmerman, and Robin Hyatt. Your collaboration, insight, and camaraderie have enriched my journey. Thank you for the stimulating discussions, the countless moments of support, and the shared passion for discovery. Your encouragement and friendship have made even the most challenging days more rewarding, and I feel truly fortunate to have been part of such an inspiring team.

I owe a big thank you to my other mentors along the road and previous mentors. To María M. Corrada and Eric Tchetgen Tchetgen—thank you for teaching me epidemiology is a way of learning and how to use rigorous, novel design to do good to the public health. To Deborah Blacker—thank you for generously sharing your insights and expertise that brightened my future paths. To Fei Jiang—thank you for seeing my potential and always believing in me.

Finally, I am endlessly grateful to my family. To my parents—Xia Liu and Zhengguo Wang—thank you for your unconditional love and for always being there for me, no matter the distance. To my wife, Shengzi Zeng, your constant love, support, patience,

and belief in me have been my greatest source of strength. When I was overwhelmed, you grounded me. When I doubted myself, you believed in me more than I believed in myself. You've celebrated my wins, lifted me through my struggles, and stood beside me with grace and unshakable faith. I truly could not have done this without you. You are my rock. And a special thank you to Bubble, my furry companion, for your love and constant source of comfort throughout this journey.

Contributions

A version of Chapter 1 of this dissertation has been published in *JAMA Network Open* (see citation below) in 2025. The published material is substantially the product of Jingxuan Wang's period of study at UCSF and was primarily conducted and written by him. The work he completed for this published manuscript is comparable to a standard dissertation chapter.

Wang J, Choi M, Buto P, Kelly JD, La Joie R, Kornak J, Zimmerman SC, Chen R, Raphael E, Schaefer CA, Blacker D, Glymour MM. (2025). Potential for detection bias in EHR-based research on clinical predictors of dementia. JAMA Network Open. 2025;8(4):e256637. doi:10.1001/jamanetworkopen.2025.6637

Methodological Innovations to Assess Detection Bias and Infectious Disease Risk Factors for Dementia in Real-World Data

Jingxuan Wang

Abstract

This dissertation applies methodological innovations in real-world data to examine three areas: (1) the extent of detection bias in electronic health record (EHR)-based research on clinical predictors of dementia, (2) the potential impact of COVID-19-related brain MRI changes on long-term dementia incidence, and (3) the association between the age at onset of herpes viral infection and dementia incidence. All three chapters utilize data from the UK Biobank, with Chapter 1 also incorporating EHR data from the All of Us Research Program. Each chapter uses a unique analytical approach. Chapter 1 presents the first empirical evidence of detection bias in EHR-based dementia research, demonstrating that healthcare encounters for common conditions increase the likelihood of dementia detection, leading to spurious associations between clinically diagnosed conditions and subsequent dementia risk. Chapter 2 finds that naturally occurring variations in imaging-derived phenotypes linked to COVID-19 are associated with higher dementia incidence, suggesting a potential neurobiological link between COVID-19 and dementia. Chapter 3 finds that individuals with a herpes simplex virus (HSV) infection before age 65 have an increased incidence of dementia compared to those without an HSV infection. This dissertation contributes to dementia epidemiology by addressing key methodological challenges and leveraging large-scale real-world data to enhance our understanding of infectious disease-related risk factors for dementia.

Table of Contents

List of Figures

List of Tables

Introduction

Dementia is a progressive syndrome associated with a decline in cognitive function severe enough to interfere with daily life and independence.[1] It is characterized by impairments in memory, reasoning, problem-solving, and other cognitive abilities, often accompanied by changes in behavior and personality. The most common subtype, Alzheimer's disease (AD) dementia, is driven by the accumulation of β-amyloid plaques and neurofibrillary tangles, with pathological changes occurring decades before clinical symptoms emerge.[2]

Currently, over 57 million individuals worldwide are living with dementia, and the number is projected to increase to 153 million by 2050 as the population ages.[3] Dementia places significant health, economic, and social costs on individuals, families, healthcare systems, and societies. While recent clinical trials have led to the development of AD drugs, their cognitive benefits remain uncertain, as does whether these small improvements outweigh potential side effects.[4–7] In addition, these drugs primarily target individuals already in advanced stages of the progression. To mitigate the growing impact of dementia, prevention strategies targeting the general population are crucial. Understanding modifiable risk factors is essential for delaying or preventing dementia onset.

Despite growing evidence on dementia risk factors, current knowledge attributes only about 45% of dementia cases to known modifiable factors, highlighting the need to identify additional factors contributing to dementia risk.[8] Infectious diseases, in

particular, may play a key role by triggering chronic inflammation and neuroinflammation, which contribute to neurodegeneration and dementia pathogenesis and progression.[9] For example, COVID-19 has been linked to cognitive decline.[10] Yet infectious diseases have received less attention than metabolic, socioeconomic, and behavioral risk factors. This gap is partly due to methodological challenges, including limited cohort studies with measurement on infectious diseases, small sample sizes, and difficulties in measuring infections accurately.

The surge of real-world datasets, for example, electronic health records (EHRs)-based studies, presents an opportunity to examine infectious disease risk factors for dementia. EHR data provide rich clinical information, including infectious disease diagnoses and antibody/antigen test histories. The use of EHR data for dementia research is particularly appealing due to large sample sizes, allowing more precise estimates and investigation of rare exposures, diverse subpopulations, and dementia subtypes.

The objective of this dissertation is to leverage methodological innovations to advance dementia research using real-world data, with a particular focus on evaluating infectious disease risk factors.

Chapter 1 investigates detection bias in EHR-based dementia research. Since EHR systems are not designed for research, they present methodological challenges for research purposes. One challenge is clinical events are recorded only when patients have healthcare encounters, leading to potential detection bias. This chapter presents

the first empirical evidence of detection bias in EHR-based dementia research using two large-scale EHR-based cohorts: UK Biobank and All of Us.

Chapter 2 explores the link between COVID-19 and dementia. Neurocognitive effects of COVID-19 may increase dementia risk, but evidence is limited. A key challenge in studying this association is underreporting of COVID-19 cases due to asymptomatic infections and limited early testing. Additionally, widespread exposure to COVID-19 further complicates finding a suitable comparison group who have not been infected. To address this, this chapter utilizes the short-term neurologic hallmarks of COVID-19, previously established with brain magnetic resonance imaging (MRI),[11] as proxies for COVID-19's neurological effects to assess its long-term impact on dementia risk.

Chapter 3 examines the association between age at onset of herpes viral infection and dementia. While prior studies have established associations between herpes viral infection and dementia,[12] findings have been inconsistent, possibly due to heterogeneity in susceptibility to neurological effects of herpes viral infections across age groups and challenges in accurately measuring infection history. Antibody test data provide precise infection history but suffer from small sample sizes, whereas clinical diagnoses offer larger cohorts but face underdiagnosis and missing records. To address this challenge, this chapter leverages UK Biobank data, augmenting clinical diagnosis records with antibody tests from a large randomly selected subsample, to investigate the association between age at onset of five types of herpes viral infection and dementia incidence.

Chapter 1

Potential for detection bias in electronic health record-based research on clinical predictors of dementia

Jingxuan Wang, Minhyuk Choi, Peter Buto, J. Daniel Kelly, Renaud La Joie, John Kornak, Scott C. Zimmerman, Ruijia Chen, Eva Raphael, Cathy A. Schaefer, Deborah Blacker, M. Maria Glymour

Abstract

Importance: Detection bias occurs when exposure is associated with a systematic difference in outcome ascertainment, and in the context of dementia research, health conditions that bring patients into frequent interactions with healthcare may increase the chance that an individual receives a dementia diagnosis. Although such detection bias is a major concern when evaluating the associations of clinical exposures with dementia incidence, no prior empirical work attempts to quantify its magnitude.

Objective: To evaluate potential detection bias in evaluation of clinical predictors of dementia incidence using EHR data.

Design: Prospective cohort study.

Setting: Population-based volunteer cohorts: UK Biobank (UKB) and All of Us (AoU).

Participants: UKB (n=137,374) and AoU (n=91,018) participants 55 years or older, dementia-free at baseline, with linked EHRs.

Exposures: Diagnoses of type 2 diabetes mellitus (T2DM), depression, hypertension, urinary tract infection (UTI), kidney stones, forearm fracture, and gastrointestinal (GI) bleeding.

Main Outcomes and Measures: Incident all-cause dementia diagnosis from EHRs.

Results: The final UKB analytic sample comprised 137,374 participants (mean [SD] age at baseline, 62.5 [4.1] years; 73,913 females [53.8%]). The AoU analytic sample included 91,018 participants (mean [SD] age at baseline, 66.9 [7.8] years; 51,946 females [57.1%]). With mean follow-up of 13.2 and 2.7 years, UK Biobank and All of Us accrued 4,095 (3.0% of eligible individuals) and 952 (1.0%) incident dementia diagnoses, respectively. In both datasets, type 2 diabetes mellitus, depression, hypertension, urinary tract infection, kidney stones, forearm fracture, and GI bleeding were all significantly associated with dementia incidence, with the hazard ratios (HR) ranging from 1.18 to 3.51. For most exposures, the highest dementia incidence occurred in the first year following exposure diagnosis (with HRs ranging from 1.33 to 7.59) with attenuated associations thereafter (with HRs ranging from 1.08 to 3.52). For example, GI bleeding had the highest dementia incidence in the first year (HR=2.17, 95% CI=1.46-3.22 in UKB; HR=2.56, 95% CI=1.62-4.04 in AoU), and the association attenuated over 1-5 years (HR=1.46, 95% CI=1.15-1.86 in UKB; HR=2.14, 95% CI=1.63-2.81 in AoU).

Conclusion and Relevance: In this cohort study of two large datasets, diagnoses of a spectrum of conditions with varying etiological influences on dementia were associated with a higher short-term likelihood of dementia diagnosis. Healthcare encounters for any condition increased the probability of dementia detection, leading to a spurious association between conditions requiring clinical care and subsequent dementia diagnoses.

Introduction

Electronic health record (EHR)-based studies on dementia diagnoses, risk prediction, risk factor associations, and other related areas have rapidly expanded.[13–16] EHR data offers rich clinical information and a real-world representation of patient health over time. The use of EHR data for dementia research is appealing due to large sample sizes,[17] allowing more precise estimates and investigation of rare exposures, diverse subpopulations, and dementia subtypes. However, EHR information systems were not built for research. EHRs engender distinct research challenges: clinical events are recorded only when patients have healthcare encounters, leading to potential detection bias.[18,19]

Detection bias occurs when exposure is associated with a systematic difference in outcome ascertainment or diagnosis. For example, otherwise healthy older adults who would meet clinical criteria for dementia if they were assessed may not frequently interact with clinicians, thus delaying the dementia diagnosis. Health conditions that bring patients into frequent interaction with healthcare, where they receive diagnoses, may increase the chance that an individual receives a dementia diagnosis. While under-detection of dementia in EHR systems is well-documented,[20] detection bias presents a distinct challenge, as many cases might remain undiagnosed unless the individual interacts with the healthcare system. This phenomenon is sometimes called “informative presence bias” or “unmasking (detection signal) bias” and can bias estimated associations of medical conditions or clinical care with subsequent dementia risk.[21–24]

Despite widespread acknowledgment of potential detection bias, there is almost no evidence on the likely magnitude of this bias in EHR-based research on determinants of dementia. Understanding the empirical impact of detection bias is essential for interpreting clinical research results. Even estimating lower or upper bounds for associations that could plausibly be entirely attributable to detection bias would be helpful to interpret results from clinical EHR research. Although it is methodologically challenging to evaluate the bias magnitude, we propose two strategies that can each offer insight (**Figure 1.1**).

Our first approach contrasts EHR-derived estimated associations to estimates derived from existing cohort studies or meta-analyses of cohort studies. Cohort studies assess dementia on a fixed schedule guided by the research design, and thus diagnosis of dementia should not suffer the same type of detection bias as EHR-based research. While cohort studies may be subject to other biases, such as confounding, these biases are likely to be similar across both cohort and EHR-based data sources. We examine a spectrum of conditions—with varying etiological influences on dementia—and their subsequent relationship to healthcare utilization. The conditions include not only well-established risk factors for dementia but also those with minimal plausible mechanisms linked to dementia, in the spirit of negative control exposures.[25]

Our second approach to quantifying detection bias is premised on the expectation that detection bias will follow a distinctive time pattern, with the strongest association with new diagnoses immediately following clinical encounters, and subsequently declining

over time. For example, although diabetes may influence dementia risk, this effect is thought to accumulate over time, with longer duration of diabetes associated with higher dementia risk. The influence of healthcare encounters on dementia diagnoses, however, would likely peak around the time of diabetes diagnoses. The time pattern of outcome diagnoses vis-a-vis clinical encounters may thus offer insights into the magnitude of detection bias.

In this study, we aimed to estimate the plausible range of values of detection bias in EHRs by examining a range of exposures: type 2 diabetes mellitus (T2DM), depression, hypertension, urinary tract infection (UTI), kidney stones, forearm fracture, and gastrointestinal (GI) bleeding. We assessed the associations of each exposure diagnosis with dementia using two real-world EHR databases. In addition to determining the extent to which detection bias in EHR-based studies may affect research findings in dementia research, we also investigated whether each exposure was associated with healthcare utilization and whether healthcare utilization predicts dementia diagnosis.

Methods

Study population and participants

The UK Biobank (UKB) is a prospective cohort that enrolled over 500,000 participants aged 40-69 years from 2006 to 2010.[26] At baseline, participants visited one of 22 UK assessment centers and completed questionnaires, interviews, and physical and medical assessments. Follow-up data included EHR-based hospital admission and primary care data and national death register data. Hospital admission data were

available for all participants while only around 45% of participants had primary care data linked to UKB. In our analysis, we included participants who were 55 years or older, free of dementia at baseline, and with both inpatient and primary care EHR data available. Participants were followed from the baseline (2006-2010) until December 2022, with a median follow-up of 13.2 years.

The All of Us (AoU) Research Program is an ongoing large-scale research initiative enrolling diverse US participants starting from 2017, with data from baseline health surveys, physical measurements, biospecimens, and linkage to EHRs. Individuals aged 18 years or older residing in the US are eligible to enroll in the program online or via 340+ recruitment sites.[27] Participants completed baseline surveys at enrollment and around 95% of them choose to provide authorization to share EHR data. In parallel with UKB data, our AoU analyses included participants who were 55 years or older, free of dementia at baseline, and had EHR data. Participants were followed from the baseline (2017-2022) until July 2022, with a median follow-up of 2.7 years.

Ascertainment of exposures and dementia outcomes

For each dataset, we identified exposure and all-cause dementia diagnoses from participants' EHRs. We identified the first documented incidence of each exposure diagnosis. Dementia outcome included Alzheimer's dementia (AD), vascular dementia, Lewy body dementia, alcohol-related dementia, other types of dementia, and unspecified dementia. In UKB, death records were also used to ascertain dementia diagnosis. Hospital inpatient and death records were coded using International

Classification of Diseases, Ninth Revision (ICD-9) and ICD-10 in UKB; primary care records were coded using Read version 2 (Read v2) and Clinical Terms Version 3 (CTV3). In AoU, dementia and exposures were ascertained using ICD-9, ICD-10, and the corresponding Systematized Nomenclature of Medicine (SNOMED) codes.

Healthcare utilization

Because increased healthcare utilization was our primary hypothesized mechanism for detection bias, we evaluated whether the exposures affected utilization, measured by the number of encounters with the healthcare system recorded in the EHR. In UKB, we tallied the total number of general practitioner (GP) and inpatient encounters. In AoU, we counted all-cause patient encounters, including primary care records, inpatient records, procedures, pre-procedural examinations, routine medical examinations, follow-up examinations, screenings, etc. Participants with an incident exposure had their healthcare utilization calculated for 1 year prior to and 1 year after the exposure diagnosis. For those without a history of the exposure, healthcare utilization was calculated for 1 year prior to and 1 year after enrollment.

Covariates

We considered two covariate sets. The base model included age, sex (female, male), race (Asian, Black, White, Other), and number of apolipoprotein (*APOE*)-ε4 alleles (0, 1, 2). The full model additionally adjusted for education (high school or above, less than high school), smoking history (ever smoked, never smoked), and continuous body mass index (BMI) as a linear term. In the UKB, age and sex were acquired from the NHS

Primary Care Trust registries and were updated by participants at baseline. AoU participants self-reported age and sex. Race, education, and smoking history were self-reported in both datasets using baseline surveys. BMI was constructed from height and weight measured during the baseline physical measurement in each dataset and calculated as weight (kg) divided by height (m) squared.

Statistical analysis

Analyses were conducted using R statistical software version 4.4.0 (R Project for Statistical Computing). All statistical tests were two-sided, with a significance level of $\alpha = 0.05$. Data were analyzed from November 2023 through February 2025. In primary analyses, we adopted two approaches to assess detection bias: 1) comparing EHR-based estimated associations of clinical exposures with dementia diagnoses to previously published estimates from cohort studies, where available; and 2) evaluating how associations between incident clinical exposures and incident dementia diagnoses evolve over time. In secondary analysis we assessed associations of healthcare utilization with our exposures and incident dementia.

Primary analysis

We summarized characteristics of each sample and calculated cumulative incidence of dementia while accounting for the competing risk of death. We used Cox proportional hazards regression models to examine associations between each exposure and incident dementia diagnosis. Participants were followed from study baseline (date of enrollment into study) to the earliest of: date of first dementia diagnosis, death from any

cause, or administrative censoring—defined as the latest date of any dementia diagnosis in the sample (UKB: December 12, 2022; AoU: July 1, 2022). Each exposure diagnosis was treated as a time-varying exposure, with the exposure status updated if participants developed an incident diagnosis after enrollment: those with an incident exposure diagnosis contributed person-time to the “unexposed” group until their date of exposure diagnosis, after which they contributed to person-time in the incident exposure group. Participants with prevalent exposure, i.e., who had a documented exposure diagnosis before baseline, were included in the primary analyses.

We compared our EHR-derived estimated associations to benchmark estimates derived from existing meta-analyses of cohort studies. The benchmark meta-analyses were selected after a PubMed search (see Supplement for search criteria) and we used the most recent or the biggest meta-analyses, or if no meta-analysis was available, the most recently published cohort study. To investigate the pattern of detection bias over time for each exposure, we calculated HRs for incident dementia over 4 time intervals after the earliest exposure diagnosis (0–1 year, >1–5 years, >5–10 years, and >10 years after exposure diagnosis), compared to participants with no prior exposure diagnosis.

In pre-planned analysis, we aimed to establish a range of detection bias magnitudes in EHR by meta-analyzing the estimated associations of all exposures with dementia diagnoses from the full model, using a random-effects model with inverse variance weights. The estimated associations used in the meta-analysis were calculated by dividing the HR 0–1 year after the exposure diagnosis by the overall HR, which was

estimated in a separate Cox model over the entire follow-up period in our primary analysis, canceling out the effects of potential residual confounding, true causal relationships, and other study biases commonly encountered in both estimates. Standard errors were calculated assuming independence.

Sensitivity analysis for primary results

We conducted four sensitivity analyses. First, we hypothesize that individuals with high healthcare utilization are less likely to have undiagnosed dementia and thus less vulnerable to detection bias when a new condition is diagnosed. Thus, we stratified our primary analysis by levels of healthcare utilization (low and high, by sample-specific median number of healthcare encounters during the year before baseline). Second, we restricted our analysis to participants with no prevalent exposure diagnosis at study baseline and evaluated the association of incident exposure with dementia diagnosis. Third, in our meta-analysis, instead of dividing the HR 0–1 year after the exposure diagnosis by the overall HR, we replaced the overall HR by the HR 1-2 years after the exposure as well as the HR derived from cohort studies or meta-analyses of cohort studies for established risk factors.[28–31] Lastly, since the benchmark meta-analysis of T2DM was stratified by sex, we also stratified our primary analysis of T2DM by sex.

Secondary analysis

We conducted secondary analyses to investigate the associations of each exposure with healthcare utilization, and the association of healthcare utilization with incident dementia. We used a difference-in-differences negative binomial regression model to

compare the change in healthcare utilization among exposed groups to a control group.[32] For each exposure, the exposed time-at-risk included participants who had already received a first-diagnosis of the exposure, while the comparison time-at-risk comprised participants who had not yet (and might never) receive an exposure diagnosis. We compared the change in healthcare utilization before and after the incidence of each exposure in the exposed group to the change in healthcare utilization before and after baseline in the control group. Further details on the difference-in-differences design can be found in the Supplement. To assess association between healthcare utilization and dementia, we fit Cox models with frequency of healthcare utilization one year prior to baseline as the exposure, adjusting for baseline age, sex, race, and education.

Results

Study Sample Characteristics

The 137,374 UKB participants who met inclusion criteria averaged 62.5 years of age (SD=4.1); 73,913 (53.8%) were female and 133,053 (96.9%) identified as White (**Table 1.1**). The proportion of participants with a documented exposure diagnosis prior to baseline ranged from 1.6% for kidney stone to 27.8% for hypertension. The proportion of participants receiving a first exposure diagnosis during follow-up ranged from 1.4% for kidney stone to 20.7% for hypertension. Over a mean follow-up of 13.2 years, 4,095 new dementia diagnoses accrued, showing a clear age gradient (**Figure 1.2**).

Among the 91,018 AoU participants who met our inclusion criteria, the mean (SD) age was 66.9 years (7.8); 51,946 (57.1%) were females, and 61,853 (68.0%) identified as White. The proportion of participants with a documented exposure diagnosis prior to baseline ranged from 2.4% for forearm fracture to 50.1% for hypertension. The proportion of participants receiving a first exposure diagnosis during follow-up ranged from 0.7% for forearm fracture to 6.1% for hypertension. Over a mean follow-up time of 2.7 years, 952 new dementia diagnoses occurred; dementia incidence followed the expected age-gradient (**Figure 1.2**).

Associations of clinical exposures with dementia: Strategies 1-2

In both datasets, all seven clinical exposures showed a statistically significant positive association with dementia incidence. For instance, T2DM was associated with a higher dementia hazard (HR=1.94, 95% CI=1.79-2.10 in UKB and HR=2.43, 95% CI=2.11-2.79 in AoU; **Figure 1.3**, **Table 1.S1**). The benchmark meta-analysis of 14 cohort studies showed that T2DM was associated with a relative risk of 1.62 (95% CI=1.45-1.80) for women and 1.58 (95% CI=1.38-1.81) for men,[28] both of which were smaller than our EHR-derived estimates.

In UKB, patients with at least one UTI episode had a 1.85 times the hazard of dementia (95% CI=1.73-1.98) compared to participants without a UTI diagnosis; in AoU, this HR was 2.32 (95% CI=2.02-2.67). Both EHR-based estimates were more pronounced than the estimated HR of 1.13 (95% CI=1.08–1.18) from the benchmark cohort study.[31]

For most exposures, the highest incidence of dementia occurred in the first year following exposure diagnosis compared to participants without exposure diagnosis; after the first post-diagnosis year, the association attenuated (**Figure 1.3**, **Table 1.S1**). For example, participants diagnosed with a GI bleeding had the highest dementia incidence in the first year after a GI bleeding diagnosis (HR=2.17, 95% CI=1.46-3.22 in UKB and HR=2.56, 95% CI=1.62-4.04 in AoU; **Figure 1.3**, **Table 1.S1**). Subsequently, during the 1-5 years following diagnosis, the hazard ratio associated with GI bleeding attenuated to 1.46 (95% CI=1.15-1.86) in UKB and 2.14 (95% CI=1.63-2.81) in AoU.

Assessment of detection bias magnitude from meta-analysis

The average ratio of each exposure estimate in the first year following the exposure to the long-term average association, when meta-analyzed across all exposures in both cohorts, was 1.60 (95% CI=1.15-2.22; **Figure 1.4**). In other words, in the first year after exposure diagnosis, the likelihood of receiving a dementia diagnosis was elevated by 60% more than the long-term average elevation in dementia risk associated with the exposure. However, ratios ranged from 0.74 to 3.53, and the I^2 statistic, i.e., the percentage of variability in estimated associations due to heterogeneity in associations rather than sample size differences, was 83.8%.

Sensitivity analyses

Among participants with high levels of healthcare utilization before baseline, the associations of clinical exposures with new dementia diagnoses were slightly smaller than overall associations (**Figure 1.S1**), resulting in a meta-analyzed estimated

association of 1.57 (95% CI=1.07-2.30; **Figure 1.S2**). Conversely, participants with low levels of healthcare utilization before baseline experienced greater estimated detection bias (meta-analyzed HR=1.70, 95% CI=1.31-2.20; **Figure 1.S2**). Excluding participants with prevalent exposure diagnosis at baseline did not substantially change the results compared to our primary analysis (**Figure 1.S3**). Changing reference hazard ratios for the meta-analysis had an impact on the estimated association: comparing dementia incidence 0-1 to 1-2 years after exposure (instead of the overall average estimate) led to a meta-analyzed estimate of detection bias of 1.41 (95% CI=1.18-1.67; **Figure 1.S4**), while comparing to existing meta-analyses for established risk factors increased the estimated association to 2.02 (95% CI=1.33-3.07; **Figure 1.S5**). Associations between T2DM and dementia were similar between males and females in both datasets (**Table 1.S2**), but both were higher than the benchmark associations.

Secondary analyses

The difference-in-differences estimates for changes in healthcare utilization between the incident exposure group and control group varied with exposure and healthcare utilization measurement. However, for most exposures, healthcare utilization of participants with an incident exposure diagnosis increased compared to their counterparts without an exposure history. For example, the frequency of healthcare utilization increased by 15.2% (95% CI=11.0%-19.6%; **Table 1.2**) within one year following a forearm fracture diagnosis in UKB and by 27.7% (95% CI=25.2%-30.2%) in AoU. Additionally, healthcare utilization during the year before baseline predicted dementia incidence, with a hazard ratio of 1.03 per additional encounter (95% CI=1.02-

1.03; **Table 1.S3**) in UKB and a hazard ratio of 1.02 per additional encounter (95% CI=1.01-1.02) in AoU.

Discussion

We evaluated the associations of seven exposures with incident dementia using two large EHR-based cohorts. Our exposures included a spectrum of conditions with varying etiological influences on dementia. We found that all exposures, including those with minimal link to dementia, were positively associated with new dementia diagnoses recorded in the EHR. The observed associations were more pronounced compared to existing evidence from previously published cohort studies. Given that most observed associations attenuated with time after exposure diagnosis, results likely reflect detection bias inherent in EHR data. However, the increased associations observed for T2DM and hypertension after longer follow-up are likely to reflect, at least partially, the cumulative effects of the exposure, rather than detection bias. Receipt of care for any condition brought the participant into contact with the healthcare system, and therefore increased the likelihood of being diagnosed with dementia. This increased diagnosis rate associated with healthcare utilization created a spurious association between any condition requiring clinical attention and subsequent incidence of dementia diagnosis. Our results suggest the magnitude of this detection bias—the ratio of the observed association under detection bias to the expected association in the absence of detection bias—may be substantial. We interpret the meta-analyzed bias estimate centered at HR of 1.60 with great caution, because the I^2 statistic of 83.8%, suggested substantial heterogeneity in detection bias across exposures.[33]

While the existence of detection bias has been well documented in prior literature,[21,22,34–38] few studies have offered a plausible range of its magnitude or derived estimates from real data. Such an approach is challenging because the ground truth is unknown: it is impossible to disentangle bias from true associations in real data. For instance, hypertension is a known risk factor for dementia.[8] Individuals with a history of hypertension may exhibit higher dementia diagnosis rates either due to an increased risk or because greater healthcare utilization increases their likelihood of receiving a diagnosis. Our approach acknowledges that no single association provides conclusive information, and that the magnitude of the bias varies by condition. It posits healthcare utilization as a common mechanism for the bias and uses unexpected associations across multiple exposures to provide insight into its magnitude. Despite significant heterogeneity in magnitudes across conditions, our results suggest detection bias can be substantial, and associations derived from EHRs merit careful scrutiny and efforts to mitigate for this bias.

Our findings may not be generalizable to other conditions. Dementia has a long preclinical phase, and overlapping and evolving symptoms and social factors may delay diagnoses.[2,39] Individuals exhibiting clinical symptoms closer to the diagnostic threshold who are concerned about their cognitive decline or dementia, or whose caregivers express concerns, are more likely to receive a dementia diagnosis.[39,40] This dynamic creates a contact-dependent nature to dementia diagnosis. This increased likelihood may arise from more opportunities to observe dementia symptoms; from the metabolic, emotional, or cognitive stress associated with illness or treatments; or from referrals to

specialists, particularly for conditions with overlapping clinical symptoms. Whether our findings extend to other outcomes whose diagnoses depend on interactions with the healthcare system, such as hypertension,[41] chronic kidney disease,[42] thyroid disorders,[43] etc., is an empirical question. The two strategies we presented for quantifying detection bias would be equally relevant for other outcomes. Although none of these strategies are foolproof or precise, in combination, they can provide an understanding of the likely scope and magnitude of detection bias.

Key limitations of this study include possible residual confounding, lack of evaluating differences across healthcare settings, potentially missing EHR records, and the convenience samples in both cohorts. We cannot rule out the presence of other study biases or alternative explanations that might have driven the observed associations. In particular, transient cognitive dysfunction due to delirium following acute conditions such as UTIs and fractures, as well as cognitive symptoms associated with depression, could contribute to the observed short-term surge in dementia diagnoses. Our analyses did not differentiate between healthcare encounters across different types of facilities, e.g., inpatient, outpatient, emergency room, or specialty care, which could influence the probability of receiving a dementia diagnosis or being referred to specialists.[44,45] This limitation restricted our ability to explore the detailed mechanisms underlying the bias. EHR data may be incomplete. Differences in the estimated associations of exposure with healthcare utilization may suggest the presence of missing records, especially for UKB (where the estimated associations are smaller), which might partly explain the discrepancies in estimated associations compared to AoU. Lastly, a potential healthy

volunteer selection bias might exist. UKB participants tended to be younger and healthier than the general population.[46] AoU participants differed from the general population in their sociodemographic, health-related, and clinical characteristics, suggesting non-representativeness and potential selection bias.[47] Indeed, the lower incidence rates of dementia compared to those in general populations suggests a healthy participant selection in both datasets.[48–50] None of these limitations, however, could plausibly explain our primary results.

In conclusion, we demonstrated a potential detection bias inherent in the analysis of EHRs when evaluating clinical risk factors for dementia. Such potential for detection bias should be directly evaluated and addressed in EHR-based research on clinical risk factors. Our strategies can be adapted to detect such bias, and analytical approaches should be considered to mitigate it, such as estimating time-varying associations. These findings also demonstrate the potential value of periodic direct cognitive assessments embedded in EHR-based follow-up studies. Direct cognitive assessments fielded at pre-specified time points allow evaluation of undiagnosed cognitive impairment or dementia. Future research leveraging more detailed clinical data may fruitfully offer more specific quantification of the magnitude of such bias across additional risk factors and samples. Studies using EHR data for epidemiological and clinical research should exercise caution and thoughtful consideration about context and potential limitations.

Tables

Table 1.1. Characteristics of participants in the analytical sample

Characteristic *n (%)*[a]	UK Biobank	All of Us
N	137,374	91,018
Mean age at baseline (SD)	62.5 (4.1)	66.9 (7.8)
Sex		
Female	73,912 (53.8%)	51,946 (57.1%)
Male	63,462 (46.2%)	39,072 (42.9%)
Race		
Asian	2,232 (1.6%)	1,561 (1.7%)
Black	862 (0.6%)	15,733 (17.3%)
White	133,053 (96.9%)	61,853 (68.0%)
Other	1,227 (0.9%)	11,871 (13.0%)
APOE-ε4 alleles		
0	98,642 (71.8%)	66,857 (73.5%)
1	35,550 (25.9%)	22,242 (24.4%)
2	3,182 (2.3%)	1,919 (2.1%)
Education		
High school or more	81,210 (59.1%)	83,372 (91.6%)
Less than high school	56,164 (40.9%)	7,646 (8.4%)
Smoking history		
Ever smoked	66,578 (48.5%)	40,913 (45.0%)
Never smoked	70,182 (51.1%)	48,252 (53.0%)
Missing	614 (0.4%)	1,853 (2.0%)
Mean BMI (SD)	27.7 (4.7)	29.6 (6.8)

Characteristic *n (%)*[a]	UK Biobank	All of Us
Mean follow-up time (SD)	13.2 (2.2)	2.7 (1.1)
Type 2 diabetes mellitus status		
No	119,713 (87.1%)	69,189 (76.0%)
Prevalent case before baseline	8,455 (6.2%)	18,811 (20.7%)
Incident case during follow-up	9,206 (6.7%)	3,018 (3.3%)
Depression status		
No	118,480 (86.2%)	67,093 (73.7%)
Prevalent case before baseline	11,511 (8.4%)	20,495 (22.5%)
Incident case during follow-up	7,383 (5.4%)	3,430 (3.8%)
Hypertension status		
No	70,721 (51.5%)	39,900 (43.8%)
Prevalent case before baseline	38,218 (27.8%)	45,563 (50.1%)
Incident case during follow-up	28,435 (20.7%)	5,555 (6.1%)
Urinary tract infection status		
No	106,008 (77.2%)	73,808 (81.1%)
Prevalent case before baseline	16,276 (11.8%)	14,146 (15.5%)
Incident case during follow-up	15,090 (11.0%)	3,064 (3.4%)
Kidney stone status		
No	133,167 (96.9%)	84,044 (92.3%)
Prevalent case before baseline	2,253 (1.6%)	5,498 (6.0%)
Incident case during follow-up	1,954 (1.4%)	1,476 (1.6%)
Forearm fracture status		
No	126,347 (92.0%)	88,160 (96.9%)
Prevalent case before baseline	6,651 (4.8%)	2,180 (2.4%)

Characteristic *n (%)*[a]	UK Biobank	All of Us
Incident case during follow-up	4,376 (3.2%)	678 (0.7%)
Gastrointestinal bleeding		
No	128,281 (93.4%)	82,095 (90.2%)
Prevalent case before baseline	3,201 (2.3%)	7,054 (7.8%)
Incident case during follow-up	5,892 (4.3%)	1,869 (2.1%)
Number of encounters with the healthcare system during the year before baseline (SD)	7.1 (7.7)	8.2 (14.7)
Number of encounters with the healthcare system during the year after baseline (SD)	7.9 (8.2)	8.4 (14.6)

Abbreviations: SD, standard deviation; BMI, body mass index (calculated as weight in kilograms divided by height in meters squared).
[a] Percentages may not sum to 100 due to rounding.

Table 1.2. Comparison of healthcare utilization change 1 year before and after exposure incidence: Estimates of percentage increase from difference-in-differences (DID) negative binomial regression models

	UK Biobank		All of Us	
Exposure	DID estimate of percentage increase[a]	95% CI	DID estimate of percentage increase[a]	95% CI
Type 2 diabetes mellitus	22.2%	19.0% to 25.5%	20.0%	18.8% to 21.2%
Depression	1.7%	-0.9% to 4.5%	13.3%	12.4% to 14.2%
Hypertension	10.0%	8.1% to 12.0%	15.0%	14.0% to 15.9%
Urinary tract infection	-2.1%	-4.2% to 0.0%	6.3%	5.3% to 7.3%
Kidney stone	0.9%	-4.4% to 6.5%	12.2%	10.8% to 13.7%
Forearm fracture	15.2%	11.0% to 19.6%	27.7%	25.2% to 30.2%
Gastrointestinal bleeding	-2.4%	-5.4% to 0.6%	17.0%	13.6% to 20.5%

[a] Adjusted for age, sex, race, and education.

Figures

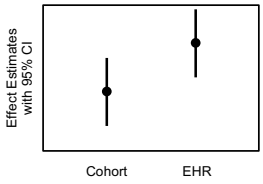
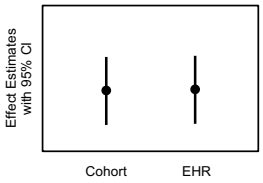
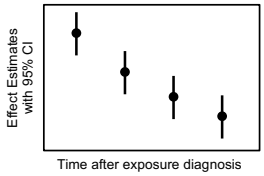
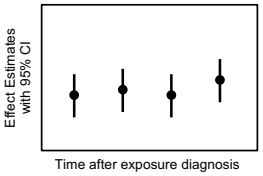
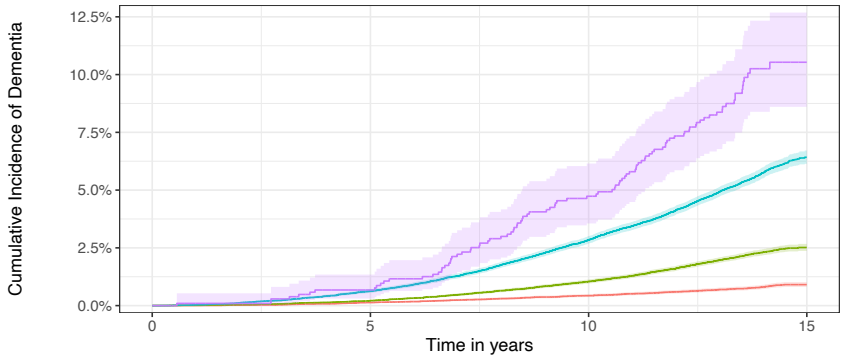
Strategy	Description	Hypothesized results under detection bias	Hypothesized results under no detection bias
Estimated associations comparison	Contrasting EHR-derived effect estimates to those derived from previously published cohort studies or meta-analyses of cohort studies, which are less susceptible to detection bias.		
Time pattern investigation	Evaluating the temporal pattern of exposure-outcome associations, which we expect to show the largest impact on new diagnoses immediately following clinical encounters, with subsequent decline over time.		

Figure 1.1. Two strategies to offer insight into the magnitude of detection bias

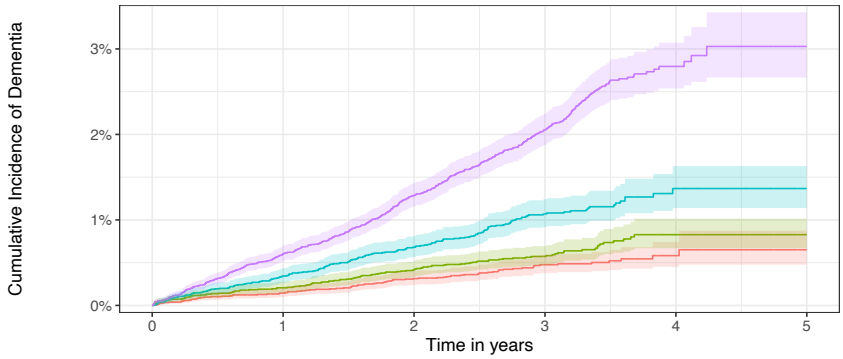
Abbreviations: EHR, electronic health record.

(A) UK Biobank



	At Risk			
55–60 years	40785	40234	39202	2917
60–65 years	54320	53271	50905	2918
65–70 years	41234	39808	36672	2158
70+ years	1035	978	876	36
	Events			
55–60 years	0	54	176	326
60–65 years	0	117	564	1209
65–70 years	0	258	1173	2297
70+ years	0	7	49	99

(B) All of Us



	At Risk					
55–60 years	20648	18454	17319	10083	1554	15
60–65 years	20539	18111	16875	9905	1675	16
65–70 years	19804	17185	15715	9136	1563	17
70+ years	30027	25768	23581	13330	2033	10
	Events					
55–60 years	0	29	59	82	87	88
60–65 years	0	41	78	101	117	117
65–70 years	0	65	120	172	185	185
70+ years	0	172	339	493	558	561

Figure 1.2. Cumulative incidence of dementia over time

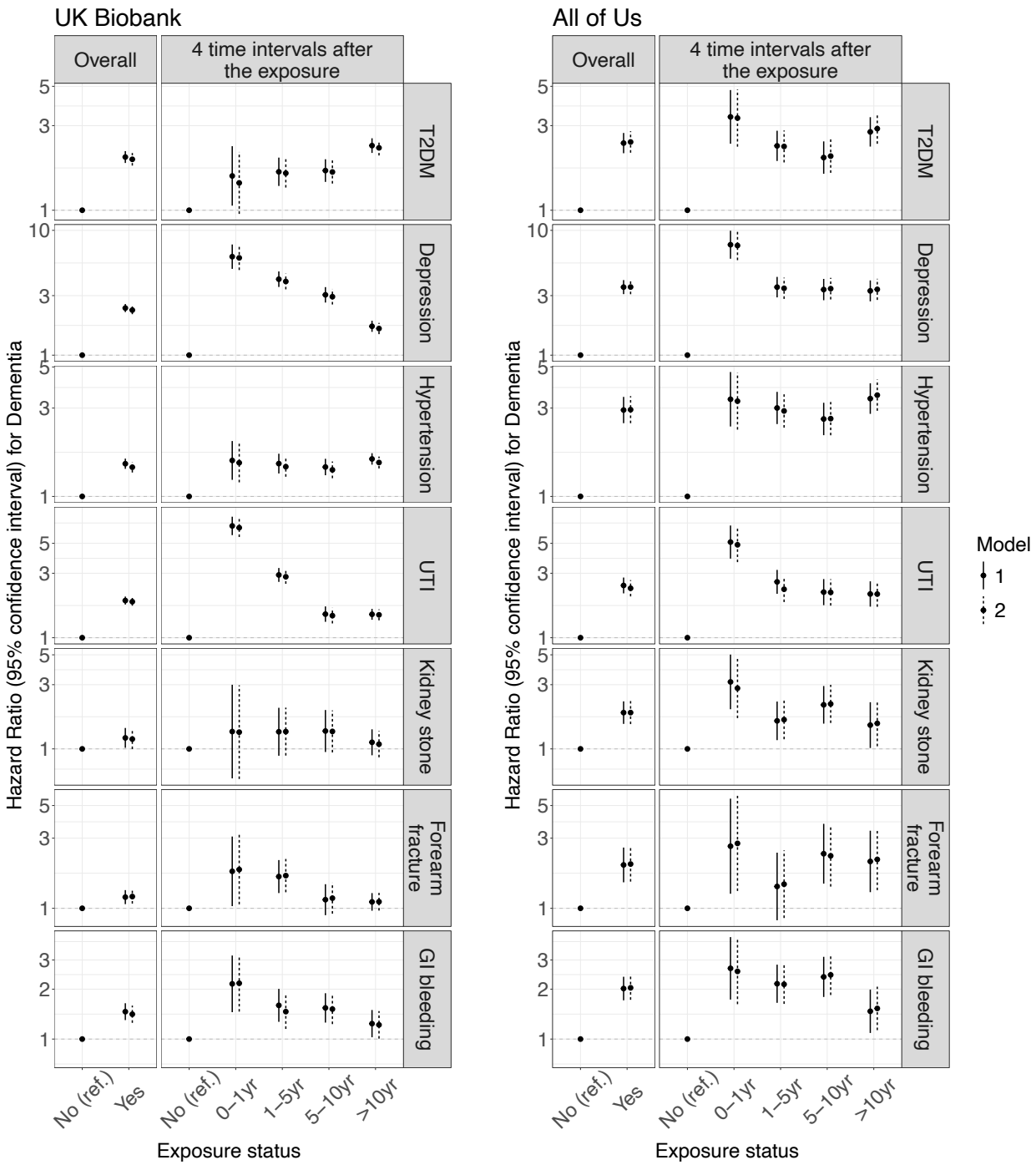


Figure 1.3. Associations of clinical exposures with incidence of first dementia diagnosis

Abbreviations: T2DM, type 2 diabetes mellitus; UTI, urinary tract infection; GI bleeding, gastrointestinal bleeding.
Note: Model 1 was adjusted for age, sex, race, and the number of APOE-ε4 alleles. Model 2 was further adjusted for education, smoking history, and BMI.

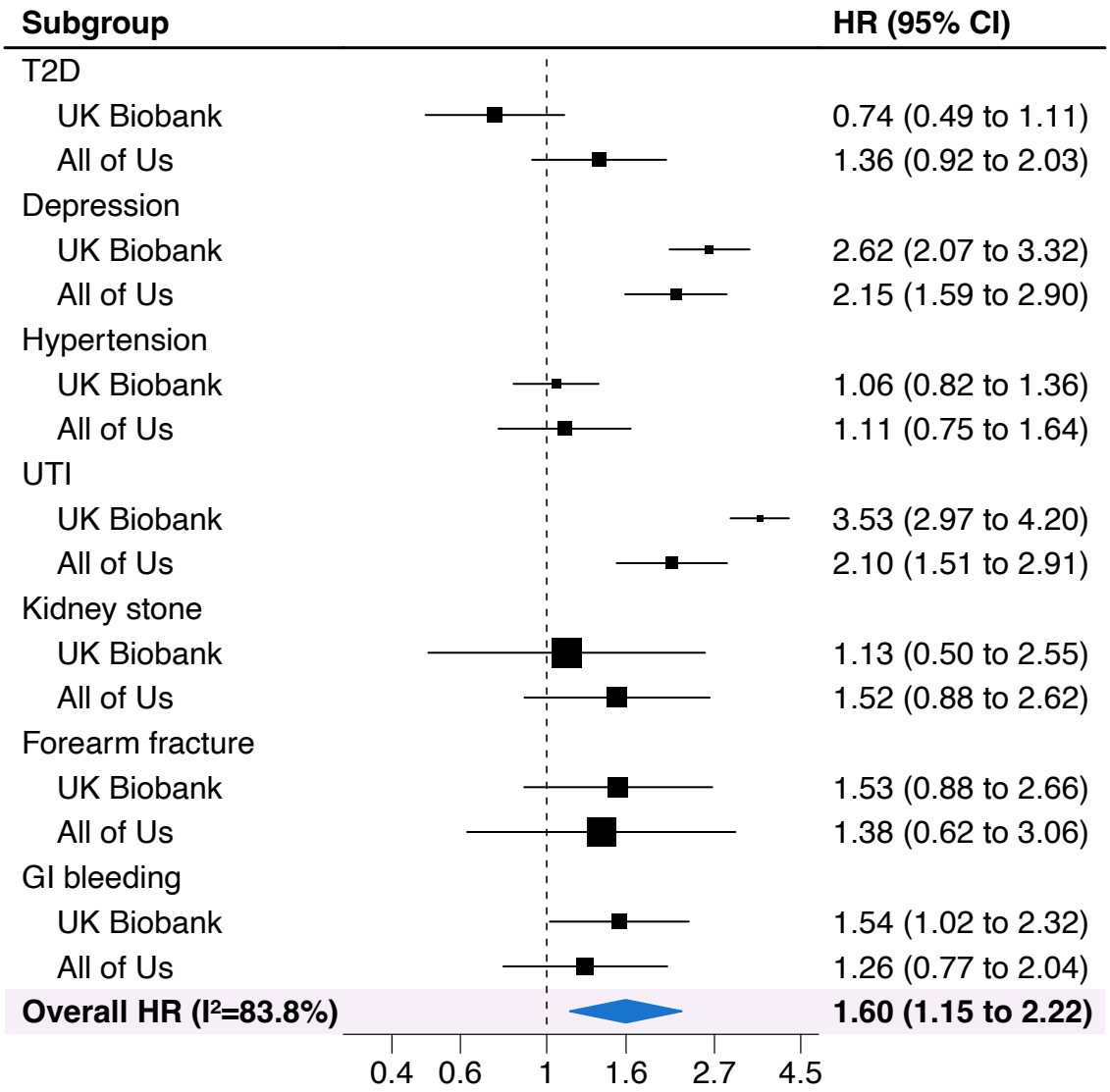


Figure 1.4. Forest plot of random-effects models for the pooled detection bias estimates

Abbreviations: T2DM, type 2 diabetes mellitus; UTI, urinary tract infection; GI bleeding, gastrointestinal bleeding.
Note: Point estimates for detection bias, which were used as inputs for the meta-analysis, were calculated by dividing the HR 0–1 year after the exposure diagnosis by the overall HR, allowing us to mitigate or cancel out the effects of potential residual confounding, true causal relationships, and other study biases commonly encountered in both estimates. We used a random-effects model with inverse variance weights. Size of squares represent the standard error of the point estimate.

Supplement

Search criteria for benchmark meta-analyses

Eligibility for the benchmark comparison estimates included: meta-analysis or cohort study if no meta-analysis exists, all-cause dementia as the outcome, and each clinical condition (type 2 diabetes mellitus, depression, hypertension, urinary tract infection, kidney stones, and forearm fracture) as the exposure.

Search criteria was defined as follows:

((meta-analysis) AND (dementia[MeSH Terms]) AND (type 2 diabetes[MeSH Terms]))

((meta-analysis) AND (dementia[MeSH Terms]) AND (depression[MeSH Terms]))

((meta-analysis) AND (dementia[MeSH Terms]) AND (hypertension[MeSH Terms]))

((meta-analysis) AND (dementia[MeSH Terms]) AND (urinary tract infection [MeSH Terms]))

((meta-analysis) AND (dementia[MeSH Terms]) AND (kidney stone[MeSH Terms]))

((meta-analysis) AND (dementia[MeSH Terms]) AND (forearm fracture[MeSH Terms]))

((meta-analysis) AND (dementia[MeSH Terms]) AND (gastrointestinal bleeding[MeSH Terms]))

Difference-in-difference (DID) design for the associations between exposure diagnosis and healthcare utilization

Due to the count data nature for the number of encounters with the healthcare system, the association of each exposure on healthcare utilization was estimated from an adjusted DID negative binomial regression model:

$$\ln(Y_{it}) = \beta_0 + \beta_1 \times X_i + \beta_2 \times D_t + \beta_3 \times X_i \times D_t$$

$$+ \beta_4 \times age_i + \beta_5 \times D_t \times age_i + \beta_6 \times sex_i + \beta_7 \times race_i + \beta_8 \times educ_i + \epsilon_{it},$$

where Y_{it} is the measurement of healthcare utilization of participant i at time t. X_i is a binary indicator (1=exposed group, 0=control group), D_t is a binary dummy variable (1=after incidence or baseline, 0=before incidence or baseline), age_i, sex_i, $race_i$, and $educ_i$ are individual baseline age, sex, race, and education, respectively. The DID estimate for the association between exposure diagnosis and healthcare utilization is represented by the parameter β_3.

Table 1.S1. Associations of exposures with dementia incidence

	UK Biobank				All of Us			
	Model1		Model2		Model1		Model2	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
T2DM								
Overall	2.00	1.85-2.15	1.94	1.79-2.10	2.39	2.10-2.73	2.43	2.11-2.79
0-1 year	1.56	1.06-2.30	1.43	0.95-2.13	3.36	2.37-4.76	3.31	2.28-4.80
1-5 years	1.65	1.37-1.98	1.62	1.34-1.95	2.31	1.90-2.81	2.29	1.86-2.82
5-10 years	1.67	1.44-1.94	1.64	1.41-1.91	1.98	1.61-2.44	2.02	1.63-2.51
>10 years	2.32	2.11-2.55	2.25	2.04-2.48	2.77	2.29-3.35	2.88	2.37-3.51
Depression								
Overall	2.38	2.21-2.57	2.30	2.13-2.48	3.51	3.08-3.99	3.51	3.08-4.01
0-1 year	6.15	4.91-7.70	6.02	4.81-7.54	7.67	5.92-9.93	7.56	5.77-9.90
1-5 years	4.06	3.52-4.69	3.89	3.37-4.50	3.51	2.90-4.24	3.43	2.82-4.17
5-10 years	3.04	2.64-3.51	2.94	2.55-3.39	3.35	2.76-4.08	3.41	2.79-4.16
>10 years	1.70	1.54-1.89	1.63	1.47-1.81	3.27	2.70-3.97	3.37	2.77-4.09
Hypertension								
Overall	1.50	1.41-1.60	1.44	1.35-1.54	2.93	2.49-3.45	2.94	2.49-3.48
0-1 year	1.56	1.23-1.99	1.52	1.19-1.94	3.35	2.39-4.70	3.27	2.30-4.66
1-5 years	1.50	1.33-1.70	1.45	1.28-1.64	3.00	2.46-3.67	2.89	2.35-3.56
5-10 years	1.44	1.30-1.60	1.39	1.25-1.54	2.62	2.14-3.20	2.63	2.14-3.24
>10 years	1.59	1.48-1.71	1.52	1.42-1.64	3.37	2.79-4.08	3.52	2.90-4.28
UTI								
Overall	1.88	1.76-2.02	1.85	1.73-1.98	2.44	2.13-2.79	2.32	2.02-2.67
0-1 year	6.73	5.75-7.87	6.52	5.57-7.65	5.11	3.85-6.77	4.87	3.63-6.55

	UK Biobank				All of Us			
	Model1		Model2		Model1		Model2	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
1-5 years	2.92	2.59-3.28	2.83	2.51-3.19	2.59	2.11-3.18	2.28	1.84-2.84
5-10 years	1.49	1.31-1.70	1.46	1.28-1.66	2.18	1.74-2.72	2.17	1.73-2.71
>10 years	1.49	1.36-1.63	1.48	1.34-1.62	2.11	1.70-2.61	2.10	1.69-2.62
Kidney stone								
Overall	1.21	1.02-1.43	1.18	1.00-1.40	1.86	1.53-2.25	1.86	1.53-2.27
0-1 year	1.35	0.60-3.00	1.33	0.60-2.97	3.15	1.97-5.02	2.82	1.69-4.71
1-5 years	1.34	0.89-2.02	1.34	0.89-2.03	1.62	1.16-2.25	1.65	1.18-2.30
5-10 years	1.36	0.95-1.94	1.35	0.94-1.93	2.13	1.54-2.93	2.16	1.56-2.99
>10 years	1.12	0.90-1.40	1.08	0.87-1.36	1.50	1.02-2.22	1.55	1.05-2.29
Forearm fracture								
Overall	1.19	1.07-1.33	1.20	1.08-1.34	1.97	1.50-2.59	2.00	1.52-2.63
0-1 year	1.78	1.03-3.08	1.84	1.06-3.17	2.64	1.26-5.57	2.76	1.31-5.82
1-5 years	1.64	1.27-2.13	1.67	1.29-2.17	1.41	0.83-2.39	1.46	0.86-2.48
5-10 years	1.14	0.90-1.46	1.17	0.92-1.49	2.35	1.47-3.75	2.27	1.41-3.68
>10 years	1.11	0.96-1.27	1.11	0.96-1.27	2.08	1.29-3.37	2.15	1.33-3.48
GI bleeding								
Overall	1.46	1.30-1.64	1.41	1.25-1.59	2.02	1.71-2.38	2.04	1.73-2.41
0-1 year	2.15	1.45-3.19	2.17	1.46-3.22	2.67	1.73-4.12	2.56	1.62-4.04
1-5 years	1.60	1.27-2.01	1.46	1.15-1.86	2.16	1.65-2.81	2.14	1.63-2.81
5-10 years	1.54	1.26-1.89	1.52	1.23-1.87	2.37	1.79-3.13	2.44	1.84-3.22
>10 years	1.24	1.03-1.50	1.22	1.01-1.47	1.47	1.09-1.99	1.53	1.13-2.07

Abbreviations: T2DM, type 2 diabetes mellitus; UTI, urinary tract infection; GI bleeding, gastrointestinal bleeding.

Table 1.S2. Association between type 2 diabetes mellitus and incident dementia, stratified by sex

	UK Biobank				All of Us			
	Model1		Model2		Model1		Model2	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Females	2.05	1.81-2.31	1.95	1.71-2.22	2.45	2.06-2.93	2.45	2.03-2.95
Males	1.97	1.79-2.17	1.95	1.76-2.17	2.32	1.90-2.83	2.42	1.96-2.98

Table 1.S3. Association between healthcare utilization one year prior to baseline and incident dementia

	UK Biobank		All of Us	
	HR	95% CI	HR	95% CI
Healthcare utilization[a]	1.03	1.02-1.03	1.02	1.01-1.02

[a] Healthcare utilization was coded linearly.

Note: Healthcare utilization was measured by the number of encounters with the healthcare system recorded in the electronic health records. Utilization is recorded differently in the two cohorts. In UK Biobank, we tallied the total number of general practitioner and inpatient encounters. In All of Us, we counted all-cause patient encounters, including primary care records, inpatient records, procedures, pre-procedural examinations, regular medical examinations, follow-up examinations, and screenings.

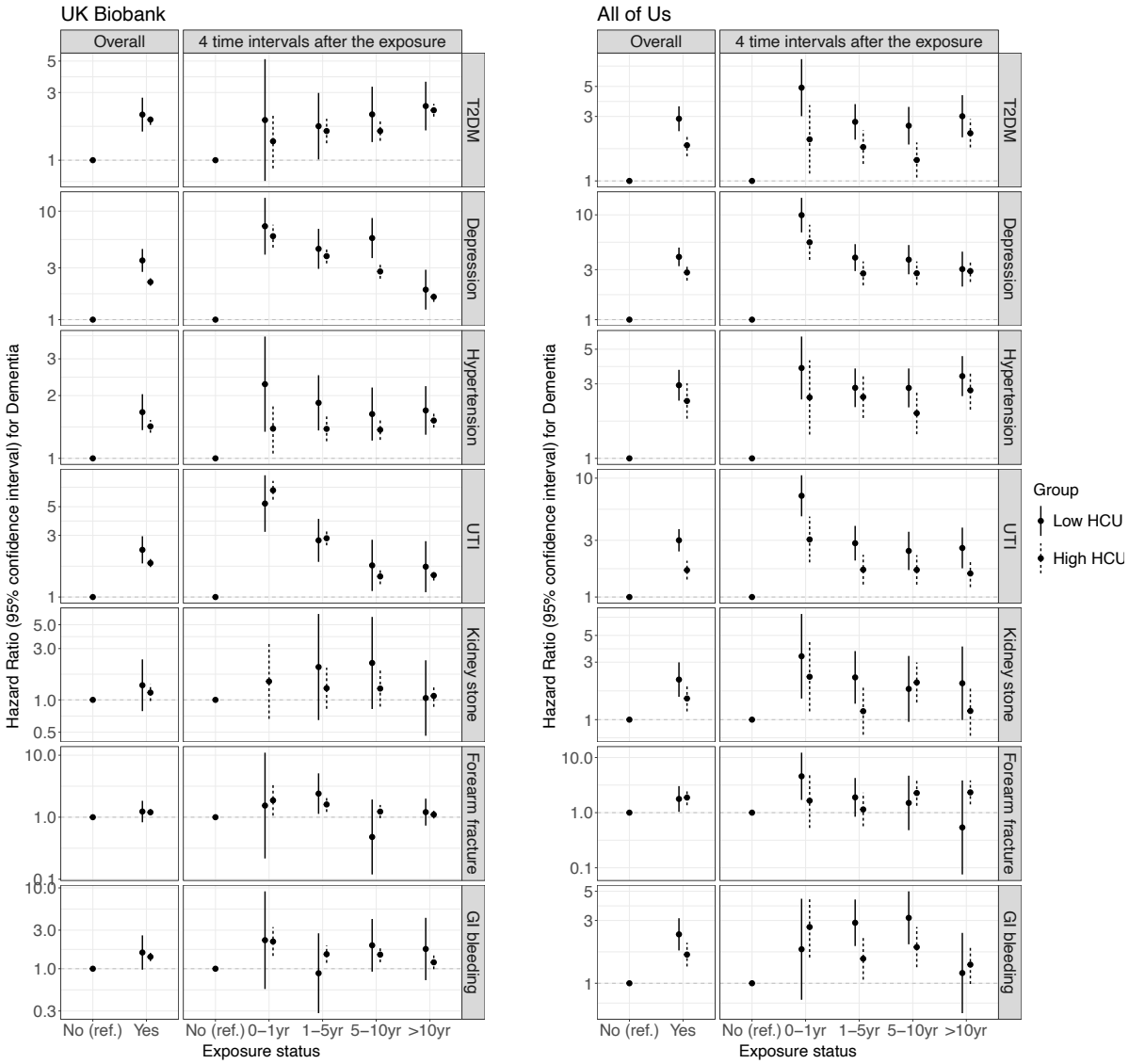


Figure 1.S1. Associations of exposures with dementia incidence, stratified by levels of healthcare utilization

Abbreviations: T2DM, type 2 diabetes mellitus; UTI, urinary tract infection; GI bleeding, gastrointestinal bleeding.
Note: All models were adjusted for age, sex, race, the number of APOE-ε4 alleles, education, smoking history, and BMI.

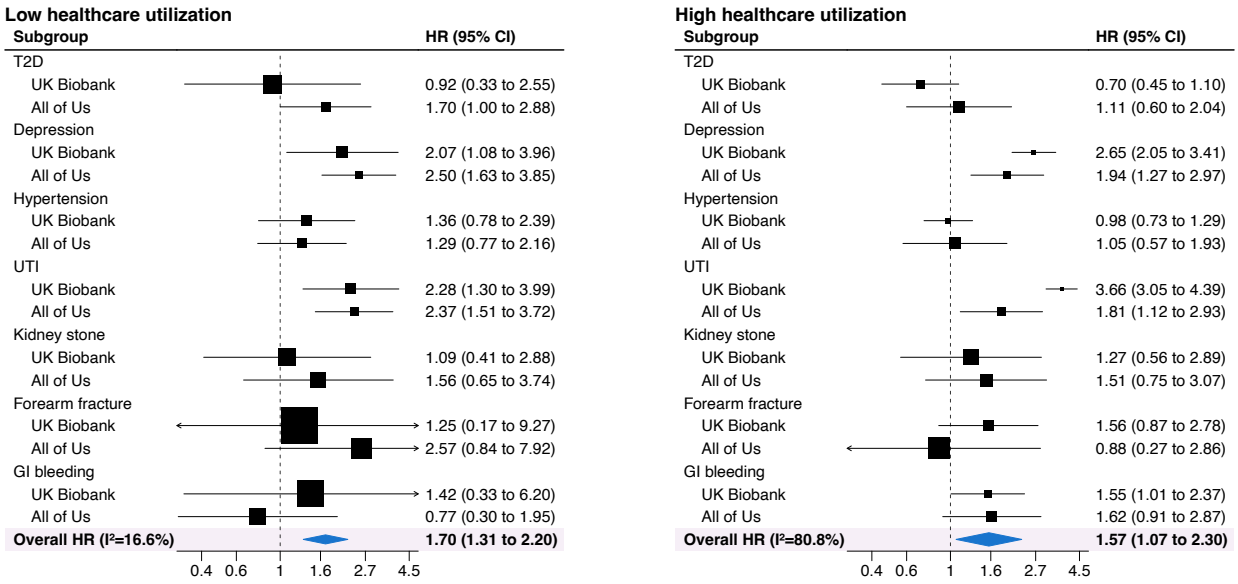


Figure 1.S2. Forest plot of random-effects models for the pooled detection bias estimates, stratified by levels of healthcare utilization

Abbreviations: T2DM, type 2 diabetes mellitus; UTI, urinary tract infection; GI bleeding, gastrointestinal bleeding.
Note: Point estimates for detection bias, which were used as inputs for the meta-analysis, were calculated by dividing the HR 0–1 year after the exposure diagnosis by the overall HR, allowing us to mitigate or cancel out the effects of potential residual confounding, true causal relationships, and other study biases commonly encountered in both estimates. We used a random-effects model with inverse variance weights. Size of squares represent the standard error of the point estimate.

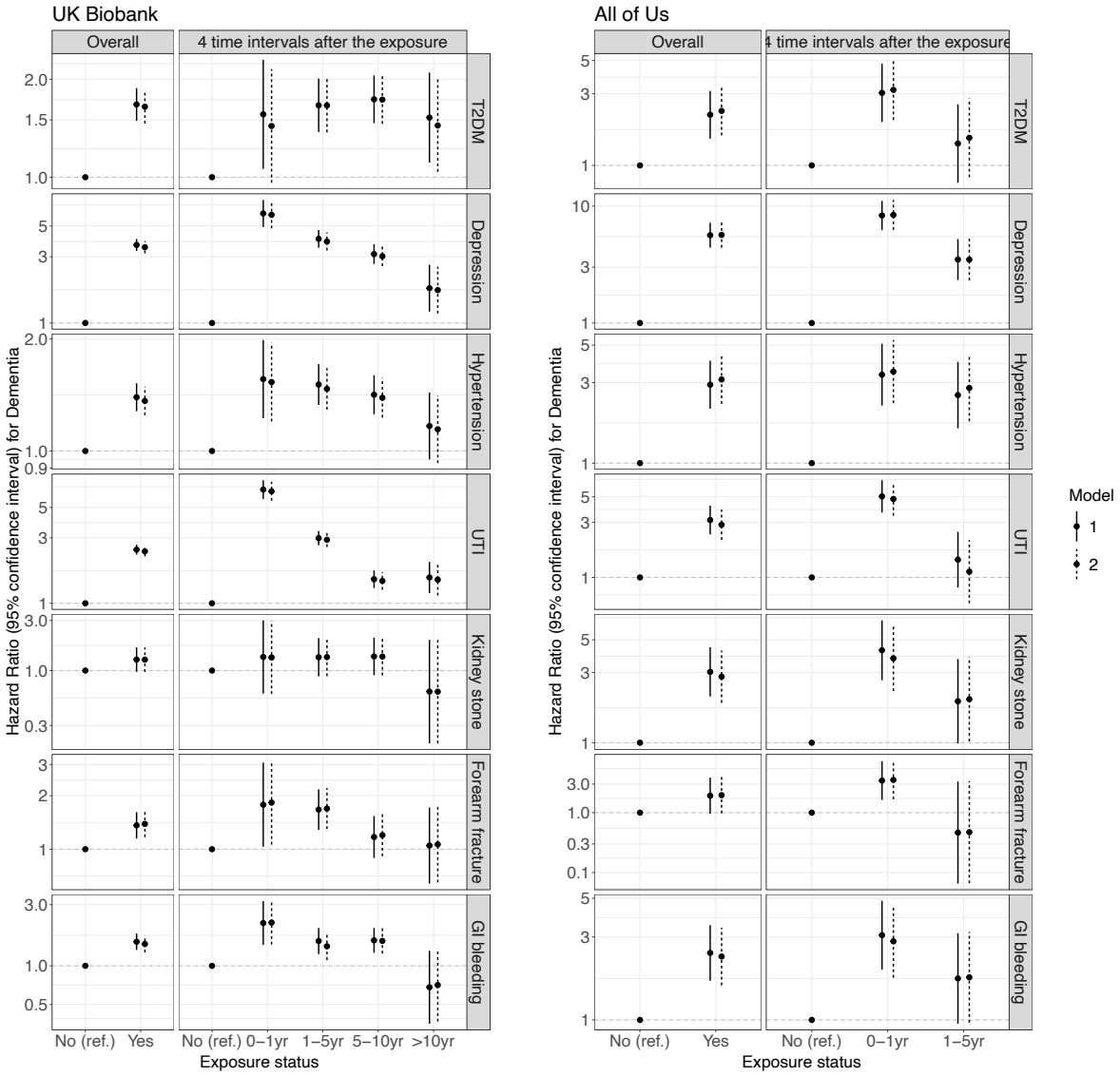


Figure 1.S3. Associations of exposures with dementia incidence, excluding prevalent exposure cases

Abbreviations: T2DM, type 2 diabetes mellitus; UTI, urinary tract infection; GI bleeding, gastrointestinal bleeding.
Note: Model 1 was adjusted for age, sex, race, and the number of APOE-ε4 alleles. Model 2 was further adjusted for education, smoking history, and BMI. In this analysis, we excluded participants with prevalent exposure diagnosis before study baseline.

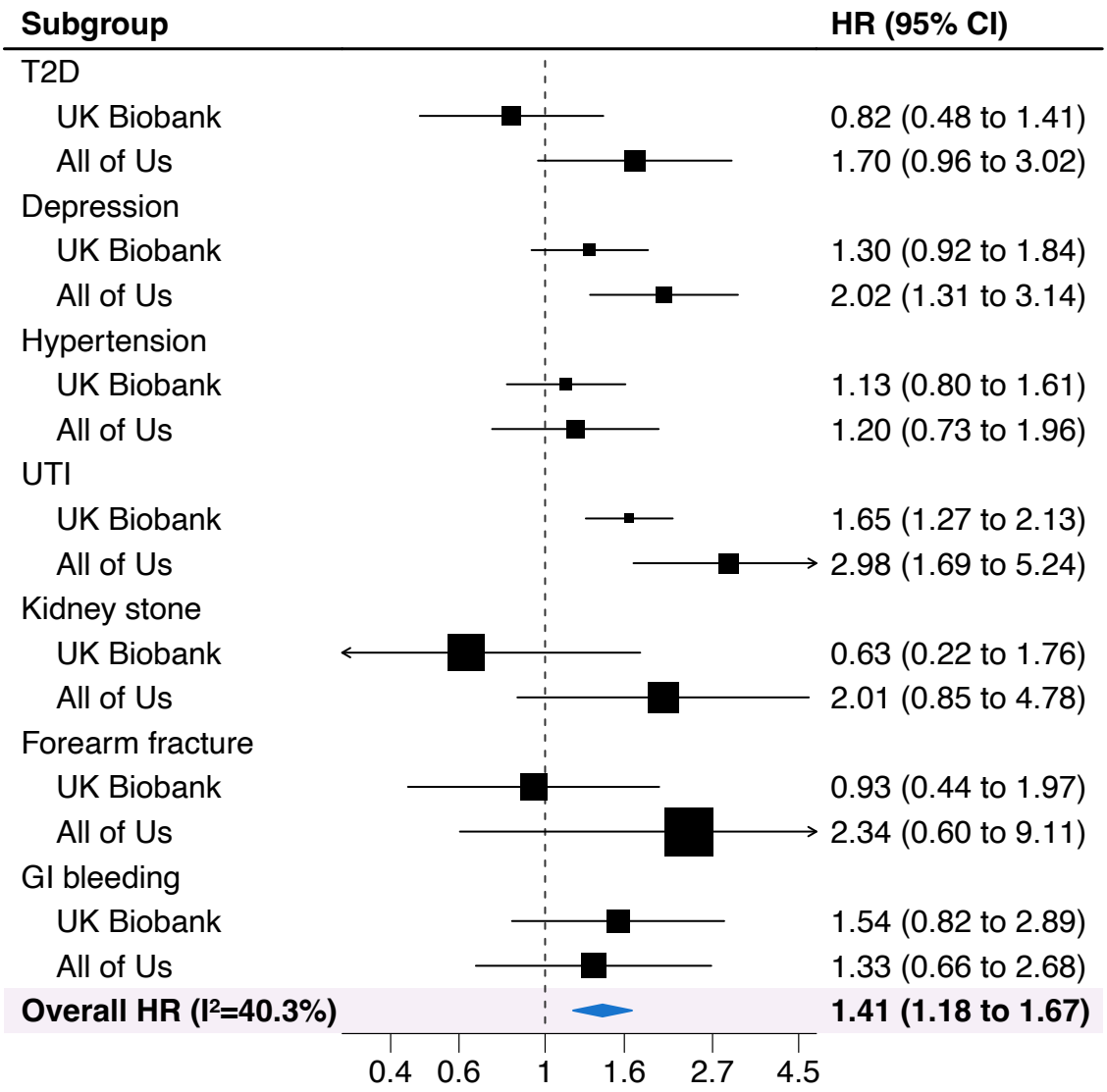


Figure 1.S4. Forest plot of random-effects models for the pooled detection bias estimates, comparing with estimates 1-2 year after exposure diagnosis

Abbreviations: T2DM, type 2 diabetes mellitus; UTI, urinary tract infection; GI bleeding, gastrointestinal bleeding.
Note: Point estimates for detection bias, which were used as inputs for the meta-analysis, were calculated by dividing the HR 0–1 year after the exposure diagnosis by the HR 0–1 year after the exposure diagnosis. We used a random-effects model with inverse variance weights. Size of squares represent the standard error of the point estimate.

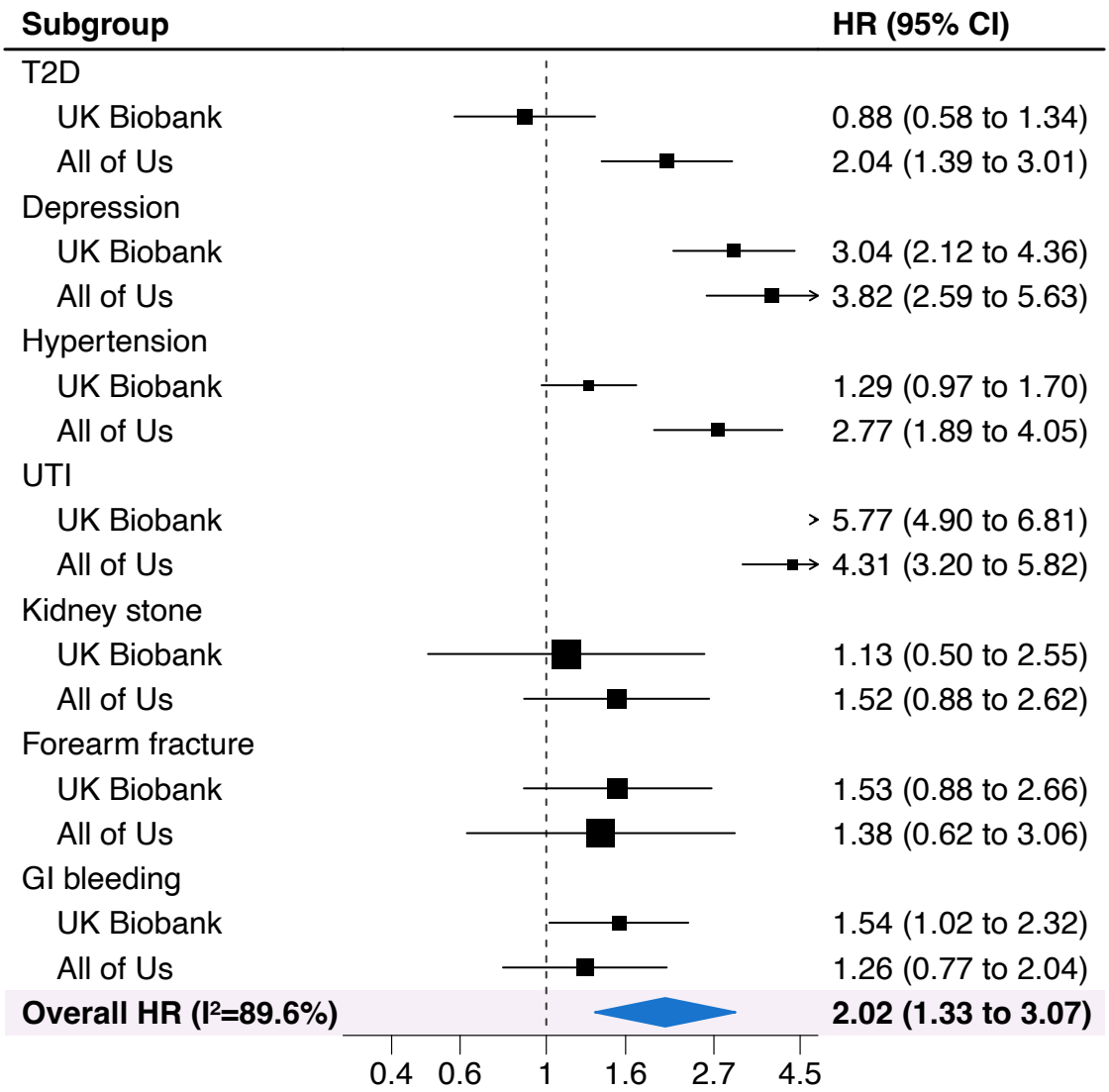


Figure 1.S5. Forest plot of random-effects models for the pooled detection bias estimates, comparing with estimates from meta-analysis of cohort studies

Abbreviations: T2DM, type 2 diabetes mellitus; UTI, urinary tract infection; GI bleeding, gastrointestinal bleeding.
Note: Point estimates for detection bias, which were used as inputs for the meta-analysis, were calculated by dividing the HR 0–1 year after the exposure diagnosis by effect sizes from existing cohort studies or meta-analysis of cohort studies, where available: T2DM,[28] depression,[29] hypertension,[30] and UTI[31]. For kidney stone, forearm fracture, and GI bleeding, there is no existing study, and we divided the HR 0–1 year after the exposure diagnosis by the overall HR in our analysis. Specifically, the comparison HR for T2DM, depression, hypertension, and UTI were 1.62 (95% CI=1.45-1.80), 1.98 (95% CI=1.50-2.63),1.18 (95% CI=1.02-1.35), and 1.13 (95% CI=1.08-1.18), respectively. We used a random-effects model with inverse variance weights. Size of squares represent the standard error of the point estimate.

Chapter 2

Anticipated effects of COVID-19 related brain MRI changes on long-term dementia risk: Linking pre-pandemic brain MRIs to dementia outcomes in UK Biobank

Jingxuan Wang, Erin L. Ferguson, Ruijia Chen, Sarah F. Ackley, Renaud La Joie, John Kornak, J. Daniel Kelly, Hossein Estiri, John Hsu, Deborah Blacker, M. Maria Glymour

Abstract

Background: Neurocognitive sequelae of COVID-19 might increase risk of dementia, but only limited evidence is available to date. This study aimed to investigate this link by examining whether naturally occurring variations in brain structures, i.e., COVID-19 brain signatures, associated with COVID-19 predict incident dementia.

Methods: Participants were enrolled in the UK Biobank from 2006 to 2010, and a subset completed a brain magnetic resonance imaging (MRI) scan. In this prospective cohort study, we included participants aged 55 years or older without prevalent dementia who had a brain MRI. We assessed associations of 31 brain MRI-derived variables previously linked to COVID-19 with incident all-cause dementia using Cox proportional-hazards models adjusted for demographic, genetic, behavioral, and clinical factors. We additionally determined the extent to which the associations were independent of general brain aging and Alzheimer's disease (AD)-related atrophy.

Results: Among 37,426 participants (mean age, 66.5 [SD, 6.3] years; 51.9% female) followed for a mean of 4.7 years (SD, 1.5 years), 112 incident dementia cases were identified. Of the 31 brain MRI-derived variables previously linked to COVID-19, 29 were associated with higher dementia incidence. Brain MRI changes of the magnitude induced on average by COVID-19 were associated with a 9% increase in dementia hazard (HR=1.09, 95% CI=1.06-1.12). Additional adjustment for cortical thinning associated with Alzheimer's disease attenuated the associations, but 14 out of the 31 remained significantly associated with incident dementia.

Conclusions: Naturally occurring variations in COVID-19 brain signatures linked with COVID-19 were associated with higher dementia risk, suggesting a neurobiological basis for a connection between COVID-19 and dementia. These signatures partially overlapped with brain regions considered hallmarks of Alzheimer's disease. Future studies should evaluate the consistency of the COVID-19 brain signatures and associated dementia risk in different samples, effect modifiers such as disease severity or vaccination status, and potential study biases.

Introduction

Long-lasting cognitive sequelae appear common among individuals who have survived coronavirus disease 2019 (COVID-19).[51,52] Neurocognitive consequences of infection may reflect direct damage of infection in the central nervous system (CNS) or hypoxic, inflammatory, immune-related, or vascular mechanisms.[53,54] In addition, neurological manifestations, including anosmia, encephalopathy, cognitive difficulties, insomnia, and stroke, have been reported in COVID-19 patients, raising concerns about shared etiologies with neurodegenerative diseases that cause dementia.[55,56] Understanding neurobiological changes induced by COVID-19 and their relationship to dementia risk is essential for guiding future preventive measures and therapeutic interventions.

However, it is methodologically challenging to assess the link between COVID-19 and dementia. First, COVID-19 cases are underreported.[57,58] Limited testing capabilities during the early stages of the pandemic and the frequency of asymptomatic infections, particularly during the Omicron period, led to underdiagnosis. Widespread exposure to COVID-19 further complicates finding a suitable comparison group who have not been infected. In many populations it is likely that the vast majority of individuals have now been infected.[59] As of November 2024, more than 776 million cases of COVID-19 have been recorded globally, but in many populations it is likely that the vast majority of individuals have been infected.[59] Furthermore, dementia is a condition that typically evolves over several years or even decades.[2] In contrast, COVID-19 emerged only recently. Directly investigating whether COVID-19 increases dementia risk would involve following infected and uninfected individuals for years. Because of the urgency of the

question, alternative approaches are needed to evaluate the hypothesis that COVID-19 affects dementia risk.

Short-term neurologic hallmarks of COVID-19 have been established with brain magnetic resonance imaging (MRI) and naturally occurring variations in these hallmarks were also measured in MRIs that were collected prior to the pandemic.[13] Evaluating whether these hallmarks influence long-term dementia risk provides an indirect route to anticipating the long-term effect of COVID-19 on dementia. The UK Biobank Study collected brain MRIs starting in 2014 and invited eligible participants for second MRIs starting in 2019 (see more details in the Supplement). Over 2,500 brain imaging-derived phenotypes, combining numerous voxels and/or images, are available for each MRI.[60] In 2021, the UK Biobank launched a COVID-19 imaging study to obtain follow-up brain MRIs from individuals who had completed an MRI prior to the pandemic, with an aim of comparing MRI changes in individuals who had COVID-19 or were infected with SARS-CoV-2 to those who had not been infected. Douaud et al.[11] examined brain changes in 785 UK Biobank participants who underwent MRI twice, including 401 individuals who had medical records of COVID-19 or tested positive for SARS-CoV-2 infection based on their primary care or hospital records, diagnostic antigen tests, and antibody tests, as well as 384 demographically matched controls with neither a COVID-19 diagnosis nor a positive test result between scans. Douaud et al. generated an additional set of MRI-based phenotypes using quantitative susceptibility mapping (QSM).[11] They reported a history of infection was statistically significantly associated with changes in 68 out of 297 imaging-derived phenotypes. Although the study used both COVID-19 and infection

history as proxy, these brain changes likely represented a direct consequence of COVID-19 illness. Specifically, comparing to individuals without a history of COVID-19 or infection, those with a history showed evidence of a greater reduction over time in cortical thickness, volume, and gray-white matter contrast in brain regions associated with smell, memory, and emotional processing compared to matched controls with no history of infection.[11]

To help understand the potential consequences of COVID-19 on dementia risk, we utilized the findings from Douaud et al. which identified structural brain changes associated with COVID-19.[11] We refer to the MRI regions identified by Douaud et al. as significantly associated with COVID-19 infection as COVID-19 brain signatures (COVIDsig). If these imaging phenotypes are linked to an increased risk of dementia, COVID-19 could potentially increase future dementia risk through structural brain changes in these regions. Our study aimed to evaluate whether naturally occurring variations in these imaging phenotypes prior to the pandemic were associated with incident dementia (**Figure 2.1**). We hypothesized that variations in COVIDsig—even when observed in a pre-pandemic population—would be associated with higher incidence of subsequent dementia. Further, we determined the extent to which the associations were independent of general brain aging and Alzheimer's disease (AD)-related atrophy.

Methods

Study population

The UK Biobank is a prospective cohort study comprising 502,490 adults who were aged 40-69 years when enrolled at assessment centers throughout the United Kingdom from 2006 to 2010. The sample for the MRI substudy was selected based on their proximity to the imaging center. We included participants who were 55 years or older and had no prior dementia diagnosis at the time of the MRI scan. We only considered the first MRI (n=37,426) for our primary analyses. 99.2% MRIs were completed prior to March 2020 (n=37,108), thus predominantly reflect naturally occurring variation of brain structures, not changes due to COVID-19.

Brain MRI data and COVID-19 signatures (COVIDsig)

All MRI scans were conducted using a standardized Siemens Skyra 3T scanner equipped with a standard Siemens 32-channel head coil. Detailed information regarding image acquisition, processing, and quality control can be found in the UK Biobank Brain Imaging Documentation and protocol publications.[61] In brief, the protocol includes diffusion MRI, resting-state, and task-based functional MRI, and three structural MRI scans, including T1, T2 with fluid attenuation inversion recovery, and susceptibility-weighted MRI. T1-weighted anatomic images were acquired using a three-dimensional magnetization prepared for rapid gradient-echo (3D MPRAGE) sequence at a resolution of 1 × 1 × 1 mm. To define COVIDsig, we exclusively considered imaging-derived phenotypes that showed a statistically significant longitudinal change caused by COVID-19 infection in Douaud et al,[11] and were extracted through the automated

processing pipeline utilized by the UK Biobank. Thus, our COVIDsig comprised 31 separate imaging-derived phenotypes, covering cortical thickness and volumetric measures in orbitofrontal cortex, parahippocampal gyrus, entorhinal cortex, and other regions. Detailed information about the phenotypes is shown in **Table 2.S1**.

Ascertainment of dementia cases

Our primary outcome was incident all-cause dementia diagnosis, including AD, vascular dementia, Lewy body dementia, alcohol-related dementia, and other types of dementia. The outcome was ascertained from coded diagnoses based on medical records linked to National Health Services hospital inpatient, primary care, and death registry data. Detailed information on obtaining dementia diagnosis was described in prior publications.[62] To avoid ambiguity in the diagnosis of dementia type, we used all-cause dementia as the primary outcome. The latest date of dementia diagnosis in our analytical sample is October 20, 2022.

Assessment of covariates

We considered three sets of covariates. Model 1 included age, sex (female, male) and race (White, Black, Asian, Other). In Model 2, we additionally adjusted for total brain volume, an indicator of global brain atrophy, to assess whether the COVIDsig has effects over and above general brain aging. Model 3 additionally controlled for potential confounders, including the indicator for number of APOE-ε4 alleles (0, 1, 2), indicator for assessment center (Bristol, Cheadle, Newcastle, Reading), education (less than A-levels, A-levels or above), and quintiles of Townsend deprivation index (a continuous

measure of area-level material deprivation). All models with volumetric measures as the exposure adjusted for intracranial volume to correct for differences in head size.[63] Detailed information about the covariate is shown in the supplement.

Statistical analyses

All exposures—including 31 imaging-derived phenotypes—were scaled based on the effect of COVID-19 on these exposures as estimated in Douaud et al,[11] such that each unit increase in the scaled value represent its corresponding change expected after COVID-19. First, we used Cox proportional-hazards models to assess the association between each imaging-derived phenotype in COVIDsig and an incident dementia diagnosis. In our primary analysis, participants were followed from time of MRI scan to the date of first all-cause dementia diagnosis, death from any cause, or administrative censoring (date of the last dementia diagnosis in the dataset), whichever came first.

To better determine the extent to which the associations were independent of general brain aging, we performed a permutation test and tested the hypothesis that a greater number of imaging-derived phenotypes in COVIDsig exhibit a significant association with dementia risk than what would be expected among randomly selected 31 imaging-derived phenotypes from all available phenotypes. To do so, we began by randomly choosing 31 phenotypes from a pool of 889 available phenotypes in the UK Biobank, which are measurements related to regional volume, gray-white matter contrast, and cortical thickness. Subsequently, we fit separate Cox models for incident dementia to each of these 31 individual imaging-derived phenotypes and tallied the number of

statistically significant associations. We used 5,000 permutations to estimate the distribution of the total number of significant associations among 31 randomly selected imaging-derived phenotypes and compared it with the number of significant associations observed for phenotypes in COVIDsig.

In addition, we calculated Spearman-rank correlation between each phenotype in COVIDsig and the cortical thickness of AD signature region (ADsig), a composite signature of AD cortical thinning, to better understand the relationship between structural brain changes due to COVID-19 and AD-related atrophy. We then extended our analysis by fitting additional Cox models based on Models 1-3 and further adjusted for the cortical thickness of ADsig to evaluate the additional effect of COVIDsig independent of atrophy associated with AD risk. The cortical thickness of ADsig was calculated by the surface area-weighted average of cortical thicknesses across entorhinal cortex, fusiform, inferior, and middle temporal regions.[64]

Lastly, we estimated the magnitude of impact of COVID-19 on dementia via structural brain changes in COVIDsig using principal component analysis (PCA). We performed a PCA on the 31 imaging-derived phenotypes, which were highly correlated (**Figure 2.S1**). Subsequently, we selected the top principal components which collectively explained at least 90% of the total variation and then fit a Cox model using these principal components. We calculated the hazard ratio associated with changes in COVIDsig by comparing hazards between two hypothetical individuals: one conditional on pre-COVID average values of all imaging-derived phenotypes and the other

conditional on post-COVID values. The post-COVID scaled variables were obtained by adding 1 to the pre-COVID scaled variables, because a unit increase in the scaled value represents the corresponding expected change due to COVID-19.

We conducted three sensitivity analyses. First, we restricted to pre-pandemic follow up (i.e., ending on March 1, 2020) to exclude the potential impact of pandemic on the observed association. Second, we repeated the primary analysis using the outcome of incident AD. Lastly, we time updated values for COVIDsig phenotypes in Cox models for participants who had repeated MRIs (n=3,748).

Results

Participants

The analytical sample (N= 37,426 individuals) had a mean age of 66.5 years (SD = 6.3 years; **Table 2.1**). During the follow-up period (mean = 4.7 years, SD = 1.5 years), 112 incident dementia cases (51 AD, 18 vascular dementia, and 43 other type) were identified, for an age-adjusted dementia incidence rate of 1.2 per 1,000 person-years.

Associations Between COVIDsig and Incident Dementia Diagnosis

Among the 31 imaging-derived phenotypes within COVIDsig, 29 were significantly associated with dementia incidence (Model 1, **Figure 2.2**). Several phenotypes showing the strongest associations included gray-white contrast in left entorhinal cortex from the Desikan-Killiany (DK) atlas (HR = 1.05, 95% CI = 1.04-1.07), cortical thickness of left perirhinal cortex from Brodmann areas (HR = 1.05, 95% CI = 1.03-1.06), and cortical

thickness of left lateral orbitofrontal cortex from Desikan-Killiany-Tourville (DKT) atlas (HR = 1.04, 95% CI = 1.02-1.07). One phenotype was no longer significantly associated with dementia incidence after further adjustment for total brain volume, and the same occurred after adjustment for potential confounders (Models 2-3, **Figure 2.2**).

Excess Effect on Dementia Beyond General Brain Atrophy

In a random selection of 31 atrophy-related imaging-derived phenotypes, an average of 20.8 were associated with dementia (95% CI = 15.7-26.0; **Table 2.2**), which is significantly lower than the 29 COVIDsig phenotypes found to be associated with dementia incidence. This pattern was consistent across covariate adjustments: an average of 17.3 (95% CI = 11.9-22.7) of 31 randomly selected phenotypes were significantly associated with dementia when adjusting for total brain volume and when adjusting for potential confounders, an average of 17.0 (95% CI = 11.8-22.3) of 31 randomly selected phenotypes were significantly associated with dementia.

No Independent Effect of AD-related Atrophy

Correlation between imaging-derived phenotypes in COVIDsig and ADsig cortical thickness were moderate to high, especially for cortical thickness phenotypes, with Spearman-rank correlation coefficients ranging from 0.32 to 0.56 (**Figure 2.3**). After controlling for cortical thickness of the ADsig in the Cox models, the magnitudes of associations were attenuated for all three covariate adjustment sets. When adjusting for ADsig cortical thickness in addition to sex and race, only 14 of the 31 COVIDsig phenotypes remained statistically significantly associated with incident dementia (Model

1, **Figure 2.3**). For example, the hazard ratio of gray-white contrast in left entorhinal cortex from the DK atlas shrank from 1.05 (95% CI = 1.04-1.07) in sex- and race-adjusted models to 1.03 (95% CI = 1.01-1.05), and that of cortical thickness of left lateral orbitofrontal cortex from DKT atlas reduced from 1.04 (95% CI = 1.02-1.07) to 0.97 (95% CI = 0.94-1.00). In regions that did not overlap with ADsig, the magnitudes of associations also decreased. For example, the hazard ratio of cortical thickness in left lateral orbitofrontal cortex from the DK atlas attenuated from 1.03 (95% CI = 1.01-1.05) to 0.97 (95% CI = 0.95-1.00) in sex- and race-adjusted models.

Magnitude of Impact of COVID-19 on Dementia via Structural Brain Changes

The first 14 components derived from a PCA on the highly correlated COVIDsig phenotypes explained 91.6% of total variation (**Figure 2.S2**). Cox regression with those 14 components showed that the brain MRI changes in COVIDsig of the average magnitude induced by COVID-19 were associated with a 9% increase in dementia hazard (HR = 1.09, 95% CI = 1.06-1.12; **Table 2.3**). The magnitudes of impact on dementia hazard were stable after further adjustments for total brain volume and potential confounders.

Sensitivity Analyses

Results were similar when we limited the follow-up period to prior to the pandemic (**Figure 2.S3**, **Table 2.S2**), or when we focused on AD incidence (**Figure 2.S4**, **Table 2.S3**), or when we time updated our exposures (**Figure 2.S5**). In first two sensitivity

analyses, the number of phenotypes associated with dementia or AD incidence significantly exceeded that of randomly selected phenotypes.

Discussion

In this large sample of UK adults with MRIs completed prior to the COVID-19 pandemic, naturally occurring variation in imaging-derived phenotypes linked to COVID-19 was associated with a small increase in the incidence of dementia diagnosis. This association exceeded what would be expected by chance, and only partially overlapped with known MRI markers of AD, such as the entorhinal cortex. The evidence suggests a link between COVID-19 and small elevations in dementia risk through structural brain changes.

Prior work has also suggested a potential link between COVID-19 infection or illness and dementia risk.[55,65] For example, using Medicare data, Cohen et al. demonstrated an excess risk for dementia observed 4 months after the acute phase of COVID-19.[65] An analysis of 2-year retrospective cohort studies in a federated network for electronic health records suggested that the elevated risk of dementia persisted up to 2 years following a COVID-19 diagnosis.[55]

Timely investigation into the link between COVID-19 and dementia can help guide early interventions after COVID-19 cases, e.g., cognitive stimulation or physical activity to potentially delay cognitive deterioration.[39,66] However, previous studies utilizing cohort designs necessitate an extended follow-up period to achieve precise and robust

estimates. A majority of the existing research relied on a follow-up duration of less than 2 years.[55,65] Our innovative approach, which leverages COVID-19 MRI signatures and pre-pandemic follow-up to assess dementia incidence (**Figure 2.1**), expedites the generation of robust results. Furthermore, prior studies may suffer from exposure misclassification bias when utilizing contemporaneous non-COVID-19 controls or potential outcome measurement bias due to evolving dementia diagnosis criteria and practices when using historical comparison controls. Our approach avoids directly comparing people with and without COVID-19, which may be biased by exposure misclassification due to negative or missing test results with asymptomatic or minimally symptomatic infection, especially in the Omicron period.[57,58]

Our findings only provided an estimation of the effect of COVID-19 on dementia risk through structural brain changes detectable on MRI, rather than the total effect of COVID-19. This is because our analysis employed neuroimaging data in lieu of disease history, focusing solely on potential associations via structural brain changes. Beyond COVID-19 MRI signatures, there may be numerous other infection-induced pathophysiologic mechanisms through which COVID-19 influences dementia risk,[67,68] such as damage due to overactive immune-inflammatory response,[69,70] ischemic injury, cerebrovascular complications, angiotensin-converting enzyme 2 (ACE2) availability in the brain.[53,70–72] SARS-CoV-2 infection has been linked to increased expression of genes related to both innate immune response and AD pathogenesis.[73] Additionally, the impact of COVID-19 on dementia risk may be mediated by behavioral factors altered by illness, including social isolation,[74] physical inactivity,[75] inadequate chronic disease

management, or depression.[76] The change in components of COVIDsig estimated by Douaud et al. ranged from -0.76% to -0.92%,[11] corresponding with a potential 9% increase in dementia hazard. Although a 9% increase may imply a large number of extra cases at a population level, the increase in risk for any individual is relatively small.

Key limitations of this study include reliance on the UK Biobank for both determination of COVID-19 signatures and the assessment of its association with dementia risk. Results should be confirmed in independent samples and the persistence of post-COVID-19 structural brain changes assessed. While Douaud et al. attempted to mitigate study biases by comparing pre- and post-infection MRIs in COVID-19 patients and matched controls,[11] and a subsequent Mendelian randomization study verified their results, some potential for misclassification bias and selection bias remains. We did not consider the heterogeneity in COVID-19 signatures, which could be influenced by demographics and clinical characteristics. Lastly, the dementia incidence rates in the UK Biobank neuroimaging sample are very low, resulting in imprecise effect estimates.

Brain structural changes associated with COVID-19 were linked to a higher dementia risk. Findings suggest a neurobiological mechanism linking COVID-19 to dementia via these structural changes. Future studies should evaluate the persistence of COVID-19 signatures in different samples and assess heterogeneity across features of the disease and vaccination history.

Tables

Table 2.1. Baseline characteristics of participants in the analytical sample

	No. (%) [a]		
Characteristic	No incident dementia (n = 37,314)	Incident dementia (n = 112)	Overall (n = 37,426)
Mean age (SD)	66.5 (6.3)	72.1 (5.2)	66.5 (6.3)
Age groups			
55-64	15,624 (41.9)	11 (9.8)	15,635 (41.8)
65-74	18,097 (48.5)	61 (54.5)	18,158 (48.5)
≥ 75	3,593 (9.6)	40 (35.7)	3,633 (9.7)
Sex			
Female	19,395 (52.0)	46 (41.1)	19,441 (51.9)
Male	17,919 (48.0)	66 (58.9)	17,985 (48.1)
Race			
Asian	466 (1.2)	0 (0)	466 (1.2)
Black	242 (0.6)	1 (0.9)	243 (0.6)
Other	322 (0.9)	0 (0)	322 (0.9)
White	36,272 (97.2)	111 (99.1)	36,383 (97.2)
Missing	12 (0.0)	0 (0)	12 (0.0)
Number of APOE-ε4 alleles			
0	26,356 (70.6)	47 (42.0)	26,403 (70.5)
1	9,187 (24.6)	50 (44.6)	9,237 (24.7)
2	797 (2.1)	11 (9.8)	808 (2.2)
Missing	974 (2.6)	4 (3.6)	978 (2.6)
Assessment center			
Bristol	56 (0.2)	0 (0)	56 (0.1)

	No. (%) [a]		
Characteristic	No incident dementia (n = 37,314)	Incident dementia (n = 112)	Overall (n = 37,426)
Cheadle	21,699 (58.2)	88 (78.6)	21,787 (58.2)
Newcastle	9,814 (26.3)	16 (14.3)	9,830 (26.3)
Reading	5,745 (15.4)	8 (7.1)	5,753 (15.4)
Education			
A-levels or above	29,237 (78.4)	75 (67.0)	29,312 (78.3)
Less than A-levels	8,077 (21.6)	37 (33.0)	8,114 (21.7)
Townsend deprivation index[b]			
1 (least deprived)	9,144 (24.5)	28 (25.0)	9,172 (24.5)
2-4	23,223 (62.2)	71 (63.4)	23,294 (62.2)
5 (most deprived)	4,915 (13.2)	13 (11.6)	4,928 (13.2)
Missing	32 (0.1)	0 (0)	32 (0.1)

Abbreviations: SD, standard deviation.
[a] Percentages may not sum to 100 because of rounding.
[b] Index quintiles, combining social class, employment, car availability, and housing.

Table 2.2. Number of significant associations among the 31 imaging-derived phenotypes of COVID-19 signatures and among randomly selected 31 imaging-derived phenotypes

Model Specification	Number of significant associations in COVIDsig	Number of significant associations in random phenotypes (95% CI)
Model 1[a]	29	20.8 (15.7-26.0)
Model 2[b]	29	17.3 (11.9-22.7)
Model 3[c]	28	17.0 (11.8-22.3)

Abbreviations: COVIDsig, COVID-19 Signature; CI, confidence interval.
[a] Adjusted for sex and race.
[b] Further adjusted for total brain volume.
[c] Further adjusted for the number of APOE-ε4 alleles, assessment center, education, and Townsend deprivation index.

Table 2.3. Hazard ratios associated with changes in top 14 principal components of the COVID-19 signature

	Model 1[a]	Model 2[b]	Model 3[c]
Parameter	HR (95% CI)	HR (95% CI)	HR (95% CI)
Joint effect of 14 PCs			
PCs1-14	1.10 (1.07-1.14)	1.10 (1.06-1.13)	1.09 (1.06-1.12)
Individual component models[d]			
PC1	1.20 (1.14-1.26)	1.18 (1.12-1.23)	1.16 (1.10-1.22)
PC2	1.03 (0.94-1.13)	1.02 (0.93-1.12)	1.01 (0.93-1.11)
PC3	1.06 (0.91-1.24)	0.91 (0.77-1.07)	0.94 (0.80-1.10)
PC4	0.84 (0.75-0.93)	0.84 (0.75-0.93)	0.82 (0.74-0.92)
PC5	0.91 (0.80-1.05)	0.93 (0.81-1.07)	0.94 (0.82-1.08)
PC6	1.07 (0.91-1.25)	1.06 (0.91-1.24)	1.02 (0.87-1.20)
PC7	1.09 (0.92-1.28)	1.03 (0.87-1.22)	1.01 (0.86-1.20)
PC8	1.02 (0.84-1.24)	1.03 (0.84-1.26)	1.04 (0.86-1.27)
PC9	0.91 (0.75-1.10)	0.86 (0.71-1.05)	0.87 (0.72-1.07)
PC10	0.90 (0.73-1.11)	0.93 (0.76-1.14)	0.94 (0.77-1.16)
PC11	0.84 (0.67-1.05)	0.89 (0.71-1.11)	0.89 (0.71-1.13)
PC12	1.59 (1.26-2.01)	1.53 (1.21-1.93)	1.42 (1.12-1.81)
PC13	0.66 (0.52-0.83)	0.64 (0.51-0.81)	0.63 (0.50-0.81)
PC14	0.50 (0.39-0.64)	0.51 (0.40-0.66)	0.55 (0.43-0.70)
Mutually adjusted model[e]			
PC1	1.22 (1.16-1.29)	1.20 (1.13-1.27)	1.18 (1.12-1.25)
PC2	1.02 (0.94-1.11)	1.01 (0.93-1.10)	1.01 (0.93-1.10)
PC3	1.35 (1.14-1.60)	1.24 (1.02-1.50)	1.28 (1.05-1.56)
PC4	0.94 (0.84-1.05)	0.93 (0.83-1.04)	0.90 (0.81-1.02)
PC5	0.88 (0.77-1.01)	0.91 (0.79-1.04)	0.91 (0.79-1.06)
PC6	1.16 (1.00-1.35)	1.15 (0.98-1.34)	1.12 (0.95-1.32)
PC7	1.19 (1.01-1.41)	1.15 (0.97-1.37)	1.14 (0.96-1.36)
PC8	0.99 (0.82-1.21)	1.01 (0.83-1.24)	1.02 (0.83-1.26)
PC9	0.95 (0.79-1.15)	0.94 (0.78-1.13)	0.95 (0.78-1.15)
PC10	0.90 (0.74-1.10)	0.92 (0.76-1.11)	0.93 (0.76-1.14)
PC11	0.81 (0.64-1.01)	0.83 (0.66-1.04)	0.86 (0.68-1.09)
PC12	1.71 (1.36-2.14)	1.64 (1.31-2.06)	1.54 (1.22-1.94)
PC13	0.72 (0.57-0.90)	0.72 (0.57-0.90)	0.72 (0.57-0.91)
PC14	0.57 (0.45-0.73)	0.58 (0.46-0.74)	0.59 (0.46-0.75)

Abbreviations: HR, hazard ratio; CI, confidence interval; PC, principal component.

[a] Adjusted for sex and race.
[b] Further adjusted for total brain volume.
[c] Further adjusted for the number of APOE-ε4 alleles, assessment center, education, and Townsend deprivation index.
[d] Cox models with each individual PC as the predictor.
[e] Cox models with all 14 PCs as predictors.

Figures

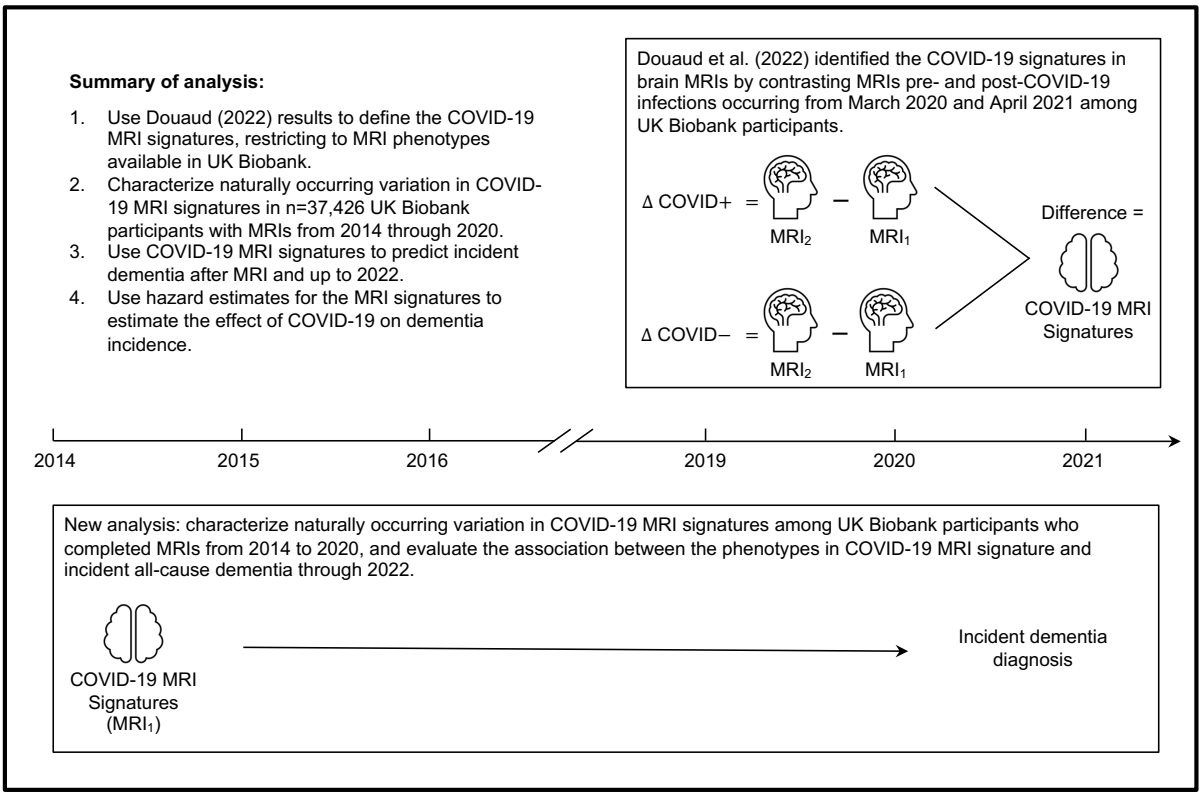


Figure 2.1. Conceptual framework of the analysis

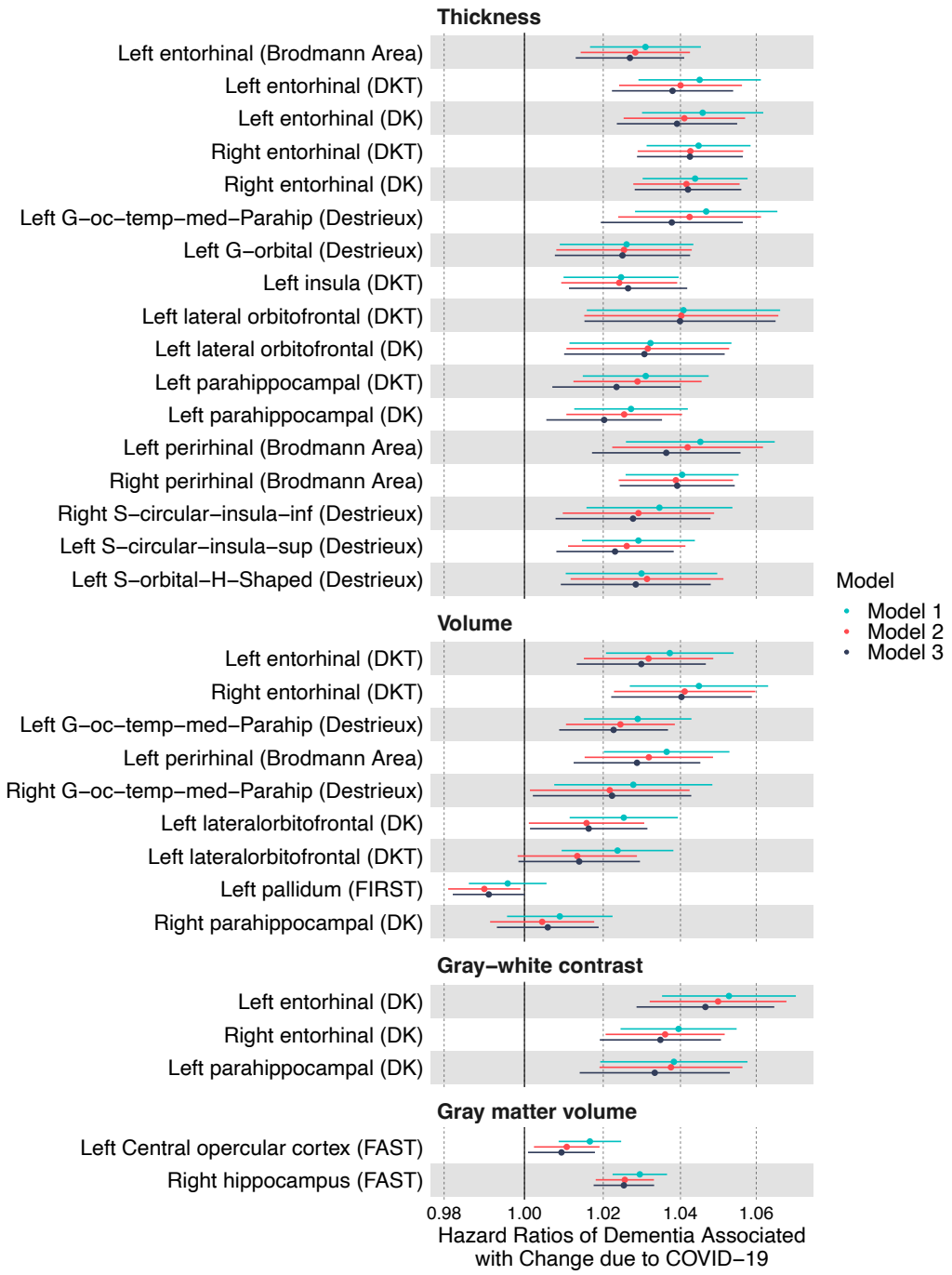


Figure 2.2. Associations between individual imaging-derived phenotypes of COVID-19 signatures and incident dementia

Abbreviations: DK, Desikan-Killiany atlas; DKT, Desikan-Killiany-Tourville atlas.
Note: Model 1 adjusted for sex and race. Model 2 further adjusted for total brain volume. Model 3 further adjusted for the number of APOE-ε4 alleles, assessment center, education, and Townsend deprivation index. All models with volumetric measures were adjusted for intracranial volume.

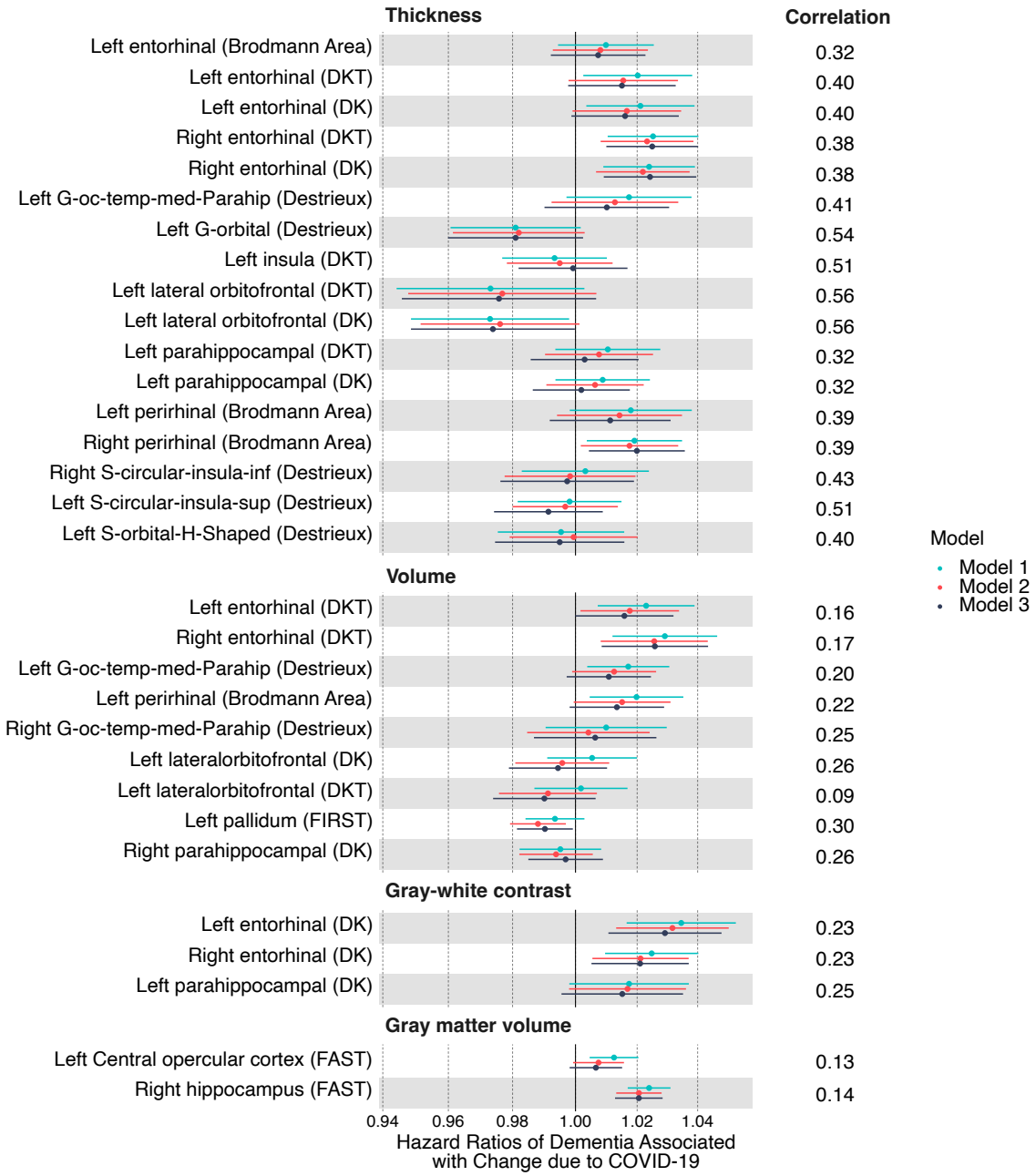


Figure 2.3. Associations between individual imaging-derived phenotypes of COVID-19 signatures and incident dementia adjusting for cortical thickness of the AD signature region and correlations between individual phenotypes and cortical thickness of the AD signature region

Abbreviations: AD, Alzheimer's disease; DK, Desikan-Killiany atlas; DKT, Desikan-Killiany-Tourville atlas.
Note: Model 1 adjusted for sex, race, and cortical thickness of AD signature region. Model 2 further adjusted for total brain volume. Model 3 further adjusted for the number of APOE-ε4 alleles, assessment center, education, and Townsend deprivation index. All models with volumetric measures were adjusted for intracranial volume.

Supplement

UK Biobank Imaging Study

In 2014, the UK Biobank initiated the imaging study, with the objective of inviting 100,000 participants who already participated in the UK Biobank study for subsequent brain, cardiac, and abdominal magnetic resonance imaging (MRI) scans. Eligible participants included all surviving UK Biobank participants who still resided in the UK and did not withdraw from the study. The initiation of invitations based on geographic locations: Invitations for the UK's Central region were first sent out in 2014, with the Northern region following in 2017, the South-East region in 2018, and the South-West center in 2020. Invitations were sent by email and then augmented with mail since 2020.

Starting in 2019, invitations for repeat MRIs were initiated for both the Central and Northern regions. However, this process was temporarily paused in 2020 due to the pandemic, during which around 3,000 repeat MRIs were already collected. Subsequently, the UK Biobank recommenced the invitation for repeat MRIs in February 2021, with an additional goal of exploring the potential impact of SARS-CoV-2 infection on internal organs.

The eligibility criteria included:

- Had already attended an imaging assessment.
- Had high-quality scans from the first imaging visit.

- Had no incidental findings identified from their scans taken at the first imaging visit.
- Still lived within the catchment area of the clinic they attended for their first imaging assessment.
- Had not withdrawn or died.
- Had a valid email and postal address.
- Lived within 60 km of the clinic (extended to 75 km in Feb 2021 and removed in May 2021), due to travel restrictions during the lockdown period.

Covariate data collection

At the time of recruitment, data on age and sex were obtained from a central registry and subsequently updated by participants. Race and education information were self-reported by participants through a touchscreen questionnaire during the baseline assessment. To assess area socioeconomic status, the Townsend deprivation index, a composite score based on preceding national census output areas, was calculated using participants' postcodes at recruitment. APOE-ε4 alleles were determined based on the single nucleotide polymorphisms rs7412 and rs429358.

Table 2.S1. Detailed information about 31 imaging-derived phenotypes of the COVID-19 signature (COVIDsig)

Imaging-derived phenotypes	UK Biobank variable ID	Estimated percentage change due to COVID-19[a]
Left entorhinal gray-white contrast (DK)	26994	-1.16%
Right entorhinal gray-white contrast (DK)	27029	-1.00%
Left parahippocampal gray-white contrast (DK)	27004	-0.92%
Left entorhinal thickness (Brodmann Area)	27086	-0.82%
Left entorhinal thickness (DKT)	27177	-1.10%
Left entorhinal thickness (DK)	26760	-1.07%
Right entorhinal thickness (DKT)	27270	-0.96%
Right entorhinal thickness (DK)	26861	-0.96%
Left G-oc-temp-med-Parahip thickness (Destrieux)	27425	-1.07%
Left G-orbital thickness (Destrieux)	27426	-0.61%
Left insula thickness (DKT)	27204	-0.48%
Left lateral orbitofrontal thickness (DKT)	27183	-0.76%
Left lateral orbitofrontal thickness (DK)	26766	-0.62%
Left parahippocampal thickness (DKT)	27187	-1.01%
Left parahippocampal thickness (DK)	26770	-0.92%
Left perirhinal thickness (Brodmann Area)	27085	-1.37%
Right perirhinal thickness (Brodmann Area)	27127	-1.00%
Right S-circular-insula-inf thickness (Destrieux)	27672	-0.85%
Left S-circular-insula-sup thickness (Destrieux)	27451	-0.46%
Left S-orbital-H-Shaped thickness (Destrieux)	27466	-0.71%
Left entorhinal volume (DKT)	27208	-1.57%
Right entorhinal volume (DKT)	27301	-1.73%

Imaging-derived phenotypes	UK Biobank variable ID	Estimated percentage change due to COVID-19[a]
Left G-oc-temp-med-Parahip volume (Destrieux)	27499	-1.30%
Right G-oc-temp-med-Parahip volume (Destrieux)	27721	-1.80%
Left lateralorbitofrontal volume (DK)	26799	-0.61%
Left lateralorbitofrontal volume (DKT)	27214	-0.62%
Left pallidum volume (FIRST)	25017	-0.76%
Right parahippocampal volume (DK)	26904	-0.96%
Left perirhinal volume (Brodmann Area)	27099	-1.52%
Left Central opercular cortex gray matter volume (FAST)	25864	-0.47%
Right hippocampus gray matter volume (FAST)	25887	-0.32%

Abbreviations: DK, Desikan-Killiany atlas; DKT, Desikan-Killiany-Tourville atlas.
[a] Estimated in Douaud et al[11].

Table 2.S2. Number of significant associations with dementia among the 31 imaging-derived phenotypes of the COVID-19 signature and among randomly selected 31 imaging-derived phenotypes: censored by onset of the pandemic

Model Specification	Number of significant associations in COVIDsig	Number of significant associations in random phenotypes (95% CI)
Model 1[a]	29	20.3 (15.2-25.3)
Model 2[b]	28	16.0 (10.7-21.3)
Model 3[c]	27	16.0 (10.5-21.4)

Abbreviations: COVIDsig, COVID-19 Signature; CI, confidence interval.
[a] Adjusted for sex and race.
[b] Further adjusted for total brain volume.
[c] Further adjusted for the number of APOE-ε4 alleles, assessment center, education, and Townsend deprivation index.

Table 2.S3. Number of significant associations with Alzheimer's disease among the 31 imaging-derived phenotypes of the COVID-19 signature and among randomly selected 31 imaging-derived phenotypes

Model Specification	Number of significant associations in COVIDsig	Number of significant associations in random phenotypes (95% CI)
Model 1[a]	29	20.0 (14.8-25.1)
Model 2[b]	28	16.5 (11.3-21.8)
Model 3[c]	28	15.8 (10.4-21.2)

Abbreviations: COVIDsig, COVID-19 Signature; CI, confidence interval.
[a] Adjusted for sex and race.
[b] Further adjusted for total brain volume.
[c] Further adjusted for the number of APOE-ε4 alleles, assessment center, education, and Townsend deprivation index.

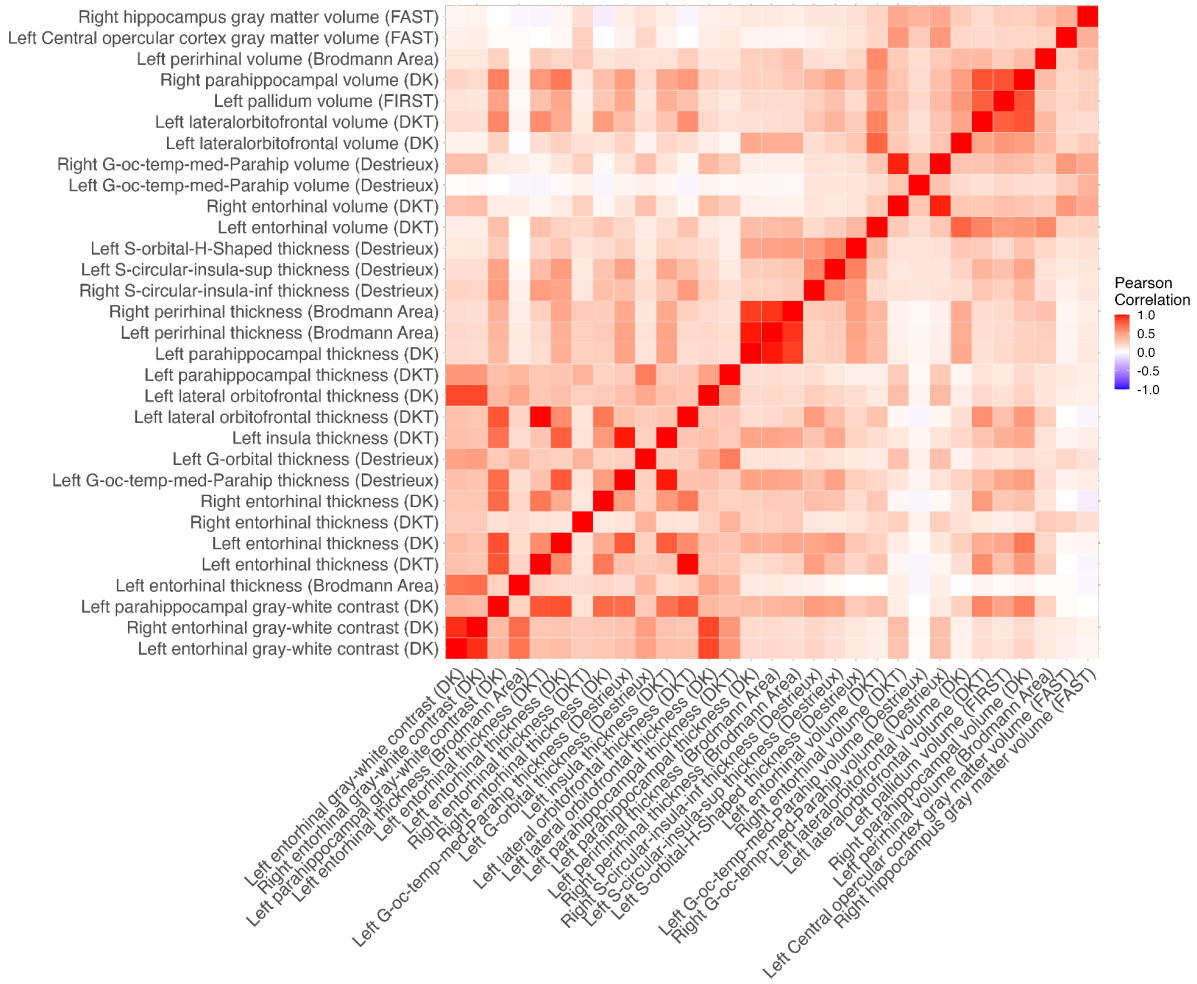


Figure 2.S1. Correlation structure of imaging-derived phenotypes of the COVID-19 signature

Figure 2.S2. Variance explained by principal components and loadings

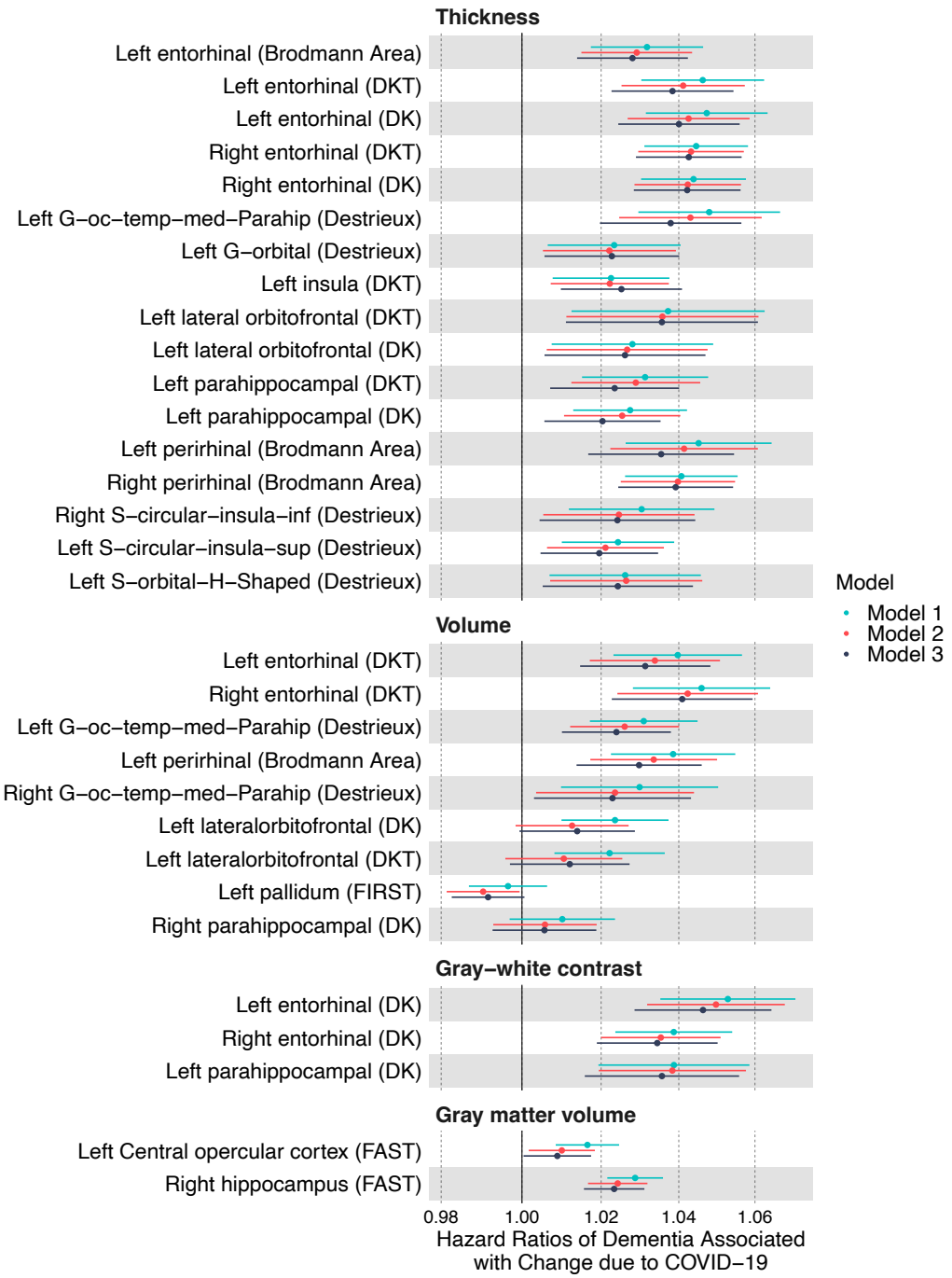


Figure 2.S3. Associations between individual imaging-derived phenotypes of the COVID-19 signature and incident dementia: censored by onset of the pandemic

Abbreviations: DK, Desikan-Killiany atlas; DKT, Desikan-Killiany-Tourville atlas.
Note: Model 1 adjusted for sex and race. Model2 further adjusted for total brain volume. Model 3 further adjusted for the number of APOE-ε4 alleles, assessment center, education, and Townsend deprivation index.

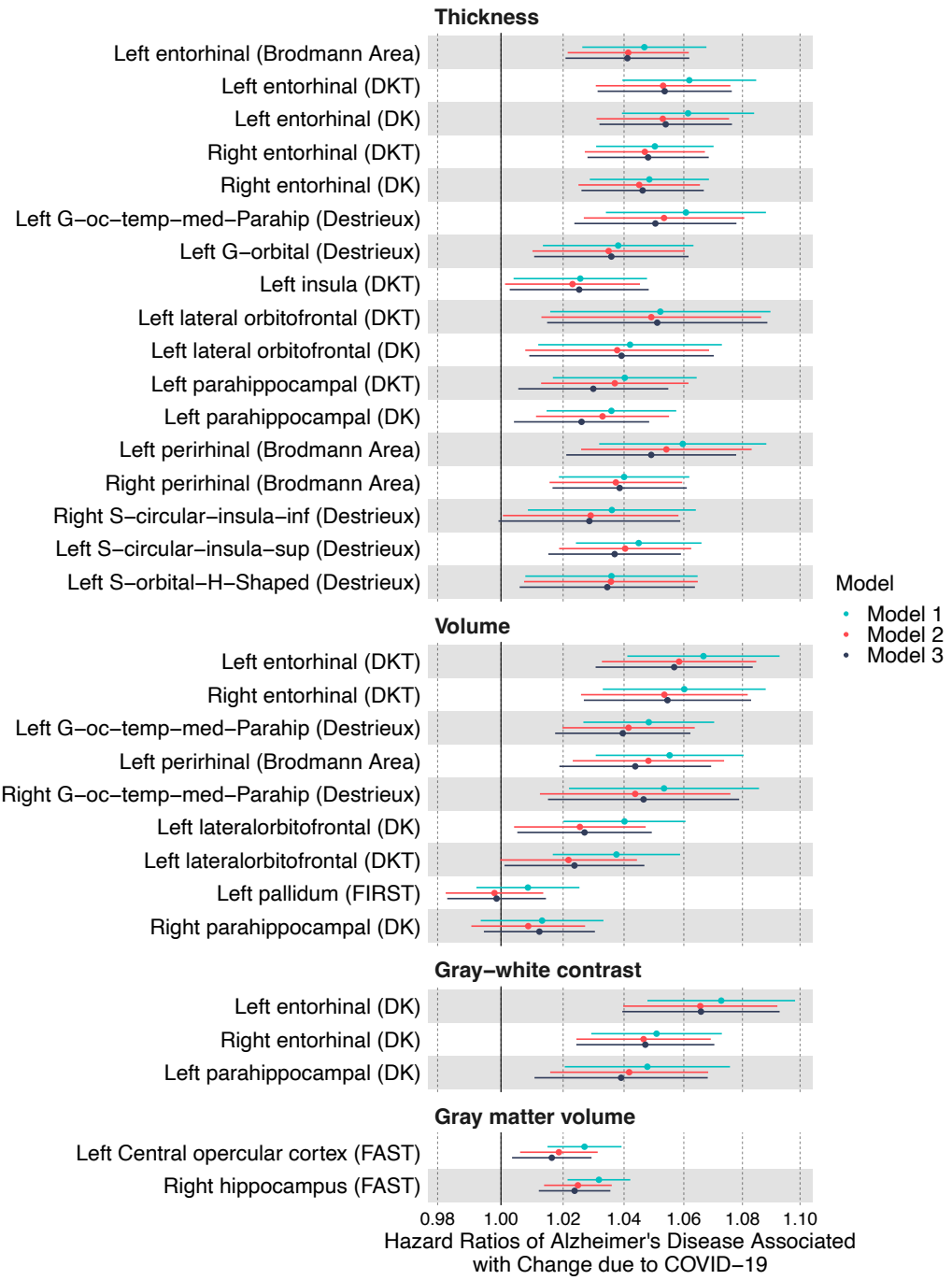


Figure 2.S4. Associations between individual imaging-derived phenotypes of the COVID-19 signature and incident Alzheimer's disease

Abbreviations: DK, Desikan-Killiany atlas; DKT, Desikan-Killiany-Tourville atlas.
Note: Model 1 adjusted for sex and race. Model2 further adjusted for total brain volume. Model 3 further adjusted for the number of APOE-ε4 alleles, assessment center, education, and Townsend deprivation index.

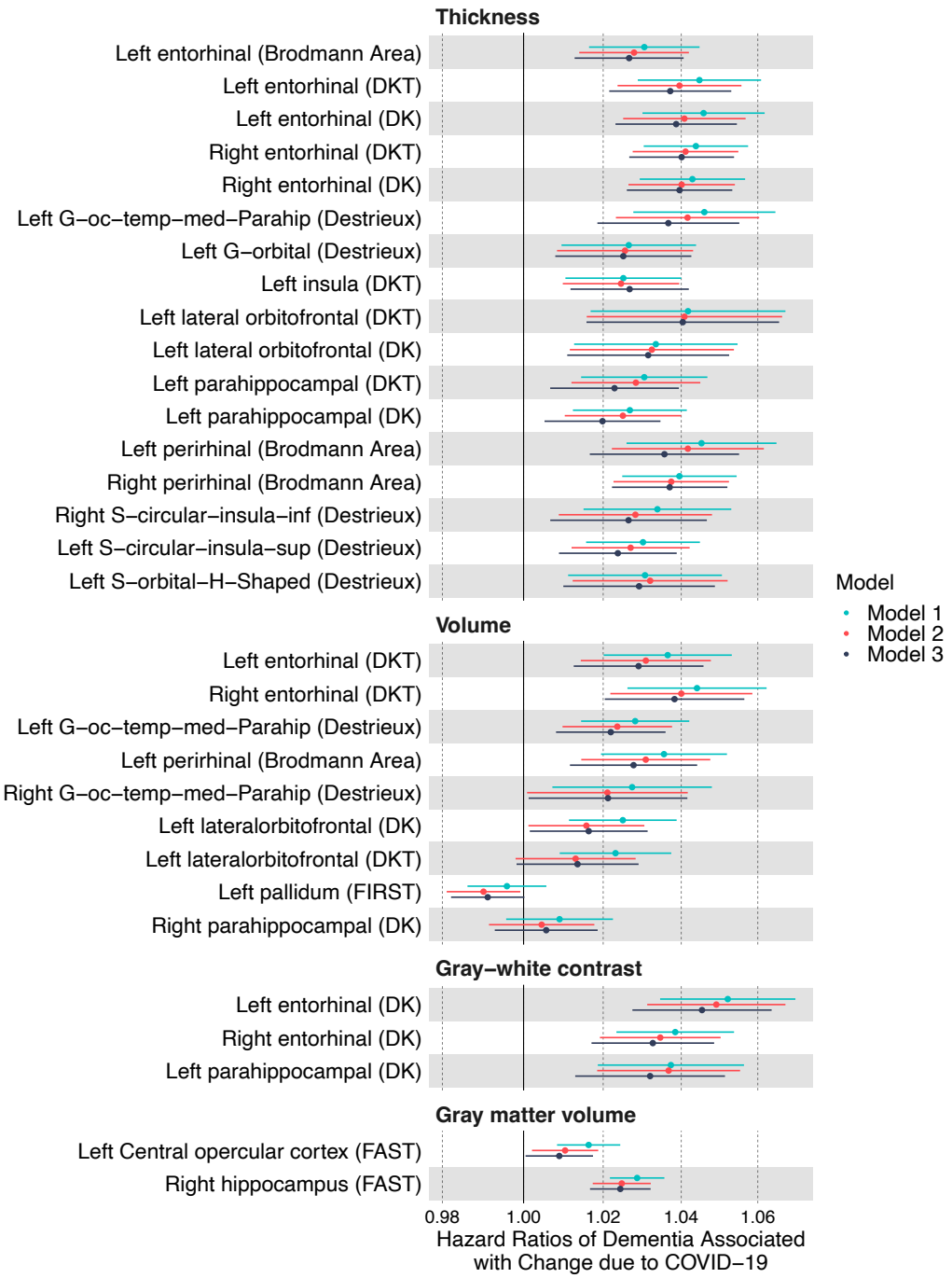


Figure 2.S5. Associations between individual imaging-derived phenotypes of the COVID-19 signature and incident dementia: time-updated Cox models

Abbreviations: DK, Desikan-Killiany atlas; DKT, Desikan-Killiany-Tourville atlas.
Note: Model 1 adjusted for sex and race. Model2 further adjusted for total brain volume. Model 3 further adjusted for the number of APOE-ε4 alleles, assessment center, education, and Townsend deprivation index.

Chapter 3

Association between age at onset of herpes viral infection and incidence of dementia

Jingxuan Wang, Peter Buto, J. Daniel Kelly, Eva Raphael, Felicia Chow,

Sarah F. Ackley, Renaud La Joie, John Kornak, M. Maria Glymour

Abstract

Background: Epidemiological studies suggest that herpes viral infection increases risk of dementia, yet evidence is inconsistent across studies. While prior studies included samples with a wide range of ages, no studies have investigated the association of age at onset of herpes viral infection with dementia.

Methods: We used data from the UK Biobank participants who had serological measurement of infectious agents and who had a clinical diagnosis for any of the following herpes viral infection: herpes simplex virus (HSV), varicella-zoster virus (VZV), Epstein-Barr virus (EBV), and human cytomegalovirus (CMV). We integrated serological data with clinical diagnoses to identify age at onset. Specifically, we identified participants without prior infection using the serological measurement and identified those with an infection history using both serological and clinical diagnosis data. We used Cox proportional hazards models to evaluate associations of age at onset of herpes viral infection with dementia.

Results: The HSV cohort included 58,676 participants, among whom 3,310 developed dementia. In the VZV cohort, 1,116 dementia cases occurred among 49,429 participants. In the EBV cohort, there were 162 cases among 9,033 participants. In the CMV cohort, 177 cases occurred among 10,092 participants. Participants with HSV onset before age 65 had a higher hazard of dementia compared to those without a history of HSV infection (HR = 1.76, 95% CI: 1.14-2.71 for onset before age 45; HR = 1.83, 95% CI: 1.23-2.71 for onset between ages 45 and 55; and HR = 1.75, 95% CI:

1.20-2.57 for onset between ages 55 and 65). None of the VZV onset groups showed a statistically significant higher dementia hazard compared to the no infection group, but there was a significant linear trend of declining hazard ratios with older onset age ($p < 0.01$). EBV and CMV associations did not reach statistical significance.

Conclusions: HSV infection before age 65 was associated with increased risk of dementia. Promoting cognitive health through the prevention and treatment of infectious diseases like HSV may complement and enhance existing biomedical and lifestyle-based approaches. Future research should further investigate the etiological links between HSV infection and dementia and leverage larger datasets with greater statistical power to examine other herpes viruses, such as VZV.

Introduction

Over 57 million individuals globally are currently living with dementia and the number is projected to increase to 153 million by 2050 as the population ages.[3] Dementia is a continuum usually with a long preclinical phase and evolving symptoms.[2] The hallmark of Alzheimer's disease (AD), the most common cause of dementia, is the accumulation of β-amyloid plaques and neurofibrillary tangles and can evolve decades before onset of clinical symptoms. Various chronic conditions including infection contribute to the onset and progression of dementia.

Herpesviridae is a large family of double-stranded DNA viruses that can enter a latent state after initial infection and reactivate from various triggers, including stress, immune activation, immune suppression, and other stimuli.[78] Herpes simplex virus type 1 (HSV-1), herpes simplex virus type 2 (HSV-2), varicella-zoster virus (VZV), Epstein-Barr virus (EBV), and human cytomegalovirus (CMV) are the most common, infecting the majority of the general population.[79] These herpesviruses can be grouped into three different subfamilies: *Alphaherpesvirinae* (HSV-1, HSV-2, VZV), *Betaherpesvirinae* (CMV) and *Gammaherpesvirinae* (EBV). HSV-1, HSV-2, and VZV consist of one subfamily that infects neurons while the other subfamilies infect immune cells throughout the body. This differential cellular infectivity forms a biological hypothesis that reactivation of HSV-1, HSV-2, and VZV is more likely to contribute to a biological pathway, leading to dementia and neurologic diseases than reactivation of EBV and CMV.

In addition to neuropathogenesis driven by cellular reactivation, herpes viral infection can elevate dementia risk through various other mechanisms, including infection-triggered inflammation,[80] direct damage from neurotropism,[81–83] promotion of β-amyloid deposition,[9,84,85] cardiovascular and metabolic complications,[9,86] and psychological disturbances.[87,88] Previous studies have established links between herpes viral infection and dementia.[12,78,89–101] For example, a large cohort study linked to national health registries in Finland found that individuals with hospital-treated herpes virus infections had a 2.3-fold increased hazard of developing dementia.[91] Yet evidence was inconsistent across studies. For example, a cohort study reported an increased risk of AD dementia associated with CMV (RR = 2.15, 95% CI: 1.42–3.27),[101] while another cohort study estimated no association (HR = 1.00, 95% CI: 0.56-1.79).[102] Inconsistent findings in previous studies may be due to heterogeneity in associations depending on age of infection with herpes virus, as different age groups may vary in their susceptibility to the long-term neurological effects of herpes viral infections.

Herpes viral infection may also contribute to dementia through cumulative effects over a lengthy preclinical phase with gradually emerging symptoms. A significant limitation of earlier studies has been their inability to explore the influence of the age at herpes virus infection onset on dementia risk, limiting the ability to assess both the heterogeneous effects across different onset ages and their potential cumulative impact over time. Another methodological challenge is accurately measuring infection history. While antibody or polymerase chain reaction (PCR) data offer precise infection detection, most prior research using these methods were limited by small sample sizes.

Conversely, studies that relied on clinical diagnoses from hospital inpatient or primary care records had larger sample sizes but faced issues of underdiagnosis and missing records. Many herpes viral infections are asymptomatic and unrecognized, and routine testing is limited, resulting in incomplete or imprecise infection measurement in clinical records.[103]

Therefore, we leveraged large-scale cohort data from the UK Biobank, which augments clinical diagnosis records with antibody tests on a large, randomly selected subsample, to investigate the association between age at onset of five types of herpes viral infection and dementia incidence using a test-negative cohort study design.

Methods

Study population and participants

The UK Biobank is a prospective cohort of 502,411 adults aged 40-69 years who were recruited from 2006-2010.[26] Participants attended one of 22 assessment centers across the United Kingdom, where they provided electronic signed consent, completed a touchscreen questionnaire and a face-to-face interview with study nurses. At enrollment, participants also underwent physical measurements and provided blood, urine, and saliva samples. Follow-up data on health outcomes included electronic health records (EHRs)-based hospital admission and primary care data and national cancer and death register data. A pilot study with randomly selected 9,695 participants was conducted to assess the seroprevalence of 20 specific infectious agents using a Multiplex Serology platform, including human herpes, hepatitis, polyoma, papilloma, and retroviruses, as

well as *Chlamydia trachomatis*, *Helicobacter pylori* and *Toxoplasma gondii*.[104] Our analysis comparing seroprevalence and clinical diagnoses of herpes viral infections indicates that clinical diagnoses have high specificity (ranging from 0.94 to 1.00) but low sensitivity (ranging from <0.01 to 0.07), as well as high positive predictive value (ranging from 0.76 to 0.94) but low negative predictive value (ranging from 0.05 to 0.42; **Table 3.S1**). This suggests that while clinical diagnoses effectively identify exposed individuals, accurately identifying unexposed individuals requires serological data.

Therefore, for the current analyses, we used a test-negative cohort study design. The control group consisted of participants who tested negative, while the exposure group included participants with a positive antibody test or a clinical diagnosis. The analytical sample included the 9,695 participants who had serological measurement of infectious agents and 65,433 participants who did not have serological measurement but had a clinical diagnosis for any of HSV-1 and HSV-2, VZV, EBV, and CMV.

Identification of herpes infection and onset age

We identified the history of herpes viral infection from both the seropositivity based on the serological measurement and clinical diagnosis recorded in EHRs. For herpes simplex virus (HSV), data on HSV-1 and HSV-2 were combined because ICD codes do not differentiate between the two types. The serological measurements were conducted using the Multiplex Serology platform, a method developed based on the principles of an enzyme-linked immunosorbent assay (ELISA), which quantifies serum immunoglobulin levels. 20 infectious agents were included in the platform, including

HSV-1 and HSV-2, VZV, EBV, and CMV. Further details can be found in previous publications.[104] The platform facilitated the detection of pathogen-specific antibodies, generating quantitative data represented as median fluorescence intensity (MFI) values for each antigen and serum sample using the Luminex reader. Seropositivity for each antigen was determined by analyzing percentile plots, with cut-off values selected to optimize sensitivity and specificity based on validation studies.

The clinical diagnosis was ascertained via linked hospital inpatient and primary care data. In the UK Biobank, hospital inpatient records are coded using ICD-9 and ICD-10, and primary care data are coded using Read version 2 (Read v2) and Clinical Terms Version 3 (CTV3). The list of ICD and Read codes to identify HSV, VZV, EBV, and CMV infection can be found in **Table 3.S2**.

The age at onset of herpes viral infection was measured by age at earliest infection diagnosis for participants with a clinical diagnosis recorded in EHR or age at baseline for those tested positive in the serological measurement but did not have any diagnosis.

Ascertainment of dementia outcomes

Our primary outcome was the incidence of all-cause dementia, identified using linked hospital inpatient, mortality, and primary care data. Dementia cases were ascertained through a comprehensive list of ICD and Read codes, including Alzheimer's disease (AD), vascular dementia, frontotemporal dementia, Lewy body dementia, alcohol-related dementia, and other dementias (**Table 3.S2**). Incident dementia was defined as a

recorded diagnosis of dementia—either as a primary or secondary diagnosis in hospital or primary care records, or as a contributory cause of death in mortality data. We also examined AD dementia as the secondary outcome.

Covariates

The following covariates were included in multivariable analyses: age, sex (female, male), race (White, Black, Asian, Other), education (lower than high school, high school or above), Townsend deprivation index (measure of area-level deprivation, categorized into quintiles), smoking status (never, previous, current), and physical activity (defined as ≥75 minutes per week of vigorous activity or ≥150 minutes per week of moderate activity). We conceptualized two sets of covariate adjustment. The ‘base’ model included covariates that could not be affected by exposure: age, sex, and race. The ‘full’ model additionally incorporated education, Townsend deprivation index, smoking status, and physical activity. Age and sex were obtained from a central registry at recruitment and updated by participants. Race, education, smoking status, and physical activity were self-reported at baseline. The Townsend deprivation index, a composite measure of area-level socioeconomic status based on employment, home ownership, car ownership, and household overcrowding,[105] was calculated prior to baseline using national census data and determined by participants' postcodes at recruitment.

Statistical analysis

To examine the association between age at onset of herpes viral infection and incidence of dementia, we used separate Cox proportional-hazards models for each herpes

infection type (HSV, VZV, EBV, or CMV). For each infection type, participants were categorized into five exposure groups: no infection history, infected before age 45 years, between 45 and 55 years, between 55 and 65 years, and after 65 years.

Participants were followed from baseline to the earliest of date of dementia diagnosis, death from any cause, or administrative censoring, which was defined as the latest date of any dementia diagnosis in the sample (December 12, 2022). For participants without a serological measurement but with a clinical diagnosis for herpes infection, we start the follow-up since the time of the diagnosis. In our analysis, herpes infection was modeled as a time-varying exposure. Participants without a history of herpes infection at baseline contributed person-time from baseline to the “no infection” group until the date of their first herpes infection diagnosis. Following diagnosis, they contributed person-time to the corresponding herpes infection group, classified based on their age at diagnosis. To test whether the associations follow a linear trend regarding the onset age, we conducted a linear trend test using herpes infection groups as a linear variable.

We repeated the primary analysis using the secondary outcome AD dementia. We conducted two sensitivity analyses. First, we excluded participants who did not have any serological measurement but had a herpes infection diagnosis after baseline. Additionally, we incorporated age at onset as a continuous variable in the model rather than age groups. All statistical analyses were performed using R version 4.4.0.

Results

The seroprevalence for HSV, VZV, EBV, and CMV among 9,695 participants who had serological measurement of infectious agents were 74.6%, 92.4%, 94.7%, and 58.3%, respectively (**Table 3.S1**). Among the 58,676 participants included in the analysis for HSV, 1,239 dementia cases were identified over a median follow-up of 13.2 years after UKB enrollment. For VZV, 1,116 dementia cases occurred among 49,429 participants during a median follow-up of 13.1 years. In the EBV analysis, 162 dementia cases were observed among 9,033 participants over a median follow-up of 13.7 years. Similarly, for CMV, 177 dementia cases were identified among 10,092 participants during a median follow-up of 13.8 years. The mean age of participants at enrollment was 58.1 years (SD = 8.4) for the HSV analysis, 59.2 years (SD = 8.4) for the VZV analysis, 57.5 years (SD = 8.6) for the EBV analysis, and 57.2 years (SD = 8.2) for the CMV analysis (**Table 3.1**). The onset of herpes viral infection predominantly occurred between the ages of 45 and 70 for all types of herpes viruses (**Figure 3.1**).

Participants with HSV onset before age 65 had a higher hazard of dementia compared to those without a history of HSV infection. In the primary analysis base model, the hazard ratio (HR) was 1.76 (95% CI: 1.14-2.71) for onset before age 45, 1.83 (95% CI: 1.23-2.71) for onset between ages 45 and 55, and 1.75 (95% CI: 1.20-2.57) for onset between ages 55 and 65 (**Figure 3.2**). However, the hazard ratio attenuated for onset after age 65 (HR = 1.43, 95% CI: 0.97-2.12). No significant linear trend in age of onset and dementia incidence was observed (p = 0.97; **Table 3.S3**). For VZV, while none of the onset groups reached a statistically significant higher dementia hazard compared to

the no VZV infection group (**Figure 3.2**), the hazard ratio declined with increasing onset age, and a statistically significant linear trend was observed ($p < 0.01$; **Table 3.S3**). EBV and CMV associations did not reach statistical significance (**Figure 3.2**), and we did not observe any linear trends.

In our secondary analysis, HSV onset ages between 45 and 65 were associated with a higher AD dementia hazard (**Figure 3.S1**). Compared to participants with no history of HSV infection, those with HSV onset between 45 and 55 had a hazard ratio of 1.82 for AD dementia (95% CI: 1.21-2.74), and those with HSV onset between 55 and 65 had a hazard ratio of 1.72 (95% CI: 1.16-2.56; **Table 3.S4**). Similar to results for dementia as the outcome, no single VZV infection group had a statistically significant higher AD dementia hazard, but there was a statistically significant linear trend ($p = 0.01$). EBV and CMV did not reach a statistically significant level of association with AD, but confidence intervals were very wide and could not rule out large harms or benefits.

Excluding participants who did not have any serological measurement but had a herpes infection diagnosis after baseline did not substantially change the results (**Figure 3.S2**, **Table 3.S5**). When operationalizing onset age as a continuous variable, each 10-year earlier onset for VZV was associated with higher dementia hazard (HR = 1.08, 95% CI: 1.02-1.15), while confidence intervals for HRs associated with other herpes types included the null (**Table 3.S6**).

Discussion

In this cohort study, we examined the associations between age at onset of five herpes viral infections and subsequent dementia incidence. We found that participants with an HSV infection prior to age 65 had a higher dementia risk compared to individuals without an HSV infection. While we did not observe statistically significant associations between VZV and dementia, the effect size point estimates decreased with onset age, suggesting a potential cumulative effect of VZV on dementia. Our data were too scant to evaluate evidence linking EBV and CMV to dementia.

Our results were consistent with prior work. Previous epidemiological studies suggest an association between HSV infection and an increased risk of dementia. A meta-analysis of 19 studies found that individuals with a history of HSV infection had a 2.23-fold higher risk of developing dementia (95% CI: 1.18-2.29).[93] Evidence on the association between VZV and dementia remains mixed. A cohort study reported a slightly increased dementia risk associated with VZV infection (HR = 1.11, 95% CI: 1.04-1.17),[94] while a systematic review found no association between VZV IgG serum antibodies and dementia in two case-control studies, and VZV DNA was not detected in post-mortem brain samples in two other case-control studies.[12,106–109] There is no strong evidence supporting a positive association between EBV or CMV and dementia or AD. For example, two cohort studies found no link between EBV or CMV seropositivity and subsequent dementia risk,[96,99] and another cohort study reported no association between CMV and AD dementia.[102]

While prior studies that examined the associations between herpes viral infection and dementia included samples with a wide range of ages, these studies have not explicitly considered age at infection onset, which could provide greater insights for the varying relationship across onset ages. For example, HSV may influence dementia risk through multiple mechanisms. First, the neurotropism of the virus directly damages the central nervous system, potentially leading to neurodegeneration and cognitive decline.[81,110,111] Second, HSV can establish lifelong latency in the central nervous system, primarily in the trigeminal ganglia, with periodic reactivation that can trigger neuroinflammatory responses contributing to dementia pathogenesis.[9,112,113] Third, HSV infection has been linked to β-amyloid accumulation and tau hyperphosphorylation, both of which lead to higher AD dementia risk.[84,112,114,115] The age of onset plays an important role in these processes, as earlier infection might impose a cumulative, long-term effect on dementia risk. In our study, we found that the strongest associations between HSV and dementia were observed in participants infected before age 65. This suggests that the influence of HSV infection on dementia risk may require time to manifest, potentially due to cumulative effects over the lifecourse.

Although the associations between VZV infection and dementia were not statistically significant in themselves, we observed a linear trend associating earlier onset with higher incidence of dementia compared to later onset. In addition, when modeling onset age as a continuous variable, we found each 10-year earlier onset was associated with a 1.08-fold higher dementia hazard. This may suggest that an earlier onset may be associated with a higher risk of dementia, though certainly merits further interrogation.

Beyond incorporating age at infection onset, our study has several additional strengths. Previous research relying solely on antibody or PCR tests was limited by small sample sizes, typically ranging from 80 to 1,000 participants,[97–100] with the exception of one study using UK Biobank data that did not account for confounder adjustment.[96] Studies based on questionnaire data or EHR-derived clinical diagnoses faced challenges of underreporting and underdiagnosis. To enhance statistical power and improve exposure measurement, we integrated serological data with clinical diagnoses of herpes infection. In our study, serological measurements provided a well-defined control group, while clinical diagnoses, which effectively identified exposed individuals, ensured a large and robust exposure group. Additionally, we leveraged longitudinal UK Biobank data with a median follow-up of over 13 years for information on dementia onset. Lastly, our study benefited from detailed information on potential confounders, which we carefully adjusted for in our analysis.

This study has several limitations. For participants who were seropositive for herpes infection but lacked a clinical diagnosis in their EHR, we assigned their age of onset as their age at UKB enrollment. This may introduce measurement error, as their actual infection could have occurred at any point before baseline. Similarly, for participants with a recorded diagnosis in their EHR, their actual onset age may have been much earlier than the date of diagnosis recorded in the EHR system. Additionally, dementia ascertainment relied on clinical diagnoses from EHR, which may not capture all cases due to underdiagnosis, missing encounters, or healthcare utilization outside the EHR system.[62] Lastly, despite the large sample size and rich phenotypic data, the UK

Biobank is subject to healthy volunteer selection bias, as it overrepresents healthy, White individuals with higher socioeconomic status.[46] This may limit the generalizability of our findings to the general population.

In conclusion, our study found that HSV infection before age 65 was associated with increased incidence of dementia. Promoting cognitive health through the prevention and treatment of infectious diseases like HSV may complement and enhance existing biomedical and lifestyle-based approaches. Future research should further investigate the etiological links between HSV infection and dementia and leverage larger datasets with greater statistical power to examine other herpes viruses, such as VZV. Additionally, studies employing diverse methodological approaches, including genetic analyses and neuroimaging biomarkers, may be necessary to fully elucidate the relationship between herpes infections and dementia.

Tables

Table 3.1. Baseline characteristics of participants included in the analysis for each herpes

Characteristic (%)	HSV (N = 58,752)	VZV (N = 49,429)	EBV (N = 9,033)	CMV (N = 10,092)
Mean age (SD)	58.1 (8.4)	59.2 (8.4)	57.5 (8.6)	57.2 (8.2)
Age of onset				
No infection	2,238 (3.8%)	684 (1.4%)	494 (5.5%)	3,930 (38.9%)
<45	13,375 (22.8%)	9,535 (19.3%)	1,713 (19.0%)	700 (6.9%)
45-55	16,199 (27.6%)	12,356 (25.0%)	2,040 (22.6%)	1,581 (15.7%)
55-65	18,227 (31.1%)	16,904 (34.2%)	3,026 (33.5%)	2,500 (24.8%)
>65	8,637 (14.7%)	9,950 (20.1%)	1,760 (19.5%)	1,381 (13.7%)
Sex				
Female	35,196 (60.0%)	28,489 (57.6%)	5,029 (55.7%)	5,621 (55.7%)
Male	23,480 (40.0%)	20,940 (42.4%)	4,004 (44.3%)	4,471 (44.3%)
Race				
Asian	1,887 (3.2%)	1,156 (2.3%)	212 (2.3%)	260 (2.6%)
Black	866 (1.5%)	589 (1.2%)	162 (1.8%)	175 (1.7%)
Other	784 (1.3%)	589 (1.2%)	138 (1.5%)	154 (1.5%)
White	55,139 (94.0%)	47,095 (95.3%)	8,521 (94.3%)	9,503 (94.2%)
Education				
Lower than high school	21,456 (36.6%)	17,980 (36.4%)	3,212 (35.6%)	3,465 (34.3%)
High school or above	36,483 (62.2%)	30,896 (62.5%)	5,716 (63.3%)	6,524 (64.6%)
Missing	737 (1.3%)	553 (1.1%)	105 (1.2%)	103 (1.0%)
Townsend deprivation index				
Least deprived	11,632 (19.8%)	10,075 (20.4%)	1,830 (20.3%)	2,114 (20.9%)

Characteristic (%)	HSV (N = 58,752)	VZV (N = 49,429)	EBV (N = 9,033)	CMV (N = 10,092)
Moderate deprived	35,538 (60.6%)	30,090 (60.9%)	5,328 (59.0%)	6,015 (59.6%)
Most deprived	11,431 (19.5%)	9,202 (18.6%)	1,867 (20.7%)	1,953 (19.4%)
Missing	75 (0.1%)	62 (0.1%)	8 (0.1%)	10 (0.1%)
Smoking status				
Never	31,811 (54.2%)	26,474 (53.6%)	4,917 (54.4%)	5,567 (55.2%)
Previous	20,893 (35.6%)	17,729 (35.9%)	3,115 (34.5%)	3,452 (34.2%)
Current	5,708 (9.7%)	4,985 (10.1%)	954 (10.6%)	1,023 (10.1%)
Missing	264 (0.5%)	241 (0.5%)	47 (0.5%)	50 (0.5%)
Physical activity				
Yes	44,583 (76.0%)	37,410 (75.7%)	6,879 (76.2%)	7,700 (76.3%)
No	12,558 (21.4%)	10,805 (21.9%)	1,939 (21.5%)	2,183 (21.6%)
Missing	1,535 (2.6%)	1,214 (2.5%)	215 (2.4%)	209 (2.1%)

Abbreviation: HSV, herpes simplex virus; VZV, varicella-zoster virus; EBV, Epstein-Barr virus; CMV, human cytomegalovirus.

Figures

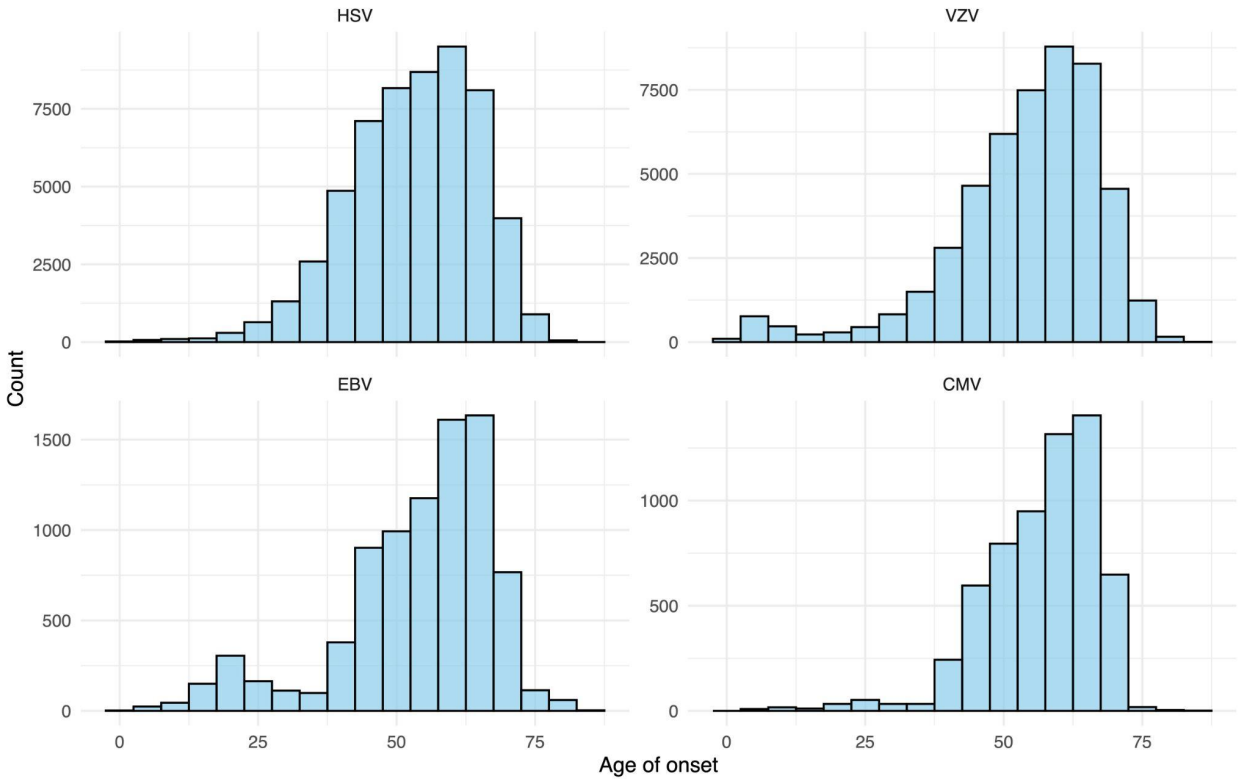


Figure 3.1. Age at onset of each herpes infection

Abbreviation: HSV, herpes simplex virus; VZV, varicella-zoster virus; EBV, Epstein-Barr virus; CMV, human cytomegalovirus.

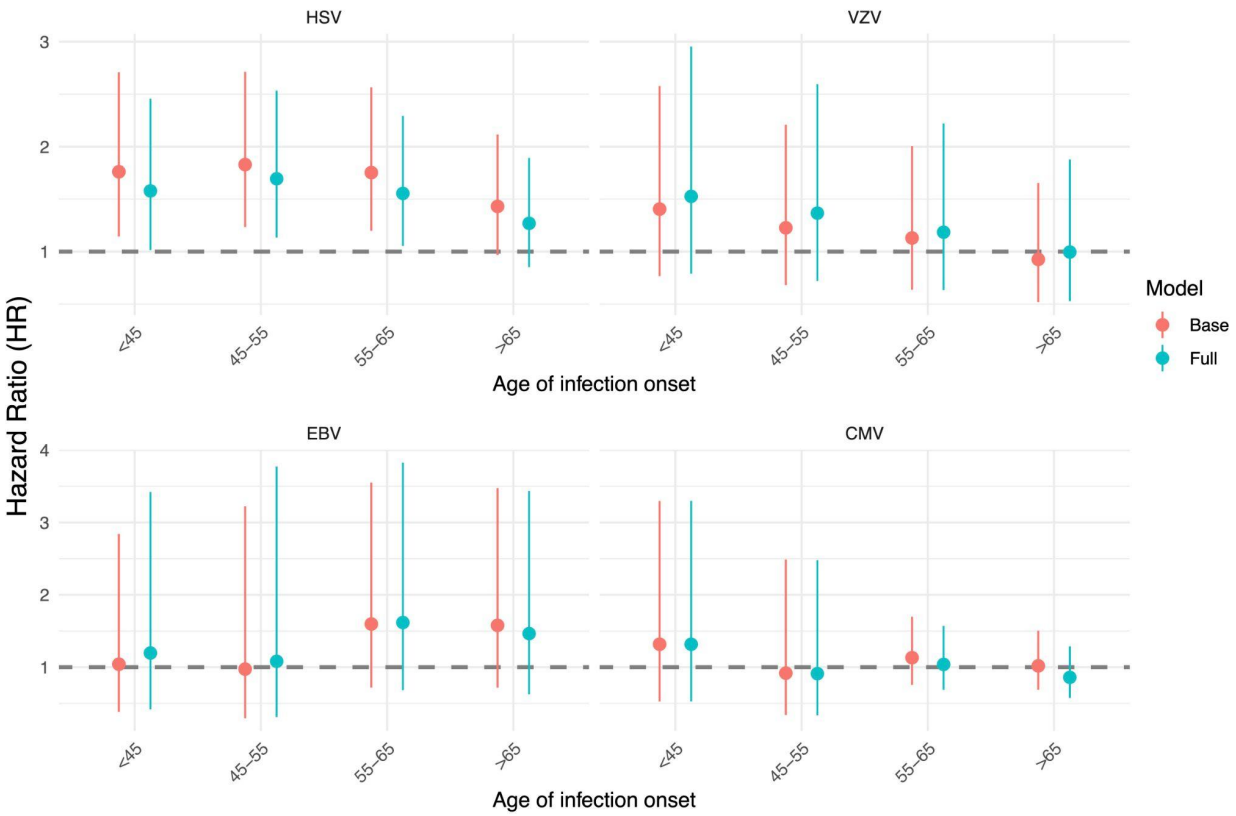


Figure 3.2. Association between age at onset of each herpes infection and subsequent dementia incidence

Abbreviation: HSV, herpes simplex virus; VZV, varicella-zoster virus; EBV, Epstein-Barr virus; CMV, human cytomegalovirus.
Note: The base model included age, sex, and race. The full model additionally incorporated education, Townsend deprivation index, smoking status, and physical activity.

Supplement

Table 3.S1. Comparison for clinical diagnosis against seroprevalence of herpes viral infections

	HSV	VZV	EBV	CMV
Seroprevalence	74.6%	92.4%	94.7%	58.3%
Sensitivity	0.07	0.05	<0.01	<0.01
Specificity	0.94	0.96	1.00	1.00
Positive predictive value	0.76	0.94	0.92	0.86
Negative predictive value	0.26	0.08	0.05	0.42

Abbreviation: HSV, herpes simplex virus; VZV, varicella-zoster virus; EBV, Epstein-Barr virus; CMV, human cytomegalovirus.
Note: The seroprevalence was estimated among 9,695 participants who had serological measurement of infectious agents. We compared the clinical diagnosis from electronic health records against seroprevalence among these 9,695 participants who had serological measurement of infectious agents. We only considered clinical diagnosis before the time of the serological measurement.

Table 3.S2. ICD, Read v2, and CTV3 codes for outcome and exposures

	ICD-9	ICD-10	Read v2	CTV3
Outcome				
Dementia	331.0, 290.4, 331.1, 046.1, 291.1, 291.2, 290.0, 290.1, 290.2, 290.3, 290.8, 290.9, 294.1, 294.8, 331.2	F00, F00.0, F00.1, F00.2, F00.9, G30, G30.0, G30.1, G30.8, G30.9, F01, F01.0, F01.1, F01.2, F01.3, F01.8, F01.9, I67.3, F02.0, G31.0, A81.0, F02.1, F02.2, F02.3, F02.4, F10.6, F02, F02.8, F03, F05.1, G31.1, G31.8	F110.,F1100,F1101,Eu00.,Eu000,Eu001,Eu002,Eu00z,Fyu30,E004.,E0040,E0041,E0042,E0043,E004z,Eu01.,Eu010,Eu011,Eu012,Eu013,Eu01y,Eu01z,F21y2,G678.,Eu020,F111.,A411.,F11x7,E011.,E0110,E0111,E0112,E011z,E012.,E0120,Eu021,A4110,Eu022,Eu023,Eu024,F11x9,Eu106,E000.,E001.,E0010,E0011,E0012,E0013,E001z,E002.,E0020,E0021,E002z,E003.,E00y.,E00z.,E041.,E04y.,Eu02.,Eu025,Eu02y,F116.,Eu02z,Eu041,F112.,F10y.,F10y0,F10y1,F10y2,F10yz,F118.,F11y.,F11y2,F11yz,Fyu31,E00..,1461.,38C13,3AE3.,3AE4.,3AE5.,3AE6.,66h..,6AB..,8BM02,8BM50,8BM60,8BPa.,8CMe0,8CMG2,8CMZ.,8CMZ0,8CMZ1,8CMZ2,8CMZ3,8CSA.,8Hla.,8IAe0,8IAe2,9hD..,9hD0.,9hD1.,9Ou..,9Ou1.,9Ou2.,9Ou3.,9Ou4.,9Ou5.	F110.,XaIKB,XaIKC,Eu00.,X002x,X002x,X002x,X002x,X002x,X002x,X002x,X002x,X0030,X0030,X0030,X0030,X0030,Eu00.,X0030,X0030,Eu002,Eu00z,Eu00z,Fyu30,XE1Xs,Xa0lH,E0040,E0041,E0042,E0043,E004z,XE1Xs,XE1Xs,XE1Xs,X003R,Xa0lH,Xa0lH,Xa0lH,X003T,X003V,Eu01y,Eu01z,F21y2,F21y2,XaIRJ,Eu020,F111.,A411.,F11x7,E011.,E011.,E011.,E011.,E0111,E0112,E011z,XE1Xu,X00Rk,E0120,Eu021,XabVp,XabVp,Eu022,Eu023,X003P,X003P,XaOfZ,Eu106,Eu106,Eu106,E000.,E001.,E0010,E0011,E0012,E0013,E001z,E002.,E0020,E0021,E002z,E003.,XE1Xt,X00R0,E00z.,E041.,E04y.,Eu02.,XaKyY,Eu02y,X003A,XE1Z6,XE1Z6,E00z.,E00z.,Eu02z,Xa1GB,Eu02z,X00R2,XE1Z6,E00z.,Eu02z,XE1Xr,Eu041,F112.,F10y.,F10y0,F10y1,XaPws,F10yz,X0037,F11..,X003m,F11yz,Fyu31,XE1Xr,X00R2,X002w,1461.,XaaeA,XaJBU,XaJBV,XaJBW,XaJBX,XaMJC,XaMGF,XabtQ,Xaefu,Xaefv,XaaiW,XacLx,XabEl,XaaBZ,XacIx,XacIy,XacIz,XacJ0,XabEk,XaYFR,XabEi,XacM2,XaLFf,XaLFo,XaLFp,XaMFy,XaMG0,XaMGG,XaMGI,XaMGJ,XaMGK

	ICD-9	ICD-10	Read v2	CTV3
Exposure				
HSV	054	B00, A60	A5..., A54.., A540., A541., A5410, A5411, A5412, A5413, A5414, A5415, A5416, A5417, A5418, A5419, A541z, A542., A543., A544., A5441, A5442, A5443, A5444, A5445, A544z, A545., A546., A548., A54x., A54x0, A54x1, A54x2, A54x3, A54x4, A54xz, A54y., A54z., Ayu4G, AyuA., AyuA0, AyuA1, F0113, F0304, F4A1., F5015, A54.., A540., A541., A5410, A5411, A5412, A5413, A5414, A5415, A541z, A542., A543., A544., A5440, A5441, A5442, A5443, A5444, A5445, A544z, A545., A546., A548., A54x., A54x0, A54x1, A54x2, A54x3, A54xz, A54y., A54z.	1C9Z., 2524., 2DC3., A54.., A540., A541., A541., A541., A5410, A5411, A5411, A5412, A5412, A5413, A5413, A5414, A5415, A5415, A5415, A5415, A5415, A5415, A541z, A542., A543., A543., A544., A544., A5441, A5441, A5442, A5442, A5443, A5443, A5444, A5444, A544z, A545., A546., A546., A54x., A54x0, A54x3, A54x3, A54xz, A54y., A54z., A70.., A795., A796., A797., A79y., Ayu4G, AyuA0, AyuA1, F011., F011., F0113, F0113, F4A1., F4A1., F501., F501., F5013, F5013, F5015, F5015, J0314, J080., K4202, K4202, X0014, X0014, X0017, X0017, X00YV, X00YV, X00Zh, X00Zo, X00Zo, X00aV, X00aV, X00ah, X00ah, X00am, X00am, X00ao, X00ao, X00iA, X00iA, X00mG, X00mH, X1002, X100g, X100g, X20Pf, X20RC, X20RD, X20RF, X20RG, X304D, X304D, X4013, X4013, X407K, X407K, X407L, X407L, X407g, X407g, X407h, X407h, X70JO, X70JO, X70JO, X70JP, X70JU, X70JU, X70JV, X70JV, X70JW, X70JW, X70JY, X70JZ, X70Ja, X70Ja, X75hn, X75hn, X75i8, X75i8, XE0RB, XE0RB, XE0RC, XE0RD, XE0RD, XE0RE, XE0Rh, XE0sE, XE168, XE16N, XE16N, XE16g, XE16g, XE2vG, XE2vH, XM02r, XM02r, XM07J, Xa0KV, Xa0KV, Xa0OP, Xa0bj, Xa0bj, Xa0ld, Xa0le, Xa0le, Xa0p8, Xa0p8, Xa0p8, Xa1sE, Xa3e4, Xa4cK, Xa4cK, Xa86V, XaFx8, XaFx8,

	ICD-9	ICD-10	Read v2	CTV3
				XaINK, XaINM, XaPEC, XaPEC, XaPEC, XaPEC, XaPEC, XaPED, XaPED, XaPED, XaPED, XaPED, XaPER, XaPER, XaPER, XaPER, XaPER, XaPES, XaPES, XaPES, XaPES, XaPES, XaY4o, 2524., A540., A541., A541., A5410, A5411, A5412, A5413, A541z, A544., A5440, A5441, A5442, A5443, A5444, A544z, A545., A546., A54x., A54x0, A54xz, A54y., AyuA1, F0113, F4A1., F501., F5015, H024., J0314, J080., K425., X000d, X0017, X00ao, X00bg, X00i6, X00i8, X00iA, X00mH, X100g, X20RC, X20RD, X20RF, X20RG, X407f, X407g, X407h, X407l, X407m, X5006, X70JN, X70JO, X70JP, X70JP, X70JS, X70JU, X70JU, X70JV, X70JV, X70JY, X70JZ, X70Ja, X75hn, X75i8, XE0RB, XE0RB, XE0RD, XE0RE, XE0Xn, XE179, XM02r, XM0tD, XM0tD, Xa0ld
VZV	052, 053	B01, B02	A52.., A520., A521., A52x., A52y., A52z., A53.., A530., A531., A5310, A5311, A5312, A5313, A5314, A5315, A531z, A532., A5320, A5321, A5322, A5323, A5324, A532z, A53x., A53x0, A53x1, A53xz, A53y., A53z., A5440, AyuA2, AyuA3, AyuA4, AyuA5, F0112, F0117, F0309,	A52.., A520., A520., A52x., A52y., A52z., A531., A5311, A5311, A5313, A5313, A531z, A532., A532., A5320, A5320, A5321, A5321, A5322, A5322, A5323, A5323, A532z, A532z, A53x., A53x1, A53xz, A53y., A53z., A5440, AyuA2, AyuA3, AyuA4, AyuA5, F011., F011., F011., F011., F0112, F0112, F0350, F0350, F300., F300., F311., F311., F3744, F3744, F501., F501., F5013, F5013, F5016, F5016, H24y7, H24y7, X0016, X0016, X001a, X001a, X002X, X002q, X002q, X009c, X009c, X00B6, X00B6, X00YV, X00YV, X00Zh, X00Zo, X00Zo, X00am, X00am,

	ICD-9	ICD-10	Read v2	CTV3
			F0350, F0370, F300., F3744, F5016, H24y7, A52.., A520., A521., A52x., A52y., A52z., A53.., A530., A531., A5310, A5311, A5312, A5313, A5314, A5315, A531z, A532., A5320, A5321, A5322, A5323, A5324, A532z, A53x., A53x0, A53x1, A53xz, A53y., A53z., F0117	X00ao, X00ao, X00iA, X00iA, X00jD, X00jD, X00ml, X20RH, X507c, X70JA, X70JA, X70JA, X70JA, X70JA, X70JF, X70JG, X70JI, X70JJ, X70JN, X70JS, X75sv, X75sv, XE0R9, XE0R9, XE0RA, XE168, XE16g, XE16g, XE2x9, XE2x9, Xa0KV, Xa0KV, Xa0p8, Xa0p8, Xa2C8, Xa2C8, Xa2C9, Xa2C9, Xa3e3, XaBsP, XaBsP, XaBsP, XaBsP, XaQj2, XaQj2, A52x., A52y., A52z., A53.., A53.., A5313, A531z, A5320, A5321, A5322, A5323, A532z, A53x., A53x1, A53xz, A53y., A53z., AyuA2, AyuA3, AyuA4, AyuA5, F0112, F0350, F300., F3010, F3744, F501., F5016, H24y7, X000d, X0016, X001a, X002X, X009c, X009e, X00B6, X00bg, X00i6, X00i8, X00iA, X00jD, X00ml, X20RH, X20RH, X5003, X5006, X5006, X70JA, X70JF, X70JG, X70JI, X70JJ, X70JN, X75sv, XE0R9, XE0RA, XE179, XM0tD, Xa2C8, Xa2C8, Xa2C9
EBV	075	B27.0	A750., A752., A75.., A750.	A54x3, A750., X70LI, XaEYX, A70.., A75.., A750., AyuD5, X00mL, X100e, X5007, X70LI, X70Lv, Xa9C8
CMV	078.5	B25, B27.1	A751., A785., A7850, A7851, A7852, A785X, A799., AyuD., AyuD0, AyuD1, F0307, H241., K0A9., A785., A7850, A7851, A7852, A785X	A70.., A70.., A785., A7852, A795., A796., A797., AyuD0, AyuD1, H241., H241., J07.., J6311, J6311, X0013, X0013, X00ca, X00k7, X303q, X308p, X308p, X30Mj, X70Lv, XE0RV, XE0RV, A70.., A7852, AyuD1, H241., J07.., Q409., Q409z, X00k7, X20T9, X306h, X308p, XE0RV, XE0RV

Abbreviation: HSV, herpes simplex virus; VZV, varicella-zoster virus; EBV, Epstein-Barr virus; CMV, human cytomegalovirus.

Table 3.S3. Association between age at onset of each herpes infection and subsequent dementia incidence

	HSV (N = 58,752)		VZV (N = 49,429)		EBV (N = 9,033)		CMV (N = 10,092)	
	Base	Full	Base	Full	Base	Full	Base	Full
Age of onset: HR (95% CI)								
No infection	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
<45	1.76 (1.14-2.71)	1.58 (1.01-2.46)	1.41 (0.77-2.58)	1.53 (0.79-2.95)	1.04 (0.38-2.84)	1.20 (0.42-3.42)	1.32 (0.53-3.30)	1.32 (0.53-3.30)
45-55	1.83 (1.23-2.71)	1.69 (1.13-2.53)	1.23 (0.68-2.21)	1.37 (0.72-2.60)	0.97 (0.29-3.22)	1.08 (0.31-3.77)	0.92 (0.34-2.49)	0.91 (0.33-2.48)
55-65	1.75 (1.20-2.57)	1.55 (1.05-2.29)	1.13 (0.64-2.01)	1.19 (0.63-2.22)	1.60 (0.72-3.55)	1.62 (0.68-3.83)	1.13 (0.75-1.70)	1.04 (0.69-1.57)
>65	1.43 (0.97-2.12)	1.27 (0.85-1.89)	0.93 (0.52-1.65)	1.00 (0.53-1.88)	1.58 (0.72-3.47)	1.46 (0.62-3.44)	1.02 (0.69-1.50)	0.86 (0.57-1.29)
Trend test: p-value								
Linear trend	0.97	0.56	<0.01	<0.01	0.13	0.30	0.82	0.56

Abbreviation: HSV, herpes simplex virus; VZV, varicella-zoster virus; EBV, Epstein-Barr virus; CMV, human cytomegalovirus.
Note: The base model included age, sex, and race. The full model additionally incorporated education, Townsend deprivation index, smoking status, and physical activity.

Table 3.S4. Association between age at onset of each herpes infection and subsequent Alzheimer's disease incidence

	HSV (N = 58,752)		VZV (N = 49,429)		EBV (N = 9,033)		CMV (N = 10,092)	
	Base	Full	Base	Full	Base	Full	Base	Full
Age of onset: HR (95% CI)								
No infection	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
<45	1.55 (0.98-2.45)	1.32 (0.83-2.10)	1.16 (0.62-2.14)	1.28 (0.65-2.50)	1.06 (0.32-3.52)	1.06 (0.32-3.53)	0.56 (0.14-2.31)	0.56 (0.14-2.32)
45-55	1.82 (1.21-2.74)	1.62 (1.07-2.44)	1.15 (0.64-2.07)	1.27 (0.67-2.43)	1.77 (0.48-6.47)	1.79 (0.48-6.65)	1.09 (0.39-3.00)	1.08 (0.39-3.01)
55-65	1.72 (1.16-2.56)	1.46 (0.98-2.18)	1.00 (0.56-1.78)	1.05 (0.56-1.97)	2.28 (0.89-5.80)	2.04 (0.80-5.25)	1.09 (0.71-1.67)	1.01 (0.66-1.57)
>65	1.44 (0.96-2.17)	1.21 (0.81-1.82)	0.84 (0.47-1.50)	0.89 (0.47-1.68)	1.98 (0.79-4.95)	1.54 (0.61-3.88)	0.97 (0.65-1.45)	0.81 (0.54-1.23)
Trend test: p-value								
Linear trend	0.64	0.74	0.01	0.01	0.05	0.22	0.90	0.50

Abbreviation: HSV, herpes simplex virus; VZV, varicella-zoster virus; EBV, Epstein-Barr virus; CMV, human cytomegalovirus.
Note: The base model included age, sex, and race. The full model additionally incorporated education, Townsend deprivation index, smoking status, and physical activity.

Table 3.S5. Association between age at onset of each herpes infection and subsequent dementia incidence, excluding participants who did not have any serological measurement but had a herpes infection diagnosis after baseline

	HSV (N = 58,752)		VZV (N = 49,429)		EBV (N = 9,033)		CMV (N = 10,092)	
	Base	Full	Base	Full	Base	Full	Base	Full
Age of onset: HR (95% CI)								
No infection	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
<45	1.77 (1.15-2.73)	1.57 (1.00-2.44)	1.44 (0.79-2.65)	1.57 (0.81-3.05)	1.01 (0.37-2.77)	1.14 (0.40-3.28)	1.32 (0.53-3.30)	1.32 (0.53-3.30)
45-55	1.81 (1.22-2.68)	1.67 (1.11-2.50)	1.24 (0.69-2.23)	1.39 (0.73-2.63)	0.90 (0.27-3.05)	0.96 (0.27-3.40)	0.94 (0.35-2.56)	0.93 (0.34-2.55)
55-65	1.78 (1.22-2.62)	1.57 (1.06-2.33)	1.16 (0.65-2.06)	1.23 (0.65-2.31)	1.55 (0.70-3.45)	1.55 (0.66-3.68)	1.07 (0.71-1.62)	0.98 (0.64-1.49)
>65	1.41 (0.94-2.12)	1.17 (0.77-1.79)	0.96 (0.53-1.73)	1.01 (0.53-1.92)	1.54 (0.70-3.40)	1.42 (0.60-3.36)	1.05 (0.71-1.55)	0.89 (0.59-1.33)
Trend test: p-value								
Linear trend	0.54	0.76	0.02	0.01	0.14	0.31	0.80	0.56

Abbreviation: HSV, herpes simplex virus; VZV, varicella-zoster virus; EBV, Epstein-Barr virus; CMV, human cytomegalovirus.
Note: The base model included age, sex, and race. The full model additionally incorporated education, Townsend deprivation index, smoking status, and physical activity.

Table 3.S6. Association between continuous age at onset (each decade) of each herpes infection and subsequent dementia incidence

	HSV (N = 58,752)		VZV (N = 49,429)		EBV (N = 9,033)		CMV (N = 10,092)	
	Base	Full	Base	Full	Base	Full	Base	Full
Age of onset: HR (95% CI)								
No infection	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Each 10-year earlier onset	1.04 (0.96-1.11)	1.04 (0.97-1.13)	1.08 (1.02-1.15)	1.09 (1.02-1.16)	0.85 (0.70-1.05)	0.89 (0.72-1.09)	0.92 (0.68-1.26)	0.96 (0.71-1.30)

Abbreviation: HSV, herpes simplex virus; VZV, varicella-zoster virus; EBV, Epstein-Barr virus; CMV, human cytomegalovirus.
Note: The base model included age, sex, and race. The full model additionally incorporated education, Townsend deprivation index, smoking status, and physical activity.

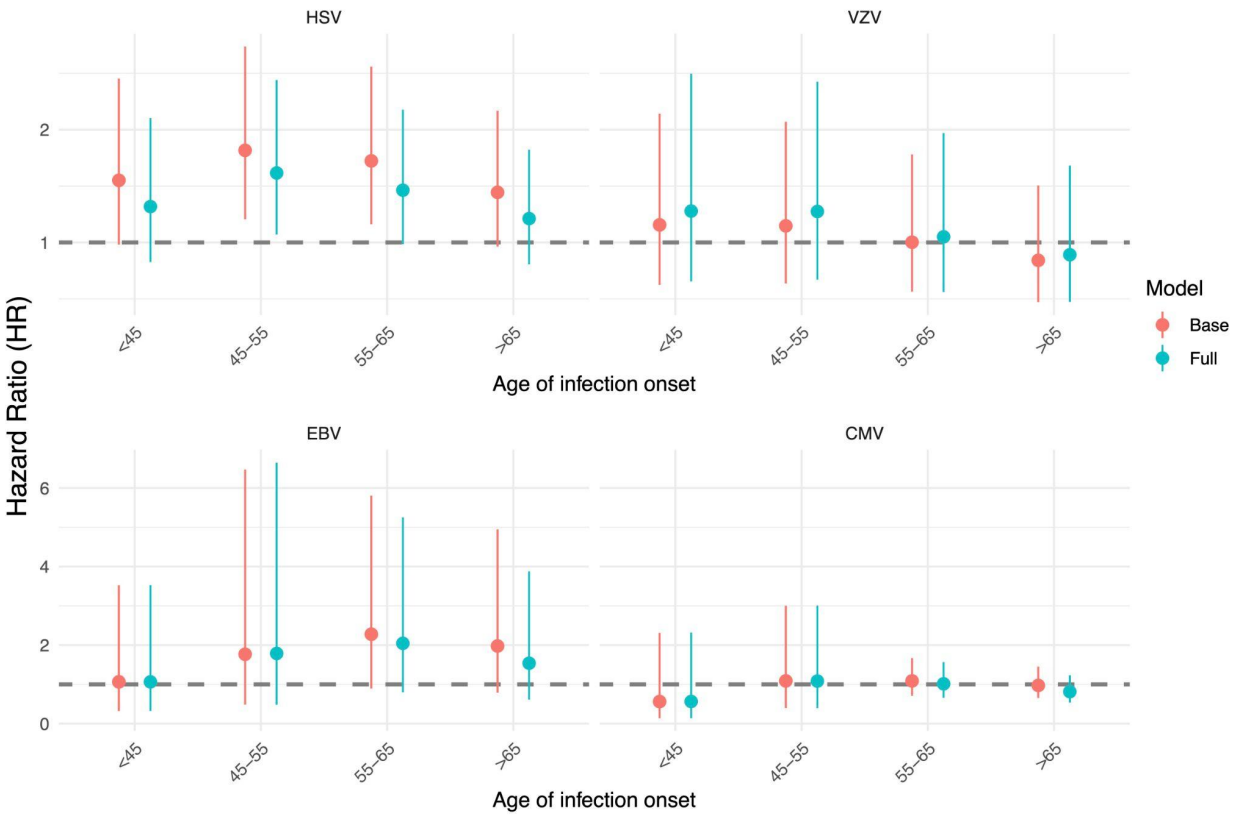


Figure 3.S1. Association between age at onset of each herpes infection and subsequent Alzheimer's disease dementia incidence

Abbreviation: HSV, herpes simplex virus; VZV, varicella-zoster virus; EBV, Epstein-Barr virus; CMV, human cytomegalovirus.
Note: The base model included age, sex, and race. The full model additionally incorporated education, Townsend deprivation index, smoking status, and physical activity.

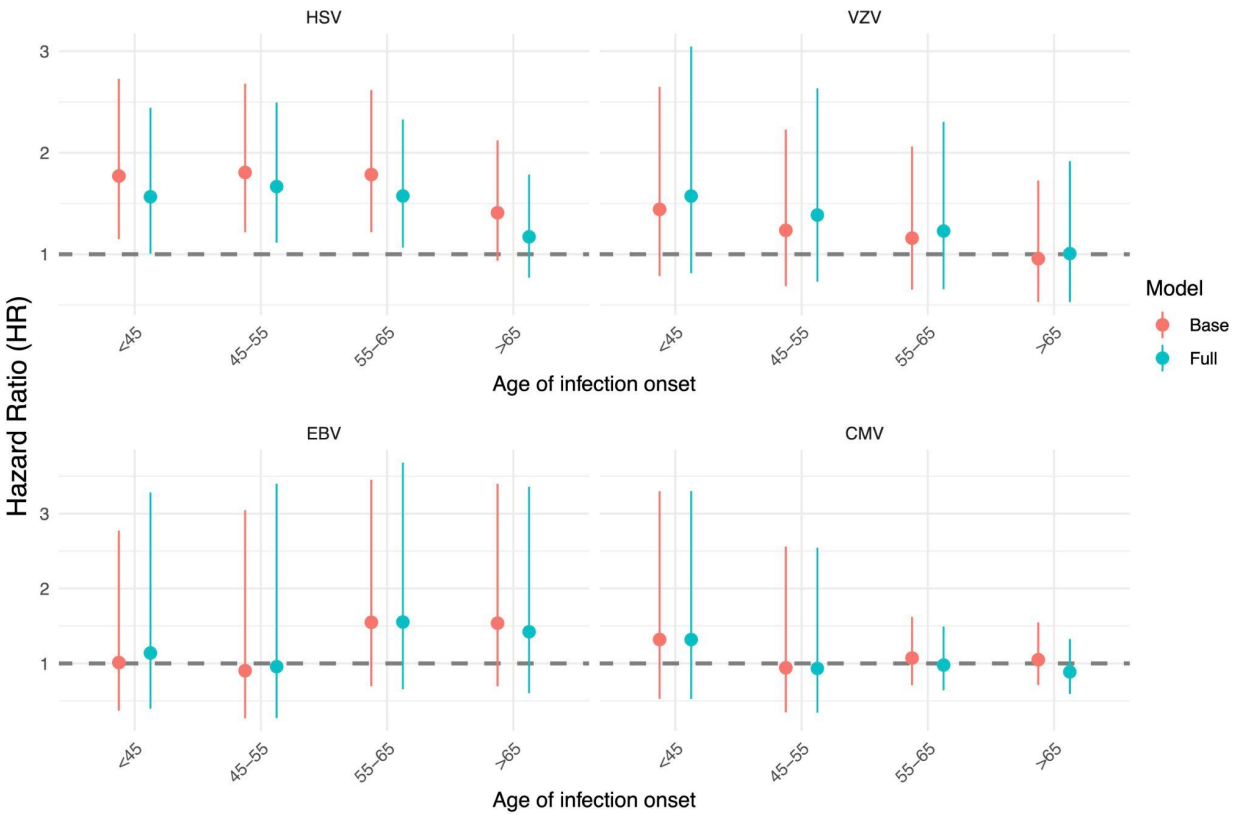


Figure 3.S2. Association between age at onset of each herpes infection and subsequent dementia incidence, excluding participants who did not have any serological measurement but had a herpes infection diagnosis after baseline

Abbreviation: HSV, herpes simplex virus; VZV, varicella-zoster virus; EBV, Epstein-Barr virus; CMV, human cytomegalovirus.
Note: The base model included age, sex, and race. The full model additionally incorporated education, Townsend deprivation index, smoking status, and physical activity.

References

1. Arvanitakis Z, Shah RC, Bennett DA. Diagnosis and Management of Dementia: Review. *JAMA*. 2019;322(16):1589-1599. doi:10.1001/jama.2019.4782
2. Elahi FM, Miller BL. A clinicopathological approach to the diagnosis of dementia. *Nat Rev Neurol*. 2017;13(8):457-476. doi:10.1038/nrneurol.2017.96
3. Nichols E, Steinmetz JD, Vollset SE, et al. 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*. 2022;7(2):e105-e125. doi:10.1016/S2468-2667(21)00249-8
4. Sevigny J, Chiao P, Bussière T, et al. The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. *Nature*. 2016;537(7618):50-56. doi:10.1038/nature19323
5. van Dyck CH, Swanson CJ, Aisen P, et al. Lecanemab in Early Alzheimer's Disease. *N Engl J Med*. 2023;388(1):9-21. doi:10.1056/NEJMoa2212948
6. Sims JR, Zimmer JA, Evans CD, et al. Donanemab in Early Symptomatic Alzheimer Disease: The TRAILBLAZER-ALZ 2 Randomized Clinical Trial. *JAMA*. 2023;330(6):512-527. doi:10.1001/jama.2023.13239
7. Manly JJ, Glymour MM. What the Aducanumab Approval Reveals About Alzheimer Disease Research. *JAMA Neurology*. 2021;78(11):1305-1306. doi:10.1001/jamaneurol.2021.3404
8. Livingston G, Huntley J, Liu KY, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. *Lancet*. 2024;404(10452):572-628. doi:10.1016/S0140-6736(24)01296-0

9. Itzhaki RF, Lathe R, Balin BJ, et al. Microbes and Alzheimer's Disease. *J Alzheimers Dis*. 2016;51(4):979-984. doi:10.3233/JAD-160152

10. Bertuccelli M, Ciringione L, Rubega M, Bisiacchi P, Masiero S, Del Felice A. Cognitive impairment in people with previous COVID-19 infection: A scoping review. *Cortex*. 2022;154:212-230. doi:10.1016/j.cortex.2022.06.002

11. Douaud G, Lee S, Alfaro-Almagro F, et al. SARS-CoV-2 is associated with changes in brain structure in UK Biobank. *Nature*. 2022;604(7907):697-707. doi:10.1038/s41586-022-04569-5

12. Warren-Gash C, Forbes HJ, Williamson E, et al. Human herpesvirus infections and dementia or mild cognitive impairment: a systematic review and meta-analysis. *Sci Rep*. 2019;9(1):4743. doi:10.1038/s41598-019-41218-w

13. Barnes DE, Zhou J, Walker RL, et al. Development and Validation of eRADAR: A Tool Using EHR Data to Detect Unrecognized Dementia. *Journal of the American Geriatrics Society*. 2020;68(1):103-111. doi:10.1111/jgs.16182

14. McCoy Jr. TH, Han L, Pellegrini AM, Tanzi RE, Berretta S, Perlis RH. Stratifying risk for dementia onset using large-scale electronic health record data: A retrospective cohort study. *Alzheimer's & Dementia*. 2020;16(3):531-540. doi:10.1016/j.jalz.2019.09.084

15. Wang J, Buto P, Ackley SF, et al. Association between cancer and dementia risk in the UK Biobank: evidence of diagnostic bias. *Eur J Epidemiol*. 2023;38(10):1069-1079. doi:10.1007/s10654-023-01036-x

16. Ferguson EL, Zimmerman SC, Jiang C, et al. Low- and High-Density Lipoprotein Cholesterol and Dementia Risk Over 17 Years of Follow-up Among Members of a

Large Health Care Plan. *Neurology*. 2023;101(21):e2172-e2184. doi:10.1212/WNL.0000000000207876

17. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012;13(6):395-405. doi:10.1038/nrg3208

18. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc*. 2017;24(1):198-208. doi:10.1093/jamia/ocw042

19. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ*. 2018;361:k1479. doi:10.1136/bmj.k1479

20. Shao Y, Zeng QT, Chen KK, Shutes-David A, Thielke SM, Tsuang DW. Detection of probable dementia cases in undiagnosed patients using structured and unstructured electronic health records. *BMC Med Inform Decis Mak*. 2019;19(1):128. doi:10.1186/s12911-019-0846-4

21. Harton J, Mitra N, Hubbard RA. Informative presence bias in analyses of electronic health records-derived data: a cautionary note. *Journal of the American Medical Informatics Association*. 2022;29(7):1191-1199. doi:10.1093/jamia/ocac050

22. McGee G, Haneuse S, Coull BA, Weisskopf MG, Rotem RS. On the Nature of Informative Presence Bias in Analyses of Electronic Health Records. *Epidemiology*. 2022;33(1):105-113. doi:10.1097/EDE.0000000000001432

23. Goldstein BA, Phelan M, Pagidipati NJ, Peskoe SB. How and when informative visit processes can bias inference when using electronic health records data for clinical research. *Journal of the American Medical Informatics Association*. 2019;26(12):1609-1617. doi:10.1093/jamia/ocz148

24. Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32(1-2):51-63. doi:10.1016/0021-9681(79)90012-2

25. Lipsitch M, Tchetgen ET, Cohen T. Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies. *Epidemiology*. 2010;21(3):383-388. doi:10.1097/EDE.0b013e3181d61eeb

26. Allen N, Sudlow C, Downey P, et al. UK Biobank: Current status and what it means for epidemiology. *Health Policy and Technology*. 2012;1(3):123-126. doi:10.1016/j.hlpt.2012.07.003

27. The “All of Us” Research Program. *New England Journal of Medicine*. 2019;381(7):668-676. doi:10.1056/NEJMsr1809937

28. Chatterjee S, Peters SAE, Woodward M, et al. Type 2 Diabetes as a Risk Factor for Dementia in Women Compared With Men: A Pooled Analysis of 2.3 Million People Comprising More Than 100,000 Cases of Dementia. *Diabetes Care*. 2016;39(2):300-307. doi:10.2337/dc15-1588

29. Cherbuin N, Kim S, Anstey KJ. Dementia risk estimates associated with measures of depression: a systematic review and meta-analysis. *BMJ Open*. 2015;5(12):e008853. doi:10.1136/bmjopen-2015-008853

30. Lennon MJ, Makkar SR, Crawford JD, Sachdev PS. Midlife Hypertension and Alzheimer's Disease: A Systematic Review and Meta-Analysis. *J Alzheimers Dis*. 2019;71(1):307-316. doi:10.3233/JAD-190474

31. Mawanda F, Wallace RB, McCoy K, Abrams TE. Systemic and localized extra-central nervous system bacterial infections and the risk of dementia among US veterans: A retrospective cohort study. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*. 2016;4:109-117. doi:10.1016/j.dadm.2016.08.004

32. Deb P, Norton EC. Modeling Health Care Expenditures and Use. *Annual Review of Public Health*. 2018;39(Volume 39, 2018):489-505. doi:10.1146/annurev-publhealth-040617-013517

33. Imrey PB. Limitations of Meta-analyses of Studies With High Heterogeneity. *JAMA Network Open*. 2020;3(1):e1919325. doi:10.1001/jamanetworkopen.2019.19325

34. Weiskopf NG, Rusanov A, Weng C. Sick Patients Have More Data: The Non-Random Completeness of Electronic Health Records. *AMIA Annu Symp Proc*. 2013;2013:1472-1477.

35. Goldstein BA, Bhavsar NA, Phelan M, Pencina MJ. Controlling for Informed Presence Bias Due to the Number of Health Encounters in an Electronic Health Record. *American Journal of Epidemiology*. 2016;184(11):847-855. doi:10.1093/aje/kww112

36. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA*

Internal Medicine. 2018;178(11):1544-1547. doi:10.1001/jamainternmed.2018.3763

37. Al-Sahab B, Leviton A, Loddenkemper T, Paneth N, Zhang B. Biases in Electronic Health Records Data for Generating Real-World Evidence: An Overview. *J Healthc Inform Res*. 2024;8(1):121-139. doi:10.1007/s41666-023-00153-2

38. Beesley LJ, Mukherjee B. Statistical inference for association studies using electronic health records: handling both selection bias and outcome misclassification. *Biometrics*. 2022;78(1):214-226. doi:10.1111/biom.13400

39. Robinson L, Tang E, Taylor JP. Dementia: timely diagnosis and early intervention. *BMJ*. 2015;350:h3029. doi:10.1136/bmj.h3029

40. Bradford A, Kunik ME, Schulz P, Williams SP, Singh H. Missed and Delayed Diagnosis of Dementia in Primary Care: Prevalence and Contributing Factors. *Alzheimer Disease & Associated Disorders*. 2009;23(4):306-314. doi:10.1097/WAD.0b013e3181a6bebc

41. Maimaris W, Paty J, Perel P, et al. The Influence of Health Systems on Hypertension Awareness, Treatment, and Control: A Systematic Literature Review. *PLOS Medicine*. 2013;10(7):e1001490. doi:10.1371/journal.pmed.1001490

42. Chen TK, Knicely DH, Grams ME. Chronic Kidney Disease Diagnosis and Management: A Review. *JAMA*. 2019;322(13):1294-1304. doi:10.1001/jama.2019.14745

43. Surks MI, Ortiz E, Daniels GH, et al. Subclinical Thyroid DiseaseScientific Review and Guidelines for Diagnosis and Management. *JAMA*. 2004;291(2):228-238. doi:10.1001/jama.291.2.228

44. Harding BN, Floyd JS, Scherrer JF, et al. Methods to identify dementia in the electronic health record: Comparing cognitive test scores with dementia algorithms. *Healthcare*. 2020;8(2):100430. doi:10.1016/j.hjdsi.2020.100430

45. Koch T, Iliffe S. Dementia diagnosis and management: a narrative review of changing practice. *Br J Gen Pract*. 2011;61(589):e513-e525. doi:10.3399/bjgp11X588493

46. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology*. 2017;186(9):1026-1034. doi:10.1093/aje/kwx246

47. Wang J, Ferguson EL, Buto P, et al. Sociodemographic, health-related, and clinical characteristics and their associations with mortality among All of Us participants compared with the United States general population. *medRxiv*. Published online 2024:2024-11.

48. Pham TM, Petersen I, Walters K, et al. Trends in dementia diagnosis rates in UK ethnic groups: analysis of UK primary care data. *Clin Epidemiol*. 2018;10:949-960. doi:10.2147/CLEP.S152647

49. Olfson M, Blanco C, Marcus SC. Treatment of Adult Depression in the United States. *JAMA Internal Medicine*. 2016;176(10):1482-1491. doi:10.1001/jamainternmed.2016.5057

50. Wang J, Chen R, Hayes-Larson E, et al. Comparison of dementia incidence and risk factor associations in All of Us against the nationally representative Health and Retirement Study. *Alzheimer's & Dementia*. 2024;20(S7):e091815. doi:10.1002/alz.091815

51. Rogers JP, Lewis G. Neuropsychiatric sequelae of COVID-19: long-lasting, but not uniform. *The Lancet Psychiatry*. 2022;9(10):762-763. doi:10.1016/S2215-0366(22)00302-9

52. Adjaye-Gbewonyo D, Vahratian A, Perrine CG, Bertolli J. Long COVID in Adults: United States, 2022. *NCHS Data Brief*. 2023;(480):1-8.

53. Pyne JD, Brickman AM. The Impact of the COVID-19 Pandemic on Dementia Risk: Potential Pathways to Cognitive Decline. *Neurodegenerative Diseases*. 2021;21(1-2):1-23. doi:10.1159/000518581

54. Toniolo S, Scarioni M, Di Lorenzo F, et al. Dementia and COVID-19, a Bidirectional Liaison: Risk Factors, Biomarkers, and Optimal Health Care. *Journal of Alzheimer's Disease*. 2021;82(3):883-898. doi:10.3233/JAD-210335

55. Taquet M, Sillett R, Zhu L, et al. Neurological and psychiatric risk trajectories after SARS-CoV-2 infection: an analysis of 2-year retrospective cohort studies including 1 284 437 patients. *The Lancet Psychiatry*. 2022;9(10):815-827. doi:10.1016/S2215-0366(22)00260-7

56. Jasti M, Nalleballe K, Dandu V, Onteddu S. A review of pathophysiology and neuropsychiatric manifestations of COVID-19. *J Neurol*. 2021;268(6):2007-2012. doi:10.1007/s00415-020-09950-w

57. Whittaker C, Walker PGT, Alhaffar M, et al. Under-reporting of deaths limits our understanding of true burden of covid-19. *BMJ*. 2021;375:n2239. doi:10.1136/bmj.n2239

58. Wu SL, Mertens AN, Crider YS, et al. Substantial underestimation of SARS-CoV-2 infection in the United States. *Nat Commun*. 2020;11(1):4507. doi:10.1038/s41467-020-18272-4

59. Geneva: World Health Organization. WHO COVID-19 Dashboard. Accessed November 6, 2024. https://covid19.who.int/info

60. Miller KL, Alfaro-Almagro F, Bangerter NK, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci*. 2016;19(11):1523-1536. doi:10.1038/nn.4393

61. Alfaro-Almagro F, Jenkinson M, Bangerter NK, et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage*. 2018;166:400-424. doi:10.1016/j.neuroimage.2017.10.034

62. Wilkinson T, Schnier C, Bush K, et al. Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. *Eur J Epidemiol*. 2019;34(6):557-565. doi:10.1007/s10654-019-00499-1

63. Wang J, Hill-Jarrett T, Buto P, et al. Comparison of approaches to control for intracranial volume in research on the association of brain volumes with cognitive outcomes. *Human Brain Mapping*. 2024;45(4):e26633. doi:10.1002/hbm.26633

64. Jack CR, Wiste HJ, Weigand SD, et al. Different definitions of neurodegeneration produce similar amyloid/neurodegeneration biomarker group findings. *Brain*. 2015;138(Pt 12):3747-3759. doi:10.1093/brain/awv283

65. Cohen K, Ren S, Heath K, et al. Risk of persistent and new clinical sequelae among adults aged 65 years and older during the post-acute phase of SARS-CoV-2 infection: retrospective cohort study. *BMJ*. 2022;376:e068414. doi:10.1136/bmj-2021-068414

66. Demurtas J, Schoene D, Torbahn G, et al. Physical Activity and Exercise in Mild Cognitive Impairment and Dementia: An Umbrella Review of Intervention and Observational Studies. *Journal of the American Medical Directors Association*. 2020;21(10):1415-1422.e6. doi:10.1016/j.jamda.2020.08.031

67. Rahman MA, Islam K, Rahman S, Alamin M. Neurobiochemical Cross-talk Between COVID-19 and Alzheimer's Disease. *Mol Neurobiol*. 2021;58(3):1017-1023. doi:10.1007/s12035-020-02177-w

68. Charnley M, Islam S, Bindra GK, et al. Neurotoxic amyloidogenic peptides in the proteome of SARS-COV2: potential implications for neurological symptoms in COVID-19. *Nat Commun*. 2022;13(1):3387. doi:10.1038/s41467-022-30932-1

69. Nalbandian A, Sehgal K, Gupta A, et al. Post-acute COVID-19 syndrome. *Nat Med*. 2021;27(4):601-615. doi:10.1038/s41591-021-01283-z

70. Zhou Y, Xu J, Hou Y, et al. Network medicine links SARS-CoV-2/COVID-19 infection to brain microvascular injury and neuroinflammation in dementia-like cognitive impairment. *Alzheimer's Research & Therapy*. 2021;13(1):110. doi:10.1186/s13195-021-00850-3

71. Solomon IH, Normandin E, Bhattacharyya S, et al. Neuropathological Features of Covid-19. *New England Journal of Medicine*. 2020;383(10):989-992. doi:10.1056/NEJMc2019373

72. South AM, Diz DI, Chappell MC. COVID-19, ACE2, and the cardiovascular consequences. *American Journal of Physiology-Heart and Circulatory Physiology*. 2020;318(5):H1084-H1090. doi:10.1152/ajpheart.00217.2020

73. Green R, Mayilsamy K, McGill AR, et al. SARS-CoV-2 infection increases the gene expression profile for Alzheimer's disease risk. *Molecular Therapy - Methods & Clinical Development*. 2022;27:217-229. doi:10.1016/j.omtm.2022.09.007

74. Peng S, Roth AR. Social Isolation and Loneliness Before and During the COVID-19 Pandemic: A Longitudinal Study of U.S. Adults Older Than 50. *The Journals of Gerontology: Series B*. 2022;77(7):e185-e190. doi:10.1093/geronb/gbab068

75. Hall G, Laddu DR, Phillips SA, Lavie CJ, Arena R. A tale of two pandemics: How will COVID-19 and global trends in physical inactivity and sedentary behavior affect one another? *Prog Cardiovasc Dis*. 2021;64:108-110. doi:10.1016/j.pcad.2020.04.005

76. Fountoulakis KN, Apostolidou MK, Atsiova MB, et al. Self-reported changes in anxiety, depression and suicidality during the COVID-19 lockdown in Greece. *Journal of Affective Disorders*. 2021;279:624-629. doi:10.1016/j.jad.2020.10.061

77. Zhou S, Wei T, Liu X, et al. Causal effects of COVID-19 on structural changes in specific brain regions: a Mendelian randomization study. *BMC Medicine*. 2023;21(1):261. doi:10.1186/s12916-023-02952-1

78. Vigasova D, Nemergut M, Liskova B, Damborsky J. Multi-pathogen infections and Alzheimer's disease. *Microb Cell Fact*. 2021;20(1):25. doi:10.1186/s12934-021-01520-7

79. Olsson J, Kok E, Adolfsson R, Lövheim H, Elgh F. Herpes virus seroepidemiology in the adult Swedish population. *Immunity & Ageing*. 2017;14(1):10. doi:10.1186/s12979-017-0093-4

80. Hernandez-Ruiz V, Letenneur L, Fülöp T, et al. Infectious diseases and cognition: do we have to worry? *Neurol Sci*. 2022;43(11):6215-6224. doi:10.1007/s10072-022-06280-9

81. Linard M, Baillet M, Letenneur L, et al. Herpes simplex virus, early neuroimaging markers and incidence of Alzheimer's disease. *Transl Psychiatry*. 2021;11(1):414. doi:10.1038/s41398-021-01532-2

82. Roos KL. Encephalitis. *Handb Clin Neurol*. 2014;121:1377-1381. doi:10.1016/B978-0-7020-4088-7.00094-8

83. Itzhaki RF. Herpes and Alzheimer's Disease: Subversion in the Central Nervous System and How It Might Be Halted. *J Alzheimers Dis*. 2016;54(4):1273-1281. doi:10.3233/JAD-160607

84. Wozniak MA, Mee AP, Itzhaki RF. Herpes simplex virus type 1 DNA is located within Alzheimer's disease amyloid plaques. *J Pathol*. 2009;217(1):131-138. doi:10.1002/path.2449

85. Fulop T, Witkowski JM, Bourgade K, et al. Can an Infection Hypothesis Explain the Beta Amyloid Hypothesis of Alzheimer's Disease? *Front Aging Neurosci*. 2018;10:224. doi:10.3389/fnagi.2018.00224

86. Whitson HE, Colton C, El Khoury J, et al. Infection and inflammation: New perspectives on Alzheimer's disease. *Brain, Behavior, & Immunity - Health*. 2022;22:100462. doi:10.1016/j.bbih.2022.100462

87. Mindel A, Marks C. Psychological Symptoms Associated with Genital Herpes Virus Infections. *CNS Drugs*. 2005;19(4):303-312. doi:10.2165/00023210-200519040-00003

88. Coughlin SS. Anxiety and Depression: Linkages with Viral Diseases. *Public Health Rev*. 2012;34(2):7, 92. doi:10.1007/BF03391675

89. Lin WR, Wozniak MA, Cooper RJ, Wilcock GK, Itzhaki RF. Herpesviruses in brain and Alzheimer's disease. *The Journal of Pathology*. 2002;197(3):395-402. doi:10.1002/path.1127

90. Levine KS, Leonard HL, Blauwendraat C, et al. Virus exposure and neurodegenerative disease risk across national biobanks. *Neuron*. 2023;111(7):1086-1093.e2. doi:10.1016/j.neuron.2022.12.029

91. Sipilä PN, Heikkilä N, Lindbohm JV, et al. Hospital-treated infectious diseases and the risk of dementia: a large, multicohort, observational study with a replication cohort. *The Lancet Infectious Diseases*. 2021;21(11):1557-1567. doi:10.1016/S1473-3099(21)00144-4

92. Readhead B, Haure-Mirande JV, Funk CC, et al. Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus. *Neuron*. 2018;99(1):64-82.e7. doi:10.1016/j.neuron.2018.05.023

93. Elhalag RH, Motawea KR, Talat NE, et al. Herpes simplex virus infection and the risk of dementia: a systematic review and meta-analysis. *Ann Med Surg (Lond)*. 2023;85(10):5060-5074. doi:10.1097/MS9.0000000000000951

94. Chen VCH, Wu SI, Huang KY, et al. Herpes Zoster and Dementia: A Nationwide Population-Based Cohort Study. *J Clin Psychiatry*. 2018;79(1):16m11312. doi:10.4088/JCP.16m11312

95. Lopatko Lindman K, Hemmingsson ES, Weidung B, et al. Herpesvirus infections, antiviral treatment, and the risk of dementia-a registry-based cohort study in Sweden. *Alzheimers Dement (N Y)*. 2021;7(1):e12119. doi:10.1002/trc2.12119

96. Mekli K, Lophatananon A, Cant R, et al. Investigation of the association between the antibody responses to neurotropic viruses and dementia outcomes in the UK Biobank. *PLOS ONE*. 2022;17(10):e0274872. doi:10.1371/journal.pone.0274872

97. Carbone I, Lazzarotto T, Ianni M, et al. Herpes virus in Alzheimer's disease: relation to progression of the disease. *Neurobiology of Aging*. 2014;35(1):122-129. doi:10.1016/j.neurobiolaging.2013.06.024

98. Lövheim H, Olsson J, Weidung B, et al. Interaction between Cytomegalovirus and Herpes Simplex Virus Type 1 Associated with the Risk of Alzheimer's Disease Development. *Journal of Alzheimer's Disease*. 2018;61(3):939-945. doi:10.3233/JAD-161305

99. Shi R, Yu S, Larbi A, Pin Ng T, Lu Y. Specific and cumulative infection burden and mild cognitive impairment and dementia: A population-based study. *Brain, Behavior, and Immunity*. 2024;121:155-164. doi:10.1016/j.bbi.2024.07.026

100. Vestin E, Boström G, Olsson J, et al. Herpes Simplex Viral Infection Doubles the Risk of Dementia in a Contemporary Cohort of Older Adults: A Prospective Study. *J Alzheimers Dis*. 2024;97(4):1841-1850. doi:10.3233/JAD-230718

101. Barnes LL, Capuano AW, Aiello AE, et al. Cytomegalovirus Infection and Risk of Alzheimer Disease in Older Black and White Individuals. *The Journal of Infectious Diseases*. 2015;211(2):230-237. doi:10.1093/infdis/jiu437

102. Ma X, Liao Z, Tan H, et al. The association between cytomegalovirus infection and neurodegenerative diseases: a prospective cohort using UK Biobank data. *EClinicalMedicine*. 2024;74:102757. doi:10.1016/j.eclinm.2024.102757

103. Singh A, Preiksaitis J, Ferenczy A, Romanowski B. The laboratory diagnosis of herpes simplex virus infections. *Can J Infect Dis Med Microbiol*. 2005;16(2):92-98.

104. Mentzer AJ, Brenner N, Allen N, et al. Identification of host–pathogen-disease relationships using a scalable multiplex serology platform in UK Biobank. *Nat Commun*. 2022;13:1818. doi:10.1038/s41467-022-29307-3

105. Townsend P, Phillimore P, Beattie A. *Health and Deprivation: Inequality and the North*. Routledge; 1988.

106. Lycke E, Norrby R, Roos BE. A Serological Study on Mentally Ill Patients: With Particular Reference to the Prevalence of Herpes Virus Infections. *The British Journal of Psychiatry*. 1974;124(580):273-279. doi:10.1192/bjp.124.3.273

107. Ounanian A, Seigneurin JM, Guilbert B, Avrameas S, Renverez JC. Antibodies to viral antigens, xenoantigens, and autoantigens in alzheimer's disease. *Journal of Clinical Laboratory Analysis*. 1990;4(5):367-375. doi:10.1002/jcla.1860040510

108. Lin WR, Shang D, Itzhaki RF. Neurotropic viruses and Alzheimer disease. *Molecular and Chemical Neuropathology*. 1996;28(1):135-141. doi:10.1007/BF02815215

109. Hemling N, Röyttä M, Rinne J, et al. Herpesviruses in brains in Alzheimer's and Parkinson's diseases. *Annals of Neurology*. 2003;54(2):267-271. doi:10.1002/ana.10662

110. Mancuso R, Baglio F, Agostini S, et al. Relationship between herpes simplex virus-1-specific antibody titers and cortical brain damage in Alzheimer's disease and amnestic mild cognitive impairment. *Front Aging Neurosci*. 2014;6. doi:10.3389/fnagi.2014.00285

111. Hogestyn JM, Mock DJ, Mayer-Proschel M. Contributions of neurotropic human herpesviruses herpes simplex virus 1 and human herpesvirus 6 to neurodegenerative disease pathology. *Neural Regen Res*. 2018;13(2):211-221. doi:10.4103/1673-5374.226380

112. Itzhaki RF. Herpes simplex virus type 1 and Alzheimer's disease: increasing evidence for a major role of the virus. *Front Aging Neurosci*. 2014;6. doi:10.3389/fnagi.2014.00202

113. Wozniak MA, Shipley SJ, Combrinck M, Wilcock GK, Itzhaki RF. Productive herpes simplex virus in brain of elderly normal subjects and Alzheimer's disease patients. *Journal of Medical Virology*. 2005;75(2):300-306. doi:10.1002/jmv.20271

114. Wozniak MA, Itzhaki RF, Shipley SJ, Dobson CB. Herpes simplex virus infection causes cellular β-amyloid accumulation and secretase upregulation. *Neuroscience Letters*. 2007;429(2):95-100. doi:10.1016/j.neulet.2007.09.077

115. Álvarez G, Aldudo J, Alonso M, Santana S, Valdivieso F. Herpes simplex virus type 1 induces nuclear accumulation of hyperphosphorylated tau in neuronal cells. *Journal of Neuroscience Research*. 2012;90(5):1020-1029. doi:10.1002/jnr.23003

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Jingyuan Wang
5355F700AEC4490...
Author Signature

3/29/2025
Date