

Common infections and neuroimaging markers of dementia in three UK cohort studies

Rebecca E Green^a, Carole H Sudre^{abcd}, Charlotte Warren-Gash^c, Julia Butt^f, Tim Waterboer^f, the Insight 46 study team, Alun D Hughes^a, Jonathan M Schott^b, Marcus Richards^a, Nish Chaturvedi^a, Dylan M Williams^a

^a MRC Unit for Lifelong Health & Ageing at UCL, University College London, London, UK

^b Dementia Research Centre, UCL Queen Square Institute of Neurology, London, UK

^c School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

^d Department of Medical Physics and Biomedical Engineering, Centre for Medical Image Computing (CMIC), University College London (UCL), London, UK

^e Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK

^f Division of Infections and Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

Corresponding authors: Rebecca Green (rebecca.e.green@ucl.ac.uk), Dylan Williams (dylan.williams@ucl.ac.uk).

Abstract

Introduction: We aimed to investigate associations between common infections and neuroimaging markers of dementia risk (brain volume, hippocampal volume, white matter lesions) across three population-based studies.

Methods: We tested associations between serology measures (pathogen serostatus, cumulative burden, continuous antibody responses) and outcomes using linear regression, including adjustments for total intracranial volume and scanner/clinic information (basic model), age, sex, ethnicity, education, socioeconomic position, alcohol, BMI, and smoking (fully adjusted model). Interactions between serology measures and *APOE* genotype were tested. Findings were meta-analysed across cohorts ($N_{\text{main}}=2632$; $N_{\text{APOE-interaction}}=1810$).

Results: Seropositivity to JC virus associated with smaller brain volumes in basic models ($\beta=-3.89\text{ml}[-5.81,-1.97]$, $p_{\text{adjusted}}<0.05$); these were largely attenuated in fully adjusted models ($\beta=-1.59\text{ml}[-3.55,0.36]$, $p=0.11$). No other relationships were robust to multiple testing correction and sensitivity analyses, but several suggestive associations were observed.

Discussion: We did not find clear evidence for relationships between common infections and markers of dementia risk. Some suggestive findings warrant testing for replication.

Key words: multiplex serology, common infections, pathogen burden, dementia, Alzheimer's disease, cerebral small vessel disease

Abbreviations

AD = Alzheimer's disease

APOE = Apolipoprotein E

BaMoS = Bayesian Model Selection

BMI = Body mass index

CMV = Cytomegalovirus

EBV = Epstein-Barr virus

HHV = Human herpesvirus

HPV = Human papillomavirus

HSV = Herpes simplex virus

IgG = Immunoglobulin G

JC = John-Cunningham virus

KSHV = Kaposi's sarcoma-associated herpesviruses

MCV = Merkel Cell virus

NSHD = National Survey of Health and Development

PBI = Pathogen burden index

QC = Quality control

SABRE = Southall and Brent Revisited

UKB = UK Biobank

VZV = Varicella zoster virus

Introduction

Common infections have been associated with a variety of non-communicable diseases, with a role in dementia suspected¹. Despite primary infection typically occurring earlier in the life course, a number of pathogens are able to cause persistent infections; for example through chronic infection or by remaining latent and reactivating at later stages². As infections may be preventable or treatable, a better understanding of their role in dementia risk could inform our knowledge of high-risk groups and possible priorities for intervention.

While the extent to which common infections contribute to dementia risk is still unclear, multiple pathways have been hypothesised. Neurotropic pathogens – such as herpes simplex virus (HSV) – are able to directly infect cells in the central nervous system, possibly triggering amyloid- β pathology³, neuroinflammation, and neuronal loss⁴. Infection or cumulative infection burden could additionally drive systemic changes, including inflammation⁵, which have been implicated in dementia aetiology^{6,7}. Nevertheless, previous studies have primarily focussed on a few hypothesis-driven pathogens or pathogen families, and findings have been mixed⁸. This may be due to challenges in defining previous infection in the absence of large-scale serology data, confounding and, as the prodromal and clinical phases of dementia may affect immune competency, reverse directionality.

Examining relationships between multiple common infections and subclinical markers of dementia prospectively – across well-characterised population-based studies – could facilitate a more comprehensive understanding of their role in dementia risk. In addition, assessing whether pathogen exposure may interact with established risk or protective factors of dementia could provide insights on possible at-risk groups. For example, *APOE* genotypes – the strongest genetic risk (for *APOE* e4 carriers) and protective (for *APOE* e2 carriers) factors for late-onset dementia – are suggested to modify associations of specific infections or burden with dementia risk⁹⁻¹¹.

Recently, a validated large-scale multiplex serology panel was applied in a subset of the UK Biobank (UKB)¹², providing a valuable resource to assess pathogen-disease relationships. This serology panel simultaneously measures quantitative antibody responses against antigens of numerous pathogens selected due to their relevance to public health, including human herpesviruses, polyomaviruses, papillomaviruses, and *C.trachomatis*, *H.pylori* and *T.gondii*. Serostatus to these pathogens (i.e. whether or not an individual has been previously infected) can

be subsequently derived using seropositivity thresholds, providing a summary of individual infection history. We have since applied this panel to two additional population-based cohorts in the UK, the MRC National Survey of Health and Development (NSHD) ¹³ and Southall and Brent Revisited (SABRE) ¹⁴, permitting a wide range of pathogens to be investigated in relation to outcomes among several settings in parallel.

In this cross-cohort research, we investigated associations between antibodies to common infections and three established neuroimaging markers of brain structure and pathology with relevance to subclinical dementia – hippocampal volume, whole brain volume (as markers of neurodegenerative disease, particularly Alzheimer’s disease), and white matter hyperintensity volume (as a marker of cerebral small vessel disease). We first examined whether seropositivity and cumulative exposure to 17 pathogens associated with neuroimaging outcomes, and further tested for interactions of these associations with *APOE* genotype. We then assessed associations of antibody levels against each pathogen (indicative of recent infection or reactivation) with the same neuroimaging outcomes.

2. Methods

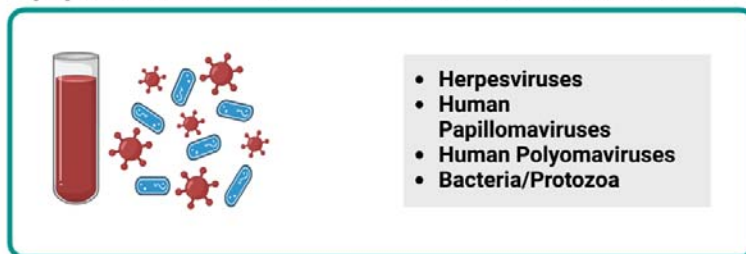
2.1 Study design

An overview of the study design is summarised in Figure 1. Analyses were conducted across three population-based cohorts: 1) Insight 46 ¹⁵, the neuroscience sub-study of the NSHD – a birth cohort originally consisting of 5362 individuals born in mainland Britain during one week of March 1946 ¹⁵, 2) SABRE, a tri-ethnic (European, South Asian, and African Caribbean) study consisting of 4972 individuals aged 40-69 at recruitment (1988-1991), randomly selected through GP practices and workplaces in two London boroughs, stratified by ethnicity, sex, and age ^{14,16}, and 3) UKB, a study comprised of >500,000 individuals approximately aged 40-69 at recruitment (2006-2010) ¹⁷. A summary of study participants can be found in Supplementary Figure 1.

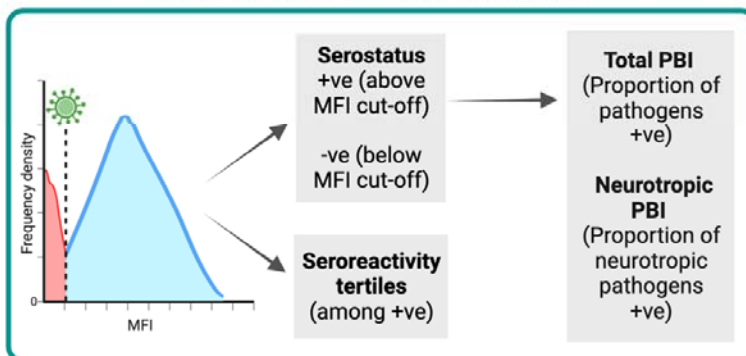
2.2 Ethical approval

All cohorts received ethical approval and all participants provided written informed consent. The NSHD and its sub-study Insight 46 received approval from the National Research Ethics Service Committee London (14/LO/1173), SABRE from the St Mary's Hospital Research Ethics Committee (07/H0712/109), and UKB from the NHS North West Research Ethics Committee (11/NW/0382).

Antigen seroreactivity values for 17 pathogens in three population-based studies: MRC NSHD, SABRE, and UKB

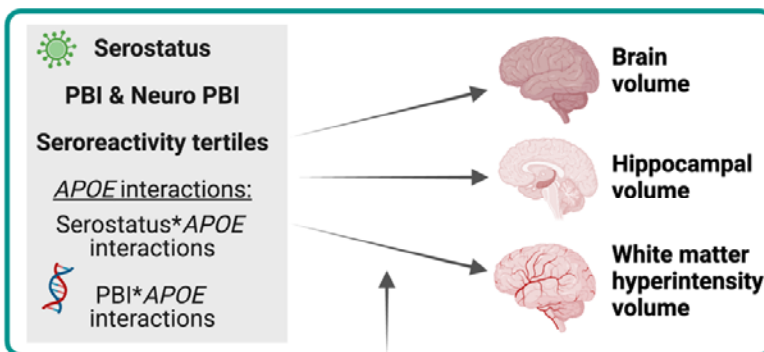


Step 1: Derive serology measures.



Step 2: Per-study analyses.

Test associations of serology measures with imaging outcomes, as well as interactions by *APOE* *e2* and *e4* genotype.



Model 1: Technical covariates (including total intracranial volume)

Model 2: Model 1 + age + sex (+ ethnicity for SABRE & UKB)

Model 3: Model 2 + alcohol intake + educational attainment + socioeconomic position + BMI + smoking

***APOE* interactions:** Models 1-3 + 10 genetic principal components.

Step 3: Meta-analyse findings using random-effects models.
(Main analyses total $N=2632$; *APOE* interactions total $N=1810$).

Figure 1. Study workflow.

Abbreviations: *APOE*=Apolipoprotein E genotype; MFI=median fluorescence intensity; MRC NSHD=Medical Research Council National Survey of Health and Development, PBI=pathogen

2.3 Exposure data

2.3.1 Multiplex serology

Serum immunoglobulin G (IgG) antibody levels against a selection of antigens were measured using a validated fluorescence bead-based multiplex serology platform developed at the German Cancer Research Center (DKFZ) in Heidelberg^{12,18}. Antibody responses (“seroreactivity”) were quantified in median fluorescence intensity units, and between one and six antigens were quantified per pathogen. A full list of pathogens assayed in each cohort are detailed in Supplementary Notes. In UKB, 21 pathogens were assayed among 9429 participants (baseline, 2006-10) and an additional 260 at follow-up only (Instance 1, 2012-13). An adaptation of the same multiplex panel assayed 18 pathogens among 1813 NSHD participants (Age 60-64 follow-up; 2006-10), and 1423 SABRE participants (Visit 2; 2008-11). Thirty-three participants in the NSHD and SABRE were excluded due to technical failures in the assaying (including pipetting errors, high background, and insufficient beadcounts).

2.3.2 Pathogen serostatus and burden scores

We derived binary serostatus variables indicating previous infection for each pathogen, using standardised median fluorescence intensity cut-offs for specific antigens, or antigen combinations where multiple antigens were quantified (see Supplementary Notes). Serostatus variables were derived in the UKB as reported previously¹². Only the 17 pathogens with relevant serology data available in all three cohorts were included in this study. To measure cumulative exposure to multiple pathogens, we additionally derived two pathogen burden index (PBI) scores: i) a proportion of positive serostatus values to all 17 pathogens (total PBI; sum of serostatus to 17 pathogens / 17), ii) a proportion of positive serostatus values to 11 neurotropic pathogens, including herpesviruses, *T.gondii*, and JC virus¹⁹ (neurotropic PBI; sum of serostatus to neurotropic pathogens / 11), which may hold more relevance to neurological phenotypes.

2.3.3 Antibody responses

Seroreactivity values formed a variety of non-normal distributions. Therefore, to group antibody responses for seroreactivity analyses, we derived tertiles of antibody titres against each antigen using recommended antigens by the German Cancer Research Center (see Supplementary Notes). Tertiles were derived in the full serology and neuroimaging sample that were seropositive for that antigen.

2.4 Neuroimaging outcomes

For NSHD, we used neuroimaging measures collected in the neuroscience sub-study Insight 46, when participants were aged 69-71 years (5-11 years after blood sampling used for serology assaying), using a single Biograph mMR 3 Tesla PET/MRI scanner (Siemens Healthcare)¹⁵. For SABRE, we used neuroimaging measures collected at the same time point as blood sampling used for serology assaying (Visit 2). Brain MRI was performed using General Electric Signa HDxt 1.5T, Signa HDx 1.5T, Signa EXCITE 1.5T, and Discovery MR750 3T scanners. Finally, for UKB, neuroimaging measures were collected at the first imaging visit (Instance 2; 1-13 years after blood sampling used for serology assaying). Details of the acquisition protocols for T1 and FLAIR sequences are available in Supplementary Notes.

In all cohorts, brain volume and hippocampal volume were quantified in-house using Geodesic Information Flows²⁰. White matter hyperintensities were derived from T2 FLAIR sequences using an unsupervised automatic algorithm Bayesian Model Selection (BaMoS) (see^{21,22} for further details), and were subsequently log_e transformed due to skewed distributions.

2.5 Genetic data for *APOE* analyses

2.5.1 Genetic principal components

We used directly genotyped data for genetic quality control and to derive genetic principal components. Details of genotyping can be found for NSHD²³, SABRE¹⁴, and UKB²⁴ elsewhere. Briefly, blood samples were collected at the age 53 visit for NSHD and baseline (visit 1) for SABRE, and participants were genotyped using the DrugDev array. Two genotyping arrays (BiLEVE Axiom and Affymetrix UK Biobank Axiom) were used to genotype UKB participants. NSHD and SABRE genetic quality control (QC) steps were performed in-house, and we used UKB provided QC criteria and genetic principal components (following the cohort's central QC) for UKB. For SABRE, QC steps were applied separately by self-reported ethnicity. QC for all cohorts included filtering for common biallelic autosomal variants that did not exhibit significant missingness or deviate from Hardy-Weinberg equilibrium. Samples that did not exhibit significant heterozygosity or missingness, with concordant genetic and self-reported sex, and that were unrelated and closely clustered with reference panel populations using genetic principal components, were retained (see Supplementary Notes for full QC parameters per cohort). Genetic principal components were generated in each study population following pruning and removal of variants in long range linkage disequilibrium regions²⁵. QC steps were performed using plink1.9 and plink2²⁶, and principal components for NSHD and

SABRE were generated using smartpca in EIGENSOFT²⁷.

2.5.2 Apolipoprotein E genotype

APOE genotypes were derived using rs7412 and rs429358 single nucleotide polymorphisms. We subsequently generated binary *APOE* e4 non-carrier/carrier and *APOE* e2 non carrier/carrier variables. Carriers were defined as heterozygous or homozygous for the e4 or e2 allele, respectively.

2.6 Covariates

Covariates included total intracranial volume and other technical variables related to neuroimaging or serology data collection (NSHD: blood clinic; SABRE: scanner; UKB: scanner co-ordinates and clinic), age at serology and neuroimaging (if not collected at the same time point), sex, ethnicity (SABRE: African-Caribbean, European, South Asian; UKB: Asian or Asian British, Black or Black British, Chinese, Mixed, Other, White), highest educational qualification (none; up to ordinary ('O') level or equivalent; advanced ('A') level or equivalent, or higher), socioeconomic position (NSHD and SABRE: current or last known occupation categorised into six groups according to the UK Registrar General; UKB: Townsend deprivation index quintiles), BMI (kg/m²), smoking status (never, ever), and alcohol intake (low <7 units per week, medium 7-14 units per week, high >14 units per week). A directed acyclic graph detailing the covariate choice can be found in Supplementary Notes.

2.7 Statistical analyses

Due to differences in exposure and outcome measurements across studies, data were not pooled and analyses were run in each individual study prior to meta-analysis. In each study, we imputed missing covariate data using multiple imputation by chained equations (10 iterations, 10 datasets; see Table 1 for details on missingness). Analyses were performed on each imputed data set and pooled using Rubin's rules²⁸. Regression assumptions were checked by examination of the residuals. We meta-analysed findings across studies (or subgroups for *APOE* interaction analyses, see 2.7.2) using random-effects models with a restricted maximum likelihood estimator. Regression coefficients for analyses with white matter lesion volumes as an outcome were multiplied by 100 to be transformed to sympercents²⁹. We used an exploratory framework for the present analyses, where multiple pathogens were tested in relation to outcomes. Due to these multiple tests, findings from meta-analyses (per outcome for serostatus, seroreactivity, and *APOE* interaction analyses) were corrected for the false discovery rate (FDR), using the

Benjamini-Hochberg procedure³⁰ with an alpha of 0.05. However, given that multiple testing correction increases the likelihood of false negative results, suggestive findings (at unadjusted $p < 0.05$) were additionally reported. Heterogeneity was assessed through Cochran's Q and I^2 statistics, and we defined significant heterogeneity where $I^2 > 50\%$ or Q-p value < 0.05 . All analyses were performed using R, version 3.6.2.

2.7.1 Primary serology analyses

We applied multiple linear regression models to test associations between serology variables (pathogen serostatus and seroreactivity tertiles) and neuroimaging outcomes. Model 1 was minimally adjusted, and included only total intracranial volume and any other technical covariates (NSHD: blood clinic; SABRE: scanner; UKB: scanner co-ordinates and clinic). Model 2 adjusted for model 1 covariates plus basic confounders: age, sex, and (for SABRE and UKB) ethnicity. Model 3 adjusted for model 1 and 2 covariates plus additional possible social, behavioural, and lifestyle confounders: BMI, highest educational qualification, socioeconomic position, smoking status, and alcohol intake. We note that primary exposure for many of these pathogens occurs earlier in life, and thus we included measures reflecting long-term characteristic differences or that may track over time. Statistical models included sampling weights where present. In seroreactivity analyses, tertiles were modelled as an ordinal variable, and pathogens with a seroprevalence over 5% in all studies were investigated (the two Human Papillomaviruses (HPV), and Kaposi's sarcoma-associated herpesvirus (KSHV) were therefore not studied). For pathogens where multiple antigens were quantified, we randomly selected one from the subset of recommended antigens (see Supplementary Notes) using the sample function in base R, with the other antigens investigated in sensitivity analyses.

2.7.2 APOE interaction analyses

To examine whether serostatus or pathogen burden relationships with outcomes differed by *APOE* genotype, we performed the same nested models listed above including a) *APOE* e4 carrier status, and b) *APOE* e2 carrier status as an interaction term. *APOE* e4 and *APOE* e2 analyses were conducted separately. All models were additionally adjusted for 10 genetic principal components, a recommended practice to minimise potential confounding by population structure, and for SABRE, analyses were further stratified by European and South Asian groups in the sample (sample size was not sufficient for inclusion of the African Caribbean group, $N=22$). Coefficients for interaction terms from each analysis were then meta-analysed using random-effects models as described in 2.7. Due to sample size (and hence power)

limitations, we did not test *APOE* interactions for seroreactivity tertile associations with outcomes, which include subsets of participants seropositive to the antigen of interest. Where suggestive evidence (unadjusted $p < 0.05$) of an interaction was observed, analyses were rerun stratified by *APOE* genotype.

2.8 Secondary and sensitivity analyses

We conducted several additional analyses, as follows: 1) we repeated serostatus analyses using alternate serostatus definitions for three pathogens (HPV16, *C.trachomatis*, *H.pylori*) that were available in the UKB; 2) for pathogens with multiple antigens assayed, we repeated seroreactivity analyses using alternate antigens, as detailed in 2.3.3; 3) we restricted samples to participants who were ≥ 60 years at the time of blood sampling used for serology assays, and performed the same analytical models as detailed above (NSHD N=438 – all aged 60-64; SABRE N=1240; UKB N=331); 4) we repeated all primary analyses excluding participants with a diagnosis of dementia or stroke (to include participants with only subclinical structural and cerebrovascular pathology only), or with known conditions affecting immune function (of which only diabetes was available in all cohorts, see Supplementary Notes), defined at or before the time of blood sampling used for serology assays (NSHD N=408; SABRE N=831; UKB N=901).

3. Results

3.1 Cohort characteristics and seroprevalence

Participant characteristics can be found in Table 1. Characteristics and seroprevalences for the full serology cohort and the serology and neuroimaging subset can be found in Supplementary Table 1. In the present study, we included participants with available serology measures and with data on at least one neuroimaging outcome (Total N: 2632; NSHD N=438, SABRE N=1259, UKB N=935). For *APOE* genotype interactions, available genetic data following QC were additionally required (Total N: 1810; NSHD N=413; SABRE = 314 (European), 279 (South Asian); UKB=804).

Table 1. Characteristics of participants with serology and neuroimaging data.

	NSHD	SABRE	UKB
N	438	1259	935
Female sex, N (%)	210 (47.9)	287 (22.8)	523 (55.9)
Age at serology in years, mean (SD)	63.3 (1.1)	69.6 (6.1)	55.3 (7.7)
Age at imaging in years, mean (SD)	70.7 (0.7)	69.6 (6.1)	63.8 (7.7)
Ethnicity, N (%)	White*	African Caribbean: 204 (16.2) European: 588 (46.7) South Asian: 467 (37.1)	Asian or Asian British: 13 (1.4) Black or Black British: 7 (0.7) Chinese: 4 (0.4) Mixed: 4 (0.4) Other: 4 (0.4) White: 902 (96.6) Missing N: 1
Brain volume in ml, mean (SD)	1141.7 (106.1)	1074.9 (109.0)	1150.5 (108.7)
Hippocampal volume in ml, mean (SD)	7.3 (0.7)	7.7 (0.8)	7.8 (0.7)
White matter hyperintensity volume in ml, median (IQR)	3.0 (1.6-6.8) Missing N: 14	4.1 (2.4-8.4) Missing N: 21	1.7 (1.0-3.6)
Total intracranial volume in ml, mean (SD)	1422.1 (132.2)	1361.9 (137.7)	1446.9 (135.6)
APOE e2 carrier (% carrier, in sample used for genetic analyses) †	68 (16.5)	82 (13.8)	153 (17)
APOE e4 carrier (% carrier, in sample used for genetic analyses) †	129 (31.2)	143 (24.1)	254 (27.4)
Body mass index in kg/m², mean (SD)	27.5 (4.1)	27.4 (4.5)	26.5 (4.2)
Alcohol intake, N (%)	Low	218 (50.2)	447 (48.1)
	Moderate	116 (26.7)	251 (27)
	Heavy	100 (23) Missing N: 4	173 (14.1) Missing N: 33

Ever smoker, N (%)		198 (45.9) Missing N: 7	533 (42.6) Missing N: 8	350 (37.5) Missing N: 2
Highest educational attainment, N (%)	None	87 (20.4)	315 (30.1)	68 (7.3)
	Up to ordinary ('O') level, or equivalent	105 (24.6)	210 (20.1)	202 (21.6)
	Advanced ('A') level or equivalent, or higher	234 (54.9) Missing N: 12	522 (49.9) Missing N: 212	664 (71.1) Missing N: 1
Socioeconomic position, N (%)	Professional	55 (12.6)	46 (3.7)	Townsend quintiles Missing N: 1
	Intermediate	227 (51.8)	258 (20.7)	
	Nonmanual skilled	91 (20.8)	187 (15)	
	Manual skilled	40 (9.1)	439 (35.3)	
	Partly skilled or Unskilled‡	25 (5.7)	315 (25.3) Missing N: 14	

Abbreviations: *APOE* = apolipoprotein E gene; NSHD = National Survey of Health and Development; SABRE = Southall and Brent Revisited; UKB = UK Biobank. *The NSHD is predominantly white British as they are representative of births in mainland Britain in 1946, which was prior to major immigration flows¹³. †Reported for sample passing genetic quality control that were used in the present analyses (NSHD N=413; SABRE N=593; UKB N=804). ‡Categories collapsed due to low cell counts for the purpose of this table only.

3.2 Serostatus and pathogen burden

Meta-analysed findings for serostatus and PBI analyses are displayed in Figure 2 and Table 2 (see Supplementary Figures 2-4 for per cohort findings, and Supplementary Tables 2-4 for full numerical results). To give magnitudes of associations context, a 1-year increase in age has been associated with a 16.5% increase in white matter lesions, 11.5ml smaller whole brain volume, and 0.08ml smaller hippocampal volume in the NSHD between 69-71 years of age³¹.

Seropositivity to JC virus was associated with smaller brain volume in model 1 (brain volume, $\beta = -3.89\text{ml}$ [-5.81, -1.97], $p = 7.3 \times 10^{-5}$, $p_{\text{FDR}} < 0.05$). We observed a notable attenuation following model 2 adjustments (age, sex, and ethnicity where relevant) and negligible further changes in the fully adjusted model ($\beta = -1.5\text{ml}$ [-3.55, 0.36], $p = 0.11$). We found no clear evidence of associations between other pathogens or PBI and outcomes after multiple testing correction, although four

further suggestive relationships were observed in fully adjusted models (seropositivity to cytomegalovirus (CMV) with larger brain volume, varicella zoster virus (VZV) and JC virus with smaller hippocampal volume, and HPV16 and more white matter lesions). There were also some inconclusive results with potentially noteworthy point estimates but wide 95% confidence intervals that crossed the null, such as KSHV in association with more white matter lesions. The pattern of results did not differ meaningfully in our sensitivity analyses (see Supplementary Figures 5-7).

3.2.1 *APOE* interactions

Meta-analysed findings for *APOE* interactions are displayed in Supplementary Table 5. We found no evidence of interactions by *APOE* e4 or *APOE* e2 carrier status following multiple testing correction. In model 3, we observed suggestive evidence of an interaction between *APOE* e2 carrier status and four serology measures (seropositivity to *T.gondii* and MCV, and both the PBI and neurotropic PBI) with smaller hippocampal volume. We subsequently tested associations of these serology measures with hippocampal volume stratified by *APOE* e2 carrier status. Results from all models can be found in Supplementary Table 6. In the final model, we report associations between PBI and smaller hippocampal volume among *APOE* e2 carriers only (per 10% increase in PBI $\beta=-0.07\text{ml}$, [-0.13,-0.01], $p=0.01$).

3.3 Seroreactivity tertiles

Meta-analysed findings for seroreactivity analyses are displayed in Figure 3 and Table 3 (see Supplementary Figures 8-10 for study-level findings, and Supplementary Tables 7-9 for full numerical results).

Seroreactivity to Epstein-Barr virus (EBV; VCA antigen) was associated with more white matter lesions in model 1 (12.1% difference [5.7, 18.6], $p=2.3\times 10^{-4}$, $p_{\text{FDR}}<0.05$) but largely attenuated in subsequent models (model 3: 4.5% difference [-1.5,10.1], $p=0.15$). We additionally found less evidence of an association between seroreactivity to the alternate EBV antigen (EBNA) and white matter lesions in sensitivity analyses (model 1: 4.6% difference [-9.6, 18.8], $p=0.53$). We found no further clear evidence of associations between the remaining antigens and outcomes following multiple testing correction, although other suggestive associations were observed in fully adjusted models (*T.gondii* with smaller brain volume (which was not robust when testing the alternate antigen); human herpesvirus (HHV)6B and *C.trachomatis* with larger hippocampal volume, and BK virus with less white matter lesions). With the exception of the *T.gondii* results

reported above, the pattern of results did not differ appreciably when running sensitivity analyses (see Supplementary Figures 11-13).

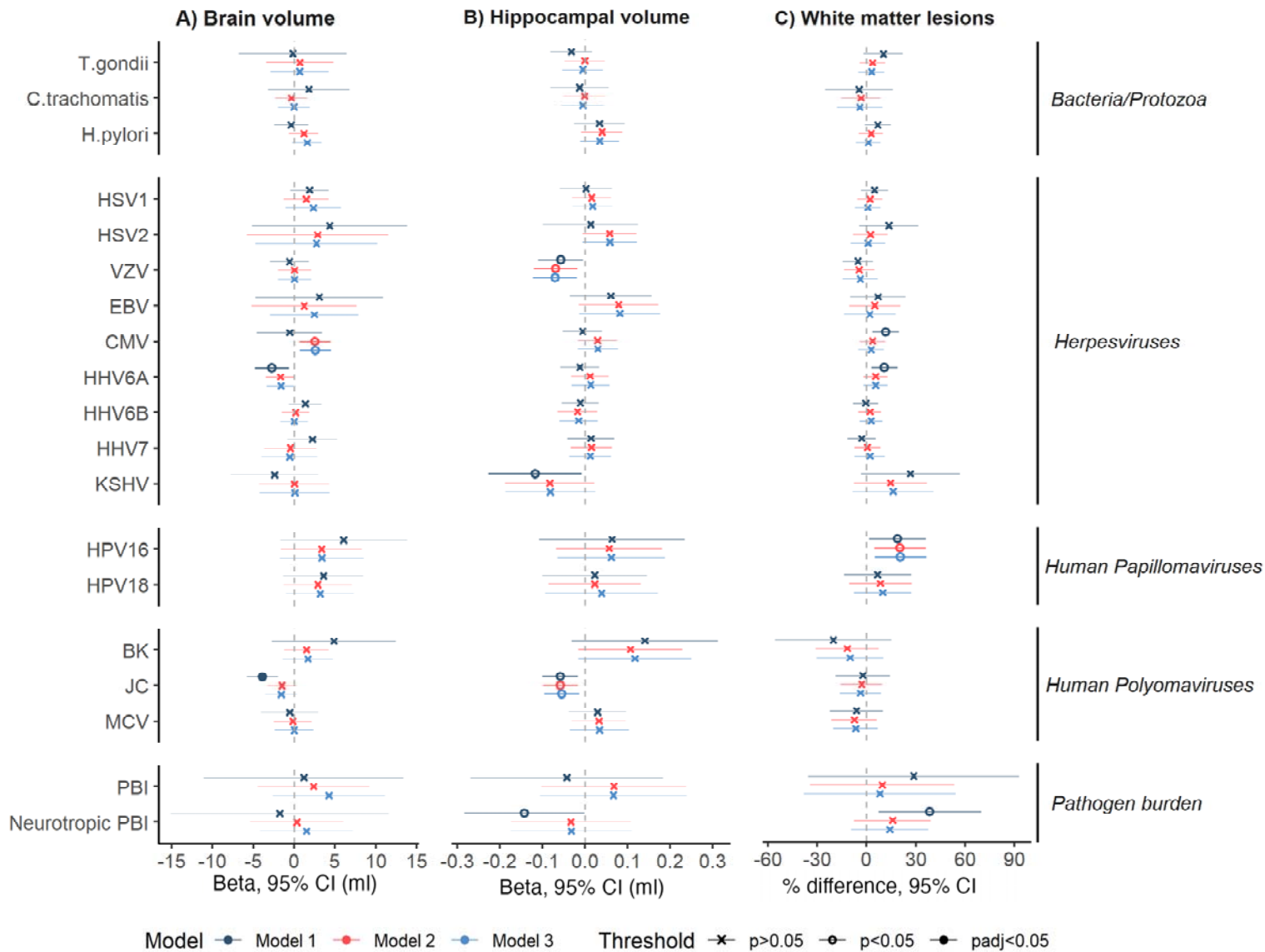


Figure 2 (left). Forest plot indicating meta-analysed associations of pathogen serostatus and burden scores neuroimaging outcomes.

Positive estimates indicate large brain and hippocampal volumes; more white matter lesions. For pathogen burden analyses, estimates represent the change in outcome (proportion of pathogens or neurotropic pathogens seropositive). Model 1 included adjustment for total intracranial volume and other technical covariates; model 2 additionally for age, sex, and ethnicity; and model 3 additionally for BMI, smoking status, education attainment, socioeconomic position, and alcohol intake. Statistical significance is indicated by colour, and significance threshold by shape. Pathogen families are further indicated on the right-hand side of the figure.

Abbreviations: BK=BK virus; CMV=Cytomegalovirus; EBV=Epstein Barr virus; HHV=Human herpesvirus; HPV=Human Papillomavirus; HSV=Herpes simplex virus; JC=John Cunningham virus; KSHV=Kaposi's sarcoma-associated herpesvirus; MCV=Merkel Cell virus; PBI=Pathogen burden index; VZV=Varicella zoster virus.

Table 2. Meta-analysed associations of serostatus and pathogen burden scores with neuroimaging outcomes.

Estimates, 95% confidence intervals, and p values are presented. For pathogen burden analyses, estimates represent the difference in outcome per 0.1 unit increase in pathogen burden (proportion of pathogens or neurotropic pathogens seropositive to). Associations significant at $p < 0.05$ are in bold, but none presented survived multiple testing correction. Significant heterogeneity (either I^2 statistic > 50 and/or Cochran's Q $p < 0.05$) are indicated by †. Due to space, only results from model 2 (adjusting for total intracranial volume and other technical covariates, age, sex, and ethnicity) and model 3 (additionally adjusting for BMI, smoking status, educational attainment, socioeconomic position, and alcohol intake) are shown.

	Brain volume		Hippocampal volume		White matter lesions	
	Model 2	Model 3	Model 2	Model 3	Model 2	Model 3
	β (ml) [95% CI]	β (ml) [95% CI]	β (ml) [95% CI]	β (ml) [95% CI]	% change [95%CI]	% change [95%CI]
HSV1	1.48 [-1.26,4.23]†	2.35 [-1.04,5.73]†	0.02 [-0.03,0.06]	0.02 [-0.03,0.06]	2.08 [-5.47,9.62]	0.8 [-6.88,8.47]
HSV2	2.87 [-5.81,11.54]†	2.73 [-4.73,10.19]†	0.06 [-0.01,0.12]	0.06 [0,0.12]	2.36 [-8.02,12.74]	0.94 [-9.47,11.35]
VZV	0.03 [-2,2.07]	0.04 [-1.97,2.06]	-0.07 [-0.12,-0.02]	-0.07 [-0.12,-0.02]	-4.47 [-13.63,4.7]	-3.82 [-14.6,6.96]
EBV	1.21 [-5.19,7.62]†	2.45 [-2.98,7.88]	0.08 [-0.01,0.17]	0.08 [-0.01,0.18]	5.03 [-10.45,20.51]	2 [-13.66,17.66]
CMV	2.57 [0.66,4.48]	2.62 [0.71,4.53]	0.03 [-0.02,0.08]	0.03 [-0.02,0.08]	3.61 [-4.15,11.36]	2.84 [-4.98,10.67]
HHV6A	-1.68 [-3.45,0.09]	-1.61 [-3.37,0.15]	0.01 [-0.03,0.06]	0.01 [-0.03,0.06]	5.5 [-1.77,12.77]	5.48 [-1.81,12.76]
HHV6B	0.17 [-1.52,1.85]	0 [-1.68,1.67]	-0.02 [-0.06,0.03]	-0.01 [-0.06,0.03]	2.06 [-4.86,8.97]	2.79 [-4.17,9.75]
HHV7	-0.46 [-3.62,2.71]	-0.56 [-3.99,2.86]	0.02 [-0.03,0.06]	0.01 [-0.04,0.06]	0.52 [-7.39,8.43]	2.02 [-7.25,11.29]
KSHV	0.02 [-4.26,4.31]	0.07 [-4.21,4.35]	-0.08 [-0.19,0.02]	-0.08 [-0.19,0.02]	14.61 [-7.55,36.77]	16.14 [-8.48,40.77]
HPV16	3.35 [-1.6,8.29]	3.37 [-1.76,8.51]	0.06 [-0.07,0.18]	0.06 [-0.06,0.19]	20.48 [4.77,36.18]	20.78 [5.03,36.53]
HPV18	2.9 [-1.27,7.07]	3.18 [-0.97,7.32]	0.02 [-0.09,0.13]	0.04 [-0.09,0.17]	8.48 [-10.49,27.46]	9.93 [-7.48,27.33]
BK	1.5 [-1.22,4.23]	1.69 [-1.37,4.74]	0.11 [-0.01,0.23]†	0.12 [-0.01,0.25]†	-11.77 [-30.99,7.45]	-10.03 [-30.32,10.26] †
JC	-1.51 [-3.31,0.29]	-1.59 [-3.55,0.36]	-0.06 [-0.1,-0.02]	-0.05 [-0.09,-0.01]	-3 [-15.5,9.5]†	-3.64 [-16.18,8.9]†
MCV	-0.16 [-2.47,2.14]	0 [-2.38,2.37]	0.03 [-0.03,0.1]†	0.03 [-0.04,0.1]†	-7.42 [-21.24,6.4]†	-6.58 [-20.09,6.93]†
Ct	-0.35 [-2.32,1.61]	-0.03 [-1.98,1.93]	0 [-0.05,0.05]	0 [-0.05,0.05]	-3.39 [-15.33,8.54]	-4.07 [-17.81,9.67]
Tg	0.69 [-3.42,4.8]†	0.67 [-2.88,4.22]†	0 [-0.05,0.05]	0 [-0.05,0.04]	3.67 [-4.19,11.52]	3.04 [-4.84,10.92]
Hp	1.18 [-0.61,2.96]	1.59 [-0.2,3.38]	0.04 [-0.01,0.09]	0.03 [-0.01,0.08]	2.76 [-4.58,10.09]	1.1 [-6.33,8.52]

PBI	0.24 [-0.45,0.92]	0.42 [-0.26,1.11]	0.01 [-0.01,0.02]	0.01 [-0.01,0.02]	0.96 [-3.42,5.34]†	0.82 [-3.81,5.44]†
Neuro PBI	0.03 [-0.54,0.6]	0.15 [-0.42,0.72]	0 [-0.02,0.01]	0 [-0.02,0.01]	1.58 [-0.75,3.91]	1.41 [-0.93,3.76]

Abbreviations: BK=BK virus; CMV=Cytomegalovirus; Ct=*C.trachomatis*; EBV=Epstein-Barr virus; HHV=Human herpesvirus; Hp=*H.pylori*; HPV=Human Papillomavirus; HSV=Herpes simplex virus; JC=John Cunningham virus; KSHV=Kaposi's sarcoma-associated herpesviruses; MCV=Merkel Cell virus; PBI=Pathogen

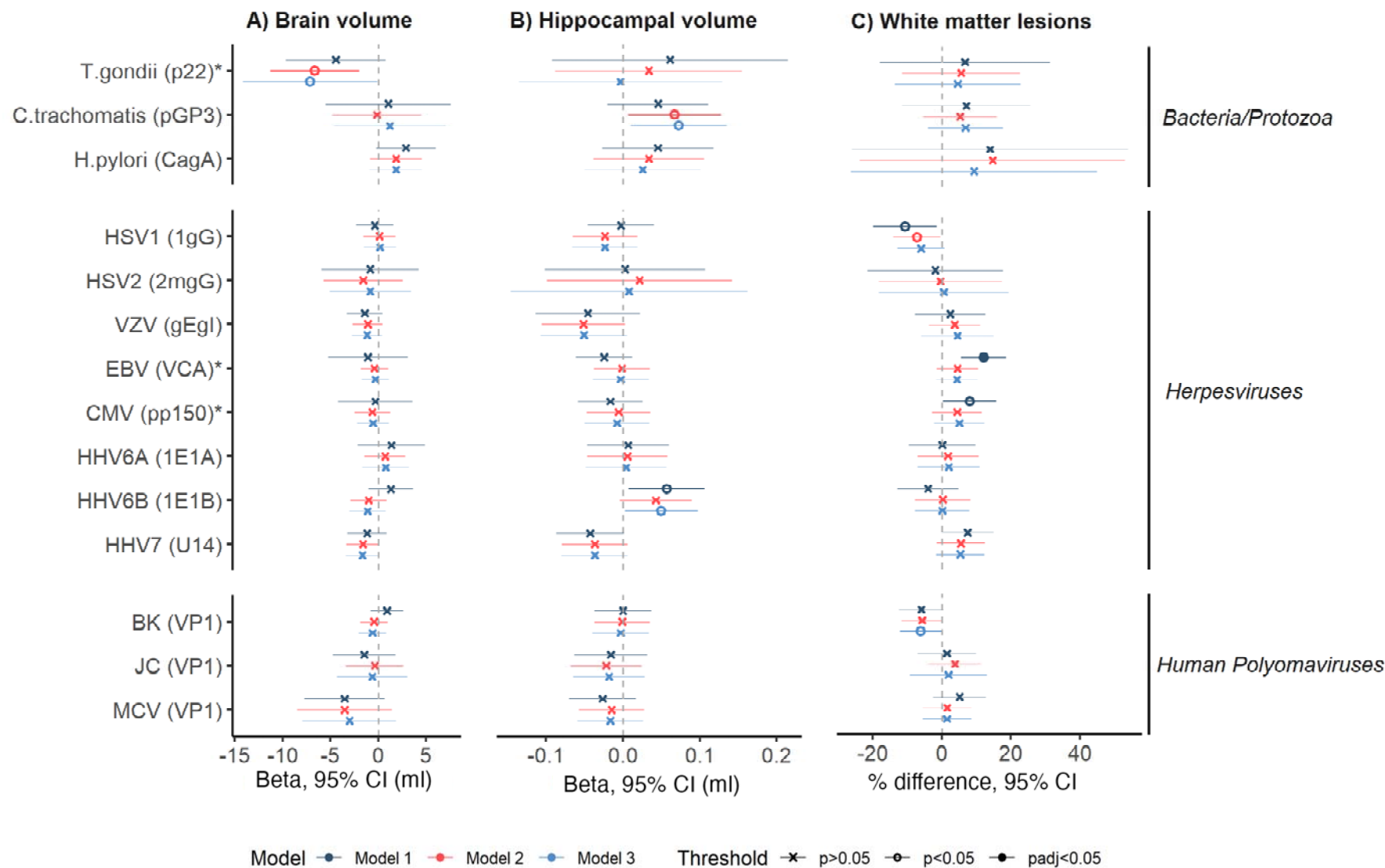


Figure 3. Forest plot indicating meta-analysed results between seroreactivity tertiles and neuroimaging outcomes.

Positive estimates indicate larger brain and hippocampal volumes, but more white matter lesions. Model 1 included adjustments for total intracranial volume and other technical covariates; model 2 additionally for age, sex, and ethnicity; and model 3 additionally for BMI, smoking status, educational attainment, socioeconomic position, and alcohol intake. Statistical models are indicated by colour, and significance threshold by shape. Pathogens and specific antigens (in brackets) are indicated on the y-axis, with pathogen families further indicated on the right-hand side of the figure. Antigens selected at random from a larger pool of recommended antigens for the pathogen are indicated by an asterisk. Abbreviations: BK=BK virus; CMV=Cytomegalovirus; EBV=Epstein-Barr virus; HHV=Human herpesvirus; HPV=Human Papillomavirus; HSV=Herpes simplex virus; JC=John Cunningham virus; MCV=Merkel Cell virus; VZV=Varicella zoster virus.

Table 3. Meta-analysed associations of seroreactivity tertiles with neuroimaging outcomes.

Estimates, 95% confidence intervals, and p values are presented. Associations significant at $p < 0.05$ are in bold, but none presented survived multiple testing correction. Significant heterogeneity (either I^2 statistic > 50 and/or Cochran's Q $p < 0.05$) are indicated by †. Due to space, only results from model 2 (adjusting for total intracranial volume, other technical covariates, age, sex, and ethnicity) and model 3 (additionally adjusting for BMI, smoking status, educational attainment, socioeconomic position, and alcohol intake) are shown.

	Brain volume		Hippocampal volume		White matter lesions	
	Model 2	Model 3	Model 2	Model 3	Model 2	Model 3
	β (ml)[95% CI]	β (ml)[95% CI]	β (ml)[95% CI]	β (ml) [95% CI]	% change [95%CI]	% change [95%CI]
HSV1	0.12 [-1.58,1.81]	0.18 [-1.52,1.87]	-0.02[-0.07,0.02]	-0.02[-0.07,0.02]	-7.19[-13.99,-0.39]	-6.01[-12.83,0.82]
HSV2	-1.58 [-5.7,2.54]	-0.85 [-5.1,3.41]	0.02 [-0.1,0.14]	0.01[-0.15,0.16]	-0.41[-18.24,17.43]	0.56[-18.21,19.32]
VZV	-1.1 [-2.67,0.47]	-1.18 [-2.74,0.39]	-0.05[-0.11,0]	-0.05[-0.11,0.01]	3.7[-3.67,11.07]	4.53[-5.93,14.99]†
EBV	-0.41 [-1.84,1.03]	-0.31 [-1.73,1.12]	0 [-0.04,0.03]	0 [-0.04,0.03]	4.58[-1.36,10.52]	4.44[-1.53,10.4]
CMV	-0.62 [-2.5,1.25]	-0.55 [-2.23,1.12]	-0.01[-0.05,0.04]	-0.01[-0.05,0.03]	4.46[-2.75,11.66]	5.05[-2.21,12.3]
HHV6A	0.71 [-1.42,2.83]	0.77 [-1.68,3.21]	0.01 [-0.05,0.06]	0 [-0.05,0.06]	1.82[-7.02,10.66]	1.99[-6.99,10.98]
HHV6B	-1.02 [-2.93,0.88]	-1.12 [-3.01,0.78]	0.04[0,0.09]	0.05 [0,0.1]	0.33[-7.7,8.36]	0.18[-7.71,8.07]
HHV7	-1.62 [-3.35,0.12]	-1.65 [-3.38,0.07]	-0.04[-0.08,0.01]	-0.04[-0.08,0.01]	5.52[-1.46,12.49]	5.37[-1.65,12.39]
BK	-0.45 [-1.9,1]	-0.59 [-2.04,0.85]	0 [-0.04,0.03]	0 [-0.04,0.03]	-5.67[-11.61,0.28]	-6.1[-12.08,-0.11]
JC	-0.37 [-3.37,2.63]†	-0.63 [-4.33,3.07]†	-0.02[-0.07,0.02]	-0.02[-0.06,0.03]	3.78[-3.76,11.31]	1.95[-9.23,13.13]
MCV	-3.52 [-8.45,1.41]†	-3.02 [-7.93,1.88]†	-0.01[-0.06,0.03]	-0.02[-0.06,0.03]	1.56[-5.41,8.54]	1.52[-5.52,8.55]
Ct	-0.14 [-4.81,4.52]	1.2 [-4.64,7.04]†	0.07 [0.01,0.13]	0.07[0.01,0.14]	5.33[-5.33,15.99]	6.87[-4.04,17.78]
Tg	-6.63 [-11.26,-2]	-7.11 [-14.13,-0.08]	0.03[-0.09,0.15]	0[-0.14,0.13]	5.58[-11.46,22.61]	4.63[-13.49,22.75]
Hp	1.86 [-0.84,4.55]	1.84 [-0.9,4.59]	0.03[-0.04,0.11]	0.03[-0.05,0.1]	14.71[-23.7,53.12]†	9.36[-26.28,44.99]†

Abbreviations: BK=BK virus; CMV=Cytomegalovirus; Ct=*C. trachomatis*; EBV=Epstein-Barr virus; HHV=Human herpesvirus; Hp=*H. pylori*; HSV=Herpes simplex virus; JC=John Cunningham virus; MCV=Merkel Cell virus; Tg=*T. gondii*; VZV=Varicella zoster virus. Kaposi's sarcoma-associated herpesviruses and Human Papillomavirus 16 and 18 were not tested as seroprevalence were $< 5\%$ in at least one cohort.

4. Discussion

In this cross-cohort study using three population-based cohorts, we examined associations of multiple common pathogens with neuroimaging markers related to subclinical dementia – brain volume, hippocampal volume, and white matter lesions – and tested whether these associations were modified by *APOE* ϵ 4 and ϵ 2 carrier status. We found little or no evidence of strong relationships in most instances. Nevertheless, several suggestive associations require examination in other cohorts with serology and neuroimaging data, given the possibility of non-negligible magnitudes of association – e.g. the difference in hippocampal volume between VZV-seropositive and seronegative individuals (-0.07ml) equates to approximately two thirds of the life course-averaged effect of *APOE* ϵ 4 carriage on hippocampal volume (-0.11ml) ³².

Among infectious agents that may have a role in the aetiology of dementia, links between HSV and Alzheimer's disease (AD) are perhaps the most studied to date ³³. Many of the other pathogens with serological markers on our panel have also been associated with AD or other causes of dementia, although findings have been contested ³⁴. Nevertheless, our null findings for HSV – alongside those for other pathogens – in relation to neuroimaging outcomes are in line with several other studies. For example, serostatuses for *C.pneumoniae*, *H.pylori* and CMV were not associated prospectively with dementia risk, whole brain volume, or white matter lesions in the Framingham cohort study ³⁵. In a study using data from the Baltimore Longitudinal Study of Ageing, no associations were found between symptomatic herpesvirus infections and changes in brain volume or AD signature regions; although some associations were reported for other markers of dementia risk, including attentional decline and astrogliosis ³⁶. Null findings between HSV1 serostatus and whole brain atrophy were additionally reported in a cohort with parental history of early-onset autosomal dominant AD ³⁷. In contrast, relationships between higher seroreactivity to HSV (either 1 or 2) and smaller hippocampal volume were reported in an analysis involving two small cohorts (N=349) ³⁸. Additionally, using the same multiplex serology data as the present study, an association of HSV1 serostatus with incident dementia diagnosis was observed in the UKB. No relationships were clearly identifiable for the other pathogens; however, with only 84 incident dementia cases available in the sample, that analysis may have lacked power to detect moderate but clinically meaningful associations ³⁹.

There could be several explanations where associations of pathogens with neuroimaging measures were not observed. Since severe infections, including those leading to hospitalisation, appear to be risk factors for AD ⁴⁰, virulent infections (rather than any exposure or seroreactivity

against pathogen antigens measured by serology) could be more relevant to dementia risk. Alternatively, pathogen exposure might only have a role later in the disease trajectory, rather than in early pathogenesis. Participants included in the present analyses are still relatively young (mean age 63.8-70.7 among the cohorts studied) and could be many years prior to developing overt signs of neurodegeneration and cerebrovascular pathology. A further possibility is that interactions between pathogens, or a combined burden of many, would be necessary to drive the development of neuropathology. In accordance with this hypothesis, pathogen burden indices derived from counts of serostatus values have been associated with dementia risk in other settings^{39,41}, but we found no convincing evidence of associations of pathogen burden scores with neuroimaging outcomes. Finally, relationships may additionally be contingent on interactions with other environmental or genetic factors. For instance, one study observed increased AD risk with higher HSV1 seroreactivity in *APOE* e4 carriers only⁹, and other data suggested that CMV and *H.pylori* serostatuses associated differently with whole brain volume according to e4 carriage³⁵. We did not find evidence of *APOE* e4 modification of serostatus associations with neuroimaging measures; however, identifying such interactions robustly in exploratory analyses with many statistical tests like ours may require much larger sample sizes. We found some suggestive evidence for associations of exposures with smaller hippocampal volume in *APOE* e2 carriers only, but are not aware of any reason why these would be specific to *APOE* e2. It is possible this is a chance finding, and would want to see whether the results are robust to replication, although we note that a previous study reported an association of increasing pathogen burden with worse cognitive function in non-*APOE* e4 carriers only¹⁰.

Among other suggestive findings that we observed, associations of seropositivity to VZV and JC virus with smaller hippocampal volumes, and of HPV16 seropositivity with more white matter lesions, have plausibility. VZV infection and its clinical manifestation as shingles have been investigated as risk factors for dementia in many settings, although findings have been mixed^{8,42}. JC virus is able to cause a rare, aggressive neurodegenerative disease under immunosuppressed conditions⁴³, and some evidence has suggested associations of HPV infection with cardiovascular outcomes⁴⁴. To our knowledge, no other studies have investigated associations of these infections with subclinical neuroimaging outcomes. No prospective associations of serostatus for these pathogens with all-cause dementia risk were detected by a study using UKB data, although, as mentioned, results in this study suffered from a lack of statistical precision³⁹; in particular, their finding for VZV clearly requires follow-up in larger data (OR=3.38 for seropositive individuals; 95% CI: 0.83, 13.78). In contrast, some of our other suggestive findings

were unanticipated based on our hypothesis that pathogen exposures would be associated with worse neuroimaging metrics. For example, the associations of CMV seropositivity with larger whole brain volumes, and three of the seroreactivity measures with larger hippocampal volume (for HHV6B and *C.trachomatis*) or fewer white matter lesions (for BK virus). Possible reasons for these associations are unclear, and we are not aware of other studies reporting relationships of these infections with dementia or related outcomes in this direction. As with any of the suggestive findings reported, these may have arisen due to chance or bias, e.g. from residual confounding, and we stress the need to see whether these results can be replicated in other data.

Strengths and limitations

This study is the largest and broadest – in terms of number of pathogens examined – to investigate associations of common infections with neuroimaging markers of dementia to date. We used harmonised exposure and neuroimaging data from three large and well-characterised population-based cohorts, with an array of data on possible confounders. Furthermore, multiplex serology data directly ascertains pathogen exposure and circumvents reliance on inaccurate diagnoses from medical records or self-report of symptomatic infections from surveys. The use of subclinical markers of dementia risk are additionally less prone (although not immune) to biases, such as survival bias, in comparison to studies with clinical endpoints as outcomes. We also note several limitations. First, these findings are observational and hence associations might still arise from reverse causality, residual confounding, or other sources of bias. While most of the infections we studied are typically acquired earlier in life and thus are likely to precede the onset of neuropathology, the long prodromes of AD and other causes of dementia may mean neurodegenerative processes could influence immune function and subsequent serological values over long periods. Second, participants remaining in NSHD are only broadly representative of the study population at recruitment and Insight 46 study members are on average of better self-reported health, cognitive function, and socioeconomic position than the full NSHD cohort^{45,46}. In addition, participants in the UKB (and particularly the imaging subset) are healthier and more socially advantaged than the general UK population at the same age range^{17,47}. We incorporated sampling weights for NSHD, but no weighting was applied in the other cohorts. Third, serology values were measured once, and some individuals may have been misclassified – particularly in NSHD and UKB if individuals were infected in the period between sample collection for serology assaying and the neuroimaging visits. Fourth, as seroreactivity analyses were conducted in the seropositive subset only, results may be biased through selection on serostatus. Fifth, we were unable to identify the age of infection onset or

severity of infection with these data, and therefore investigations that addressed the timing and severity of infections were not possible, although the use of seroreactivity values as exposures may indicate more recent infection or reactivation. Sixth, as primary infections for many of the studied pathogens typically occur earlier in life, some of the behavioural and lifestyle variables included in model 3 may occur following infection and be inappropriate as confounders. Nevertheless, model 3 adjustments did not result in large changes to most results. Finally, we restricted the neuroimaging outcomes to those available in all three cohorts (measures of brain structure and white matter lesions), and thus did not address potential relationships of these exposures with more specific pathologies of dementia subtypes, such as amyloid.

Future directions

There are several ways to build upon the evidence presented in this study. The generation of equivalent serology data (or other molecular measures of infection), in studies with neuroimaging data and/or clinical follow-up, will help to estimate associations with more precision to establish or refute some of the suggestive findings observed in this study. Given many of these infections are preventable or treatable, this could help inform public health strategies, such as the prioritisation and coverage of vaccination programmes. Furthermore, expanding this work to include longitudinal outcomes, such as rates of atrophy and change in white matter lesion burden could provide more granular insights into relationships with neurodegeneration and cerebrovascular pathology. Finally, incorporating other measures of neuropathology (e.g. cerebral amyloidosis), and fluid-based biomarkers of neurodegeneration will help to triangulate possible effects of pathogens on multiple pathways affecting dementia risk.

Acknowledgements

We are grateful to study participants of the NSHD (and Insight 46), SABRE, and UKB. We thank Lee Hamill Howes, Andrew Wong, Kenan Direk, and Felicia Huang, as well as other members of the Study Support Team at the MRC Unit for Lifelong Health and Ageing at UCL for their invaluable help towards the generation and curation of NSHD and SABRE data used in this project. We are also grateful to the radiographers and nuclear medicine physicians at the UCL Institute of Nuclear Medicine and to the staff at the Leonard Wolfson Experimental Neurology Centre at UCL. We are particularly indebted to the support of the late Dr Chris Clark of Avid Radiopharmaceuticals who championed the Insight 46 study from its outset. This research has been conducted using the UK Biobank Resource under Application Number 71702. JMS acknowledges the support of the National Institute for Health Research University College

London Hospitals Biomedical Research Centre, Wolfson Foundation, Alzheimer's Research UK, Brain Research UK, Weston Brain Institute, Medical Research Council, British Heart Foundation, UK Dementia Research Institute and Alzheimer's Association. We further acknowledge the use of BioRender.com for the creation of Figure 1.

Sources of funding

This research was supported by funding from the British Heart Foundation (PG/21/10776), the UK Medical Research Council (MC_UU_00019/1; MC_UU_00019/2; MC_UU_00019/3) and Open Philanthropy. CHS is supported by an Alzheimer's Society Junior Fellowship (AS-JF-17-011). CWG is supported by a Wellcome Career Development Award (225868/Z/22/Z).

Competing interests

JMS has received research funding and PET tracer from AVID Radiopharmaceuticals (a wholly owned subsidiary of Eli Lilly) and Alliance Medical; has consulted for Roche, Eli Lilly, Biogen, AVID, Merck and GE; and received royalties from Oxford University Press and Henry Stewart Talks. He is Chief Medical Officer for Alzheimer's Research UK. NC receives funds from AstraZeneca for serving on data safety and monitoring committees for clinical trials of glucose lowering agents.

Data availability

Per-study results are available on request. Analytical code will be made publicly available.

References

1. Coates, M. M. *et al.* Burden of non-communicable diseases from infectious causes in 2017: a modelling study. *Lancet Glob. Health* **8**, e1489–e1498 (2020).
2. Traylen, C. M. *et al.* Virus reactivation: a panoramic view in human infections. *Future Virol.* **6**, 451 (2011).
3. Qiao, H. *et al.* Cerebral Organoids for Modeling of HSV-1-Induced-Amyloid β Associated Neuropathology and Phenotypic Rescue. *Int. J. Mol. Sci.* **23**, (2022).
4. Lotz, S. K., Blackhurst, B. M., Reagin, K. L. & Funk, K. E. Microbial Infections Are a Risk Factor for Neurodegenerative Diseases. *Front. Cell. Neurosci.* **15**, (2021).

5. Hodel, F. *et al.* The combined impact of persistent infections and human genetic variation on C-reactive protein levels. *BMC Med.* **20**, 416 (2022).
6. Amor, S., Puentes, F., Baker, D. & van der Valk, P. Inflammation in neurodegenerative diseases. *Immunology* **129**, 154–169 (2010).
7. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer’s disease identifies new risk loci and implicates A β , tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
8. Warren-Gash, C. *et al.* Human herpesvirus infections and dementia or mild cognitive impairment: a systematic review and meta-analysis. *Sci. Rep.* **9**, 4743 (2019).
9. Linard, M. *et al.* Interaction between APOE4 and herpes simplex virus type 1 in Alzheimer’s disease. *Alzheimers Dement.* **16**, 200–208 (2020).
10. Zhao, C. *et al.* APOE ϵ 4 modifies the relationship between infectious burden and poor cognition. *Neurol. Genet.* **6**, (2020).
11. Kuhlmann, I., Minihane, A. M., Huebbe, P., Nebel, A. & Rimbach, G. Apolipoprotein E genotype and hepatitis C, HIV and herpes simplex disease risk: a literature review. *Lipids Health Dis.* **9**, 8 (2010).
12. Mentzer, A. J. *et al.* Identification of host–pathogen-disease relationships using a scalable multiplex serology platform in UK Biobank. *Nat. Commun.* **13**, 1818 (2022).
13. Stafford, M. *et al.* Using a birth cohort to study ageing: representativeness and response rates in the National Survey of Health and Development. *Eur. J. Ageing* **10**, 145–157 (2013).
14. Tillin, T., Forouhi, N. G., McKeigue, P. M. & Chaturvedi, N. Southall And Brent REvisited: Cohort profile of SABRE, a UK population-based comparison of cardiovascular disease and diabetes in people of European, Indian Asian and African Caribbean origins. *Int. J. Epidemiol.* **41**, 33–42 (2012).
15. Lane, C. A. *et al.* Study protocol: Insight 46 – a neuroscience sub-study of the MRC National Survey of Health and Development. *BMC Neurol.* **17**, 75 (2017).

16. Jones, S. *et al.* Cohort Profile Update: Southall and Brent Revisited (SABRE) study: a UK population-based comparison of cardiovascular disease and diabetes in people of European, South Asian and African Caribbean heritage. *Int. J. Epidemiol.* **49**, 1441–1442e (2020).
17. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* **12**, e1001779 (2015).
18. Waterboer, T. *et al.* Multiplex Human Papillomavirus Serology Based on In Situ–Purified Glutathione S-Transferase Fusion Proteins. *Clin. Chem.* **51**, 1845–1853 (2005).
19. Le Govic, Y., Demey, B., Cassereau, J., Bahn, Y.-S. & Papon, N. Pathogens infecting the central nervous system. *PLoS Pathog.* **18**, e1010234 (2022).
20. Cardoso, M. J. *et al.* Geodesic Information Flows: Spatially-Variant Graphs and Their Application to Segmentation and Fusion. *IEEE Trans. Med. Imaging* **34**, 1976–1988 (2015).
21. Sudre, C. H. *et al.* Bayesian Model Selection for Pathological Neuroimaging Data Applied to White Matter Lesion Segmentation. *IEEE Trans. Med. Imaging* **34**, 2079–2102 (2015).
22. Lane, C. A. *et al.* Associations between blood pressure across adulthood and late-life brain structure and pathology in the neuroscience substudy of the 1946 British birth cohort (Insight 46): an epidemiological study. *Lancet Neurol.* **18**, 942–952 (2019).
23. Bann, D., Wright, L., Hardy, R., Williams, D. M. & Davies, N. M. Polygenic and socioeconomic risk for high body mass index: 69 years of follow-up across life. *PLOS Genet.* **18**, e1010233 (2022).
24. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
25. Price, A. L. *et al.* Long-Range LD Can Confound Genome Scans in Admixed Populations. *Am. J. Hum. Genet.* **83**, 132–135 (2008).
26. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

27. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
28. Rubin, D. B. *Multiple imputation for nonresponse in surveys*. vol. 81 (John Wiley & Sons, 2004).
29. Cole, T. J. Sympercents: symmetric percentage differences on the 100 loge scale simplify the presentation of log transformed data. *Stat. Med.* **19**, 3109–3125 (2000).
30. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
31. Lane, C. A. *et al.* Associations Between Vascular Risk Across Adulthood and Brain Pathology in Late Life: Evidence From a British Birth Cohort. *JAMA Neurol.* **77**, 175–183 (2020).
32. Walhovd, K. B. *et al.* Genetic risk for Alzheimer disease predicts hippocampal volume through the human lifespan. *Neurol. Genet.* **6**, (2020).
33. Harris, S. A. & Harris, E. A. Molecular Mechanisms for Herpes Simplex Virus Type 1 Pathogenesis in Alzheimer’s Disease. *Front. Aging Neurosci.* **10**, (2018).
34. Seaks, C. E. & Wilcock, D. M. Infectious hypothesis of Alzheimer disease. *PLoS Pathog.* **16**, e1008596 (2020).
35. Zilli, E. M. *et al.* Herpes Labialis, Chlamydomphila pneumoniae, Helicobacter pylori, and Cytomegalovirus Infections and Risk of Dementia: The Framingham Heart Study. *J. Alzheimers Dis. JAD* **82**, 593–605 (2021).
36. Duggan, M. R. *et al.* Herpes Viruses in the Baltimore Longitudinal Study of Aging. *Neurology* **99**, e2014–e2024 (2022).
37. Warren-Gash, C. *et al.* Herpes simplex virus and rates of cognitive decline or whole brain atrophy in the Dominantly Inherited Alzheimer Network. *Ann. Clin. Transl. Neurol.* **9**, 1727–1738 (2022).
38. Linard, M. *et al.* Herpes simplex virus, early neuroimaging markers and incidence of Alzheimer’s disease. *Transl. Psychiatry* **11**, 1–9 (2021).

39. Mekli, K. *et al.* Investigation of the association between the antibody responses to neurotropic viruses and dementia outcomes in the UK Biobank. *PLoS ONE* **17**, e0274872 (2022).
40. Levine, K. S. *et al.* Virus exposure and neurodegenerative disease risk across national biobanks. *Neuron* **111**, 1086-1093.e2 (2023).
41. Bu, X.-L. *et al.* A study on the association between infectious burden and Alzheimer's disease. *Eur. J. Neurol.* **22**, 1519–1525 (2015).
42. Chen, V. C.-H. *et al.* Herpes Zoster and Dementia: A Nationwide Population-Based Cohort Study. *J. Clin. Psychiatry* **79**, 8164 (2017).
43. Ferenczy, M. W. *et al.* Molecular Biology, Epidemiology, and Pathogenesis of Progressive Multifocal Leukoencephalopathy, the JC Virus-Induced Demyelinating Disease of the Human Brain. *Clin. Microbiol. Rev.* **25**, 471–506 (2012).
44. Tonhajzerova, I. *et al.* Novel Biomarkers of Early Atherosclerotic Changes for Personalised Prevention of Cardiovascular Disease in Cervical Cancer and Human Papillomavirus Infection. *Int. J. Mol. Sci.* **20**, 3720 (2019).
45. Kuh, D. *et al.* The MRC National Survey of Health and Development reaches age 70: maintaining participation at older ages in a birth cohort study. *Eur. J. Epidemiol.* **31**, 1135–1147 (2016).
46. James, S.-N. *et al.* Using a birth cohort to study brain health and preclinical dementia: recruitment and participation rates in Insight 46. *BMC Res. Notes* **11**, 885 (2018).
47. Lyall, D. M. *et al.* Quantifying bias in psychological and physical health in the UK Biobank imaging sub-sample. *Brain Commun.* **4**, fcac119 (2022).