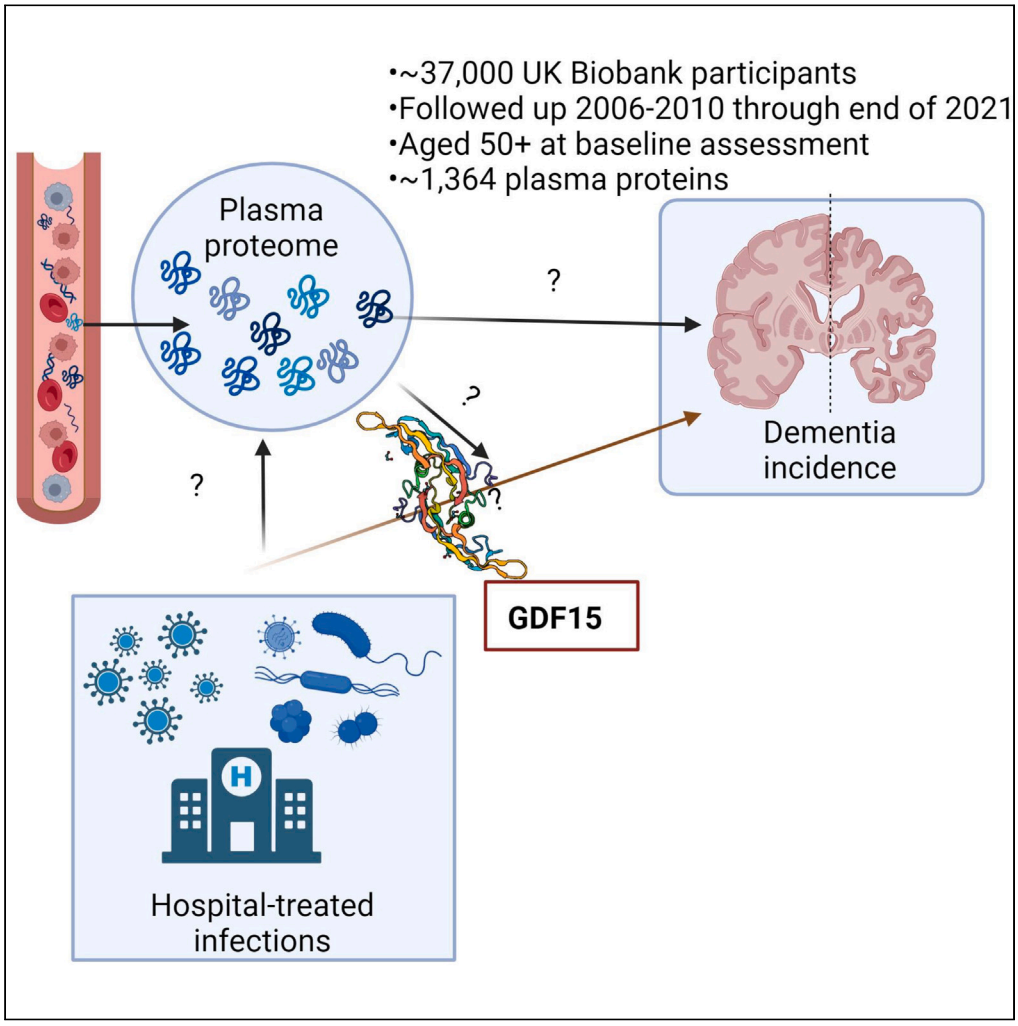
iScience

CellPress
OPEN ACCESS

Article

Hospital-treated prevalent infections, the plasma proteome and incident dementia among UK older adults



May A. Beydoun, Hind A. Beydoun, Nicole Noren Hooten, ..., Lenore J. Launer, Michele K. Evans, Alan B. Zonderman

baydounm@mail.nih.gov

Highlights

Up to 37,269 UK Biobank participants (50+ years) had plasma proteomic data

Hospital-treated infections were associated with two-folds increased dementia risk

Growth differentiation factor 15 (GDF15) protein was the strongest mediator

Beydoun et al., iScience *26*, 108526
December 15, 2023 © 2023
https://doi.org/10.1016/j.isci.2023.108526

iScience

CellPress
OPEN ACCESS

Article

Hospital-treated prevalent infections, the plasma proteome and incident dementia among UK older adults

May A. Beydoun,[1,5,*] Hind A. Beydoun,[1,2] Nicole Noren Hooten,[1] Osorio Meirelles,[1] Zhiguang Li,[1] Ziad W. El-Hajj,[3] Jordan Weiss,[4] Christian A. Maino Vieytes,[1] Lenore J. Launer,[1] Michele K. Evans,[1] and Alan B. Zonderman[1]

SUMMARY

The plasma proteome can mediate the association of hospital-treated infections with dementia incidence. We screened up to 37,269 UK Biobank participants aged 50–74 years for the presence of a prevalent hospital-treated infection, subsequently tested as a predictor for ≤1,463 plasma proteins and dementia incidence. Four-way decomposition models decomposed infection-dementia total effect into pure mediation, pure interaction, neither or both through the plasma proteome. Hospital-treated infections increased dementia two-fold. The strongest mediation effect was through the growth differentiation factor 15 (GDF15) protein. Top 17 proteomic mediators explained collectively 5% of the total effect, while pathway analysis of all mediators (k = 221 plasma proteins) revealed top pathways including the immune system, signal transduction, metabolism, disease and metabolism of proteins, with the GDF15 cluster reflecting most strongly the "transmembrane receptor protein tyrosine kinase signaling pathway". The association of hospital-treated infections with dementia was partially mediated through GDF15 and other plasma proteomic markers.

INTRODUCTION

Dementia is among the leading causes of mortality and disability in older populations, particularly in developed countries.[1,2] Exposure to viral and bacterial infections (e.g., herpesviruses, gastrointestinal microorganisms, oral bacterial species, Chlamydia pneumonia, spirochetes) has been associated with Alzheimer's disease (AD) related dementia.[1] Host immune response to these agents can trigger or accelerate AD processes, including accumulation of amyloid beta (Aβ), tau protein, and neuroinflammation.[1,2] Previous studies have implicated infectious disease and inflammation as potential culprits because genes linked to inflammatory processes were also linked to dementia and systemic inflammation related to cognitive decline and infection was associated with new-onset dementia.[2,3] Alternative explanations include the antimicrobial protection model of AD suggesting that Aβ accumulates in response to infectious agents and the inflammation hypothesis whereby systemic inflammation contributes to AD and dementia development.[3] The burden of infectious disease has been explored in a handful of studies in relation to all-cause and AD-specific dementias.[3–6] Thus far, studies examining the plasma proteome and infections have mostly analyzed a small sub-set of proteins in small cohorts.[7,8] Using targeted and large-scale proteomics, the plasma proteome has been linked to all-cause dementia.[9–13] Furthermore, various types of infections have been shown to predict dementia risk differentially by sex.[14,15]

In this analysis of retrospective cohort data from the UK Biobank, we examined hospital-treated infection burden in relation to the incidence of all-cause dementia, while attempting to explain putative associations through the plasma proteome. Sex differentials in the hospital-treated infections versus dementia risk were also tested.

RESULTS

Study sample characteristics

The sample included 37,269 UK Biobank participants aged 50 years or older at baseline with complete data on covariates of interest and no prevalent dementia (Figure 1). Table 1 shows the baseline characteristics of the study sample as a whole and by sex. There was no sex

[1]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Intramural Research Program, NIA/NIH/IRP, Baltimore, MD 21224, USA
[2]AT Augusta Military Medical Center, Fort Belvoir, VA 22060, USA
[3]Department of Biology, McGill University, Montreal, QC, Canada
[4]Stanford Center on Longevity, Stanford University, Palo Alto, CA 94305, USA
[5]Lead contact
*Correspondence: baydounm@mail.nih.gov
https://doi.org/10.1016/j.isci.2023.108526

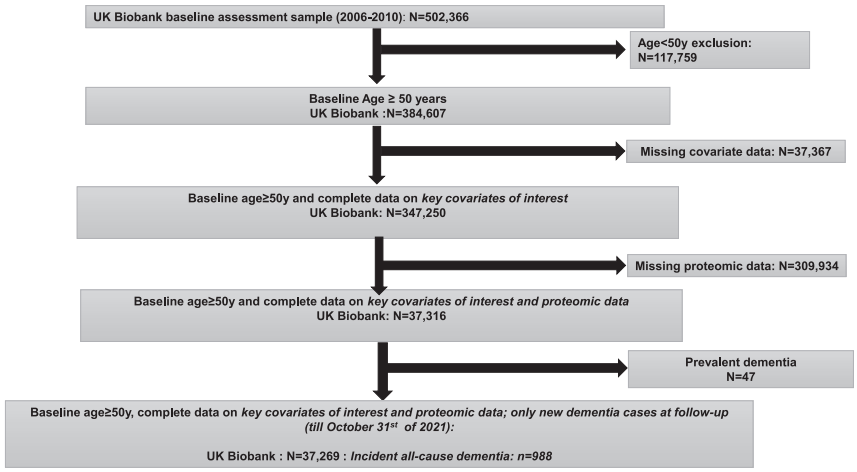


Figure 1. Participant flowchart
UK, United Kingdom.

difference in hospital-treated infection, but non-hospital treated infection burden was higher among females. All socio-demographic factors, infection load, Life's Essential 8 scores, and cumulative incidence rates showed sex differences.

Hospital-treated infections and dementia incidence

Figure S1 depicts Kaplan-Meier curves for all-cause dementia based on three levels of infection burden. The risk table shows the number of participants at risk at each age entry point and within each infection burden category. The log rank test indicates significant variations in dementia-free proportion, with hospitalized infections having higher risk of dementia, followed by having "no infections", while the lowest dementia risk was observed in the "non-hospital treated infections" category.

Table 2 shows the overall association between hospital-treated infection burden and dementia incidence, as well as the link after we stratified by AD PRS tertiles. Hospital-treated infection burden was associated with both all-cause dementia in men, women, and both sexes combined. The HR revealed a roughly 2-fold elevated risk of all-cause dementia risk in both men and women who had a hospital-treated infection as opposed to those who did not. AD PRS tertiles had no statistically significant interaction effect on these associations.

Participants with a follow-up time of less than three years were removed from a sensitivity analysis. Our findings in that sub-sample (N = 36,745; incident dementia cases n = 937) were similar to our findings in the main analyses. Most notably, the overall HR for all-cause dementia for hospital-treated infections was 2.21, with 95% CI: 1.91–2.55, $p < 0.001$. In both main and sensitivity analyses, the average number of years between the total infection burden and baseline assessment dates was around 14 years, and the estimate was 11 years for hospital-treated infections. Time elapsed distribution from infection until baseline assessment for the main sample (N = 37,269) is provided in Figure S2.

Hospital-treated infections and the plasma proteome

A series of 1,463 multiple linear regression models were conducted to examine the strongest relationships between hospital-treated infections and the plasma proteome. 583 plasma proteins were significantly predicted by hospital-treated infections after Bonferroni correction. Among those, only 22 had an effect size beta that was either <-0.25 or >+0.25, reflecting a change of $^1/_4$ standard deviation in plasma protein between the "hospital-treated infections" vs. "all others" groups. Those main findings were visualized as a volcano plot (Figure 2). Selected strongly associated proteins with hospital-treated infections (k = 22) were included into a four-way decomposition model with the outcome being incident dementia modeled with Cox regression.

Four-way decomposition models: Top mediators

Table 3 (fully presented in Table S6), Figure 3 and Datasheet S1 present key findings and indicate that the strongest plasma proteome mediator was growth differentiation factor 15 (GDF15). The 17 mediators of 22 which showed a significant pure indirect effect are shown in Table S7 with explanation of main functions and links to infection and dementia based on previous studies. In addition to exhibiting 7% pure indirect effect, GDF15 also had a marginally statistically significant mediated interaction ($p < 0.10$) accounting for 4% of the total effect of hospital-treated infection on dementia risk. When examining the full results of the four-way decomposition for GDF15, the overall proportion mediated (op_m) was 11.1% with 95% CI: 5.8%–16.5% with $p < 0.001$. Other proteins with statistically significant op_m at type I error of 0.05 included TFF3, PLAUR and EDA2R, generally explaining less than 10% of the total effect each.

A single principal components analysis (PCA) factor was extracted from the top 17 significant proteomic mediators, |beta|>0.25, approximating the 90th percentile of effect sizes; after Bonferroni correction. This PCA factor explained 5% of the total effect of hospital-treated

Table 1. Study sample characteristics by sex: UK Biobank 2006-2021

	Overall (N = 37,269)	Men (N = 17,480)	Women (N = 19,789)	P_{sex}
Demographic				
Baseline age, y	60.2 ± 0.03	60.5 ± 0.04	59.9 ± 0.04	<0.001
Sex, % female	53.1%	–	–	
Race/ethnicity				
White	95.3%	95.8%	95.8%	(Ref)
Black	1.5%	1.4%	1.6%	0.050
South Asian	1.4%	1.6%	1.2%	0.001
Other	1.8%	1.6%	1.9%	0.014
Non-White, %	4.66	4.56	4.75	0.40
Household size	2.217 ± 0.006	2.293 ± 0.009	2.149 ± 0.008	<0.001
Socioeconomic				
Education				
Low	22.1%	24.2%	20.3%	<0.001
Intermediate	39.5%	34.7%	43.6%	(Ref)
High	38.4%	41.1%	36.1%	<0.001
Income				
Less than £18,000	27.3%	25.0%	29.6%	<0.001
£18,000–£29,999	28.5%	27.6%	29.3%	(Ref)
£30,000–£51,999	23.7%	24.3%	23.1%	<0.001
£52,000–£100,000	16.3%	18.3%	14.4%	<0.001
greater than £100,000	4.1%	4.8%	3.5%	<0.001
Townsend Deprivation Index	−1.417 ± 0.016	−1.395 ± 0.024	−1.435 ± 0.021	<0.001
SES	−0.075 ± 0.004	−0.057 ± 0.005	−0.091 ± 0.005	<0.001
Infection burden, Mean ± SE				
Total	0.818 ± 0.007	0.732 ± 0.010	0.893 ± 0.011	<0.0001
Hospital-treated	0.291 ± 0.005	0.276 ± 0.007	0.304 ± 0.007	1.00
Infection burden, three-level				
None	62.6	63.9	61.4	(Ref)
Non-hospital treated only	23.5	22.1	24.7	<0.001
Hospital-treated	13.9	13.9	13.9	0.19
Life's essential 8, Mean ± SE				
Total score	500.5 ± 0.5	492.4 ± 0.7	507.6 ± 0.7	<0.001
Lifestyle score	254.1 ± 0.3	248.2 ± 0.5	259.3 ± 0.4	<0.001
Biological score	246.4 ± 0.3	244.2 ± 0.5	248.3 ± 0.5	<0.001
AD PRS				
Tertile, %				
T1	33.3	33.4	33.3	—
T2	33.3	33.2	33.4	0.78
T3	33.3	33.4	33.3	0.94
Cumulative incidence, %				
All-cause dementia	2.7%	3.1%	2.3%	<0.001
AD dementia	1.2%	1.2%	1.2%	0.91

AD, Alzheimer's Disease; LE8, Life's essential 8; PRS, Polygenic Risk Score; SE, Standard Error; UK, United Kingdom.
No multiple imputation was carried out in this analysis. p-value is associated with the parameter for sex in bivariate linear and multinomial logistic regression analyses, with the main outcome being a continuous or categorical characteristic, respectively and sex as the only predictor variable. (Ref) is the referent category in the multinomial logistic regression model. Values are means ± SE or percentages.

Table 2. Hospital-treated infections and dementia outcomes, overall, by sex and stratified by AD PRS tertile: UK Biobank 2006-2021

	Overall (N = 37,269)		By AD PRS tertiles Hazard Ratios, 95% CI			
	IB_{hosp}, yes vs. no Hazard Ratios, 95% CI	P_{IBhosp}	T_1 (N = 12,423)	T_2 (N = 12,423)	T_3 (N = 12,423)	$P_{IBhospx\ ADPRStert}$
Overall, N = 37,269	2.24 (1.94–2.56)	<0.001	2.61 (1.84–3.72)	2.65 (2.02–3.49)	2.07 (1.73–2.49)	0.15
Men, N = 17,480	2.19 (1.81–2.64)	<0.001	2.43 (1.55–3.82)	2.52 (1.77–3.59)	2.03 (1.57–2.63)	0.32
Women, N = 19,789	2.29 (1.87–2.81)	<0.001	2.85 (1.62–5.01)	2.90 (1.88–4.48)	2.11 (1.63–2.72)	0.30

AD, Alzheimer's Disease; LE8, Life's Essential 8; IB_{hosp}, Hospital-treated infection burden; PRS, Polygenic Risk Score; T, Tertile; UK, United Kingdom.
All Cox proportional hazards models were adjusted for baseline age, sex, race/ethnicity, household size, SES *Z* score and LE8 total score. Interaction between IB_{hosp} and AD PRS tertiles was tested, by including a 2-way interaction term in the reduced model.

infections on dementia risk, as depicted in Figure S3. No clear patterns were observed in analyses stratified by sex or by AD PRS tertiles. Nevertheless, mediated interaction was statistically significant in the lowest AD PRS tertile, while PIE was shown to be significant in the highest tertile for AD PRS, indicating pure mediation within that group. Additionally, among men with low AD genetic risk, the op_m was 33.7% with 95% CI: 7.4%–60.1%, p = 0.012, that attributable to interaction (op_ati) was 40.4%, with 95% CI: 1.2%–80% (p = 0.043), with a proportion of the risk eliminated (TE-CDE/TE) being 46.1% on average, with a 95% CI: 6.1%–86.1%, p = 0.024.

Four-way decomposition models: Olink insight and STRING analysis for all mediators

As shown in Figure S4, we performed Olink Insight pathway analysis including all statistically significant mediators (k = 221 proteins). The top pathways were immune system, signal transduction, metabolism, disease (including infectious and metabolic diseases), metabolism of proteins, developmental biology, hemostasis and gene expression (transcription), among others. Detailed pathways (Pathways_olink = 568) are provided in Datasheet S2. All supplementary datasheets, detailed code and related result datasets used to generate the Figures and Tables are provided in: https://github.com/baydounm/UKB-paper8-supplementarydata. More specifically, the full list of those mediators (in alphabetical order) is shown in Data S1.zip, along with the direction of the PIE (positive vs. negative) and whether mediated interaction was statistically significant at type I error of 0.05, in addition to the full Output of the four-way decomposition models (FOURWAYDECOMP_SUPPLEMENT.pdf). Both files were placed under the "FIGURES4\INSIGHT_PATHWAY_OLINK" directory in that repository.

Those key mediators were also entered into the STRING database with kmeans clustering (10 clusters) to build a functional protein-protein interaction network.[16] Each node (circle) indicates a different protein, the colors distinguish clusters, and the lines identify the type of functional (direct or indirect) relationship. The 10 clusters are listed in Figure S5. In addition, results for cluster # 10 (41 proteins), which included GDF15 at the center of the network, are presented in Figures S5 and S6. Furthermore, we also show the relationship of each of the plasma protein mediators within the top cluster containing the top mediator, with exposure (hospital-treated infection) and outcome (dementia incidence) in Table S8, using detailed results from the four-way decomposition models and the two equations (one linear model and one Cox PH model), adjusted for exogenous variables. Of the total 221 plasma proteins that were deemed statistically significant mediators, 41 were in cluster # 10 which included GDF15. Of those, four other plasma proteins (plasminogen activator, urokinase receptor (PLAUR), ephrin receptor A2 (EPHA2), leukocyte associated immunoglobulin like receptor 1 (LAIR1) and cytoskeleton associated protein 4 (CKAP4) had been selected as top mediators given their strong association with hospital-treated infections and a statistically significant PIE.

GO enrichment analysis for biological processes was performed on all significant mediators which were part of the GDF15 cluster #10 (k = 41) to examine the biological functions and pathways. Full results (194 biological processes) are shown in Datasheet S3 and top processes with -Log10(FDR) ≥ 6 are visualized in Figure S7. "Transmembrane receptor protein tyrosine kinase signaling pathway" (GO:0007169) was the pathway showing the highest strength among processes that included GDF15 (highlighted in yellow, Datasheet S3), while the most significant process was positive regulation of phosphorylation (highlighted in green, Datasheet S3; Figure S7), Other signal transduction pathways including signaling by receptor tyrosine kinases and posttranslational phosphorylation of proteins were also among the top OLINK Insight pathways as shown in Datasheet S2.

Similarly, insulin like growth factor binding protein 4 (IGFBP4) was clustered with 45 other plasma proteins deemed statistically significant mediators (Cluster # 7), none of which were included among the top 17 mediators; while WAP four-disulfide core domain 2 (WFDC2) was clustered with 20 other mediators (Cluster # 1), a cluster than also included trefoil factor 3 (TFF3) among the 17 top mediators that were strongly associated with the exposure (Figures S8 and S9). For the proteins cluster # 7 (Figure S8), the dominant GO pathway with lowest FDR and highest strength was "positive regulation of leukocyte activation" (GO:0002694), while for the proteins in cluster # 1 (Figure S9), the only GO term was "maintenance of gastrointestinal epithelium" (GO:0030277). Figure S10 shows the proteins for cluster # 8 and the strongest GO term for this cluster with the smallest FDR was "Extracellular matrix structural constituent" (GO:0005201); for proteins in cluster # 4 (Figure S11), it was "Pyrimidine metabolism" (hsa00240); for proteins in cluster # 9 (Figure S12) (including top mediator SCARB2), it was "Susceptibility to T cell mediated cytotoxicity" (GO:0060370); for proteins in cluster # 3 (Figure S13), it was "Glutathione metabolic process" (GO:0006749); for proteins in cluster # 6 (Figure S14) (including top mediators TNFRSF1A, TNFRSF1B, TNFRSF9 and EDA2R), it was "Positive

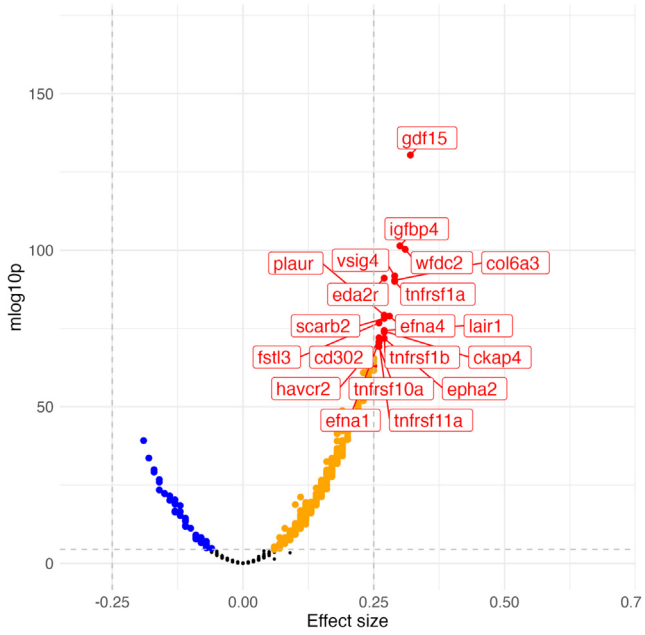


Figure 2. Volcano plot of plasma proteomic biomarkers in relation to hospital-treated prevalent infections at baseline assessment: UK biobank 2006–2010

See list of abbreviations for protein abbreviations.

Based on a series of multiple linear regression models, with main predictor being prevalent hospital-treated infection (1 = yes, 0 = no) and the outcome being each of 1,463 plasma proteomic biomarkers (Log2 transformed, z-scored). The y axis is the predictor's associated p value on a -Log10 scale and the X axis is the β coefficient (effect of hospital-treated infection (yes vs. no) on standardized z-scores of plasma proteomic markers) from the multiple linear regression models. An estimate with a Bonferroni corrected p value<0.05 and a lower confidence limit for the 95% CI of effect size >0.25 in absolute value is marked by the plasma proteomic marker abbreviation (See UKB showcase URL: https://biobank.ndph.ox.ac.uk/showcase/). Selected proteins (k = 22) for further mediation analysis have a corrected p value<0.05 and a point estimate >0.25 in absolute value (red). All other points are shown in blue (corrected p value<0.05 but effect size<0 but >-0.25), in orange (corrected p < 0.05 but effect size>0 but <0.25), and in black (corrected p value>0.05). All infections occurred prior to baseline assessment. Details are provided on GitHub: https://github.com/baydounm/UKB-paper8-supplementarydata.

regulation of I-kappaB kinase/NF-κB signaling" (GO:0043123); for proteins in cluster # 2 (Figure S15) (including top mediator CD302), the only term was "Alpha crystallin/Hsp20 domain" (CL:1870); and finally for the proteins in cluster # 5 (Figure S16) (including top mediator VSIG4), the only GO term (cellular component) was "Secretory granule" (GO:0030141). Detailed GO (or other alternate results, when GO was not available) results for each cluster are provided on Github: https://github.com/baydounm/UKB-paper8-supplementarydata, under FIGURES4_S16/STRING.

DISCUSSION

We examined hospitalized infection burden in relation to the incidence of all-cause dementia, and the potential mediating and/or moderating effects of the plasma proteome. Hospital-treated infections were associated with a 2-fold increased risk of all-cause dementia, with the strongest mediation effect being the GDF15 protein. A PCA factor encompassing key mediators (k = 17 plasma proteins), explained 5% of the total effect of hospital-treated infections on dementia risk, with the largest overall proportion mediated observed among men with low AD polygenic risk. Olink Insight pathway analysis using all significant mediator plasma proteins (k = 221) revealed top pathways related to the immune system, signal transduction, metabolism, disease, metabolism of proteins, developmental biology, hemostasis and gene expression. STRING analysis indicated that GDF15 clustered with other top mediators, including PLAUR, EPHA2, LAIR1 and CKAP4, while WFDC2 clustered with TFF3. IGFBP4 was a top mediator that did not interact functionally or physically with any of the other top mediators within its cluster. A few other clusters contained some of the remaining top mediators, and their strongest GO or other functional/biological pathway with lowest FDR included "positive regulation of I-kappaB kinase/NF-κB signaling", "Alpha crystallin/Hsp20 domain" and "secretory granule".

Several types of infections have been found in the brains of AD patients, but no one infection has been proven to be exclusively and causally associated with the disease.[17] There are several biological pathways that could occur, such as various neurotropic viral agents having a synergistic function, patients being more susceptible to infections, infections causing neuroinflammation, and infections playing a causal role in AD pathogenesis.[17] A hypothetical scenario suggests that systemic and bacterial amyloids, as well as other PAMPs, accelerate brain pathology in AD, leading to increased AD pathology.[2] This confluence of variables could then cause neurodegeneration.

Table 3. Four-way decomposition of the association between hospital-treated infections and all-cause dementia through selected plasma proteomic biomarkers (k = 17 with significant PIE): UK Biobank 2006-2021

FOURWAYDECOMP	β or %	SE	z	P	LCL	UCL	PROTEIN
tereri	**1.29**	**0.17**	**7.52**	**<0.001**	0.95	1.62	gdf15
ereri_cde	**1.14**	**0.18**	**6.54**	**<0.001**	0.80	1.49	gdf15
ereri_intref	0.00	0.01	−0.06	0.95	−0.02	0.02	gdf15
ereri_intmed	***0.06***	***0.03***	***1.76***	***0.078***	***−0.01***	***0.12***	gdf15
ereri_pie	**0.09**	**0.01**	**6.33**	**<0.001**	**0.06**	**0.11**	gdf15
tereri	**1.27**	**0.17**	**7.55**	**<0.001**	**0.94**	**1.60**	igfbp4
ereri_cde	**1.23**	**0.17**	**7.04**	**<0.001**	**0.89**	**1.57**	igfbp4
ereri_intref	0.00	0.00	−1.1	0.273	0.00	0.00	igfbp4
ereri_intmed	0.02	0.03	0.46	0.649	−0.05	0.08	igfbp4
ereri_pie	**0.03**	**0.01**	**2.08**	**0.038**	**0.00**	**0.05**	igfbp4
tereri	**1.30**	**0.17**	**7.4**	**<0.001**	**0.95**	**1.64**	wfdc2
ereri_cde	**1.25**	**0.18**	**6.84**	**<0.001**	**0.89**	**1.61**	wfdc2
ereri_intref	−0.01	0.00	−1.48	0.139	−0.02	0.00	wfdc2
ereri_intmed	0.00	0.04	−0.02	0.985	−0.07	0.07	wfdc2
ereri_pie	**0.05**	**0.01**	**4.15**	**<0.001**	**0.03**	**0.08**	wfdc2
tereri	**1.26**	**0.17**	**7.65**	**<0.001**	**0.94**	**1.59**	vsig4
ereri_cde	**1.24**	**0.17**	**7.2**	**<0.001**	**0.90**	**1.58**	vsig4
ereri_intref	0.00	0.00	−1.13	0.258	−0.01	0.00	vsig4
ereri_intmed	0.00	0.04	−0.03	0.98	−0.07	0.07	vsig4
ereri_pie	**0.03**	**0.01**	**2.2**	**0.028**	0.00	0.05	vsig4
tereri	**1.33**	**0.18**	**7.47**	**<0.001**	**0.98**	**1.68**	eda2r
ereri_cde	**1.25**	**0.19**	**6.67**	**<0.001**	**0.88**	**1.62**	eda2r
ereri_intref	0.00	0.01	−0.46	0.643	−0.02	0.01	eda2r
ereri_intmed	0.03	0.03	0.82	0.414	−0.04	0.09	eda2r
ereri_pie	**0.06**	**0.01**	**4.5**	**<0.001**	**0.03**	**0.08**	eda2r
tereri	**1.25**	**0.16**	**7.69**	**<0.001**	**0.94**	**1.57**	col6ar
ereri_cde	**1.21**	**0.17**	**7.2**	**<0.001**	**0.88**	**1.54**	col6ar
ereri_intref	0.00	0.00	−0.64	0.521	−0.01	0.00	col6ar
ereri_intmed	0.02	0.03	0.48	0.631	−0.05	0.08	col6ar
ereri_pie	**0.03**	**0.01**	**2.4**	**0.016**	**0.01**	**0.05**	col6ar
tereri	**1.26**	**0.16**	**7.68**	**<0.001**	**0.94**	**1.59**	tnfrsf1a
ereri_cde	**1.23**	**0.17**	**7.23**	**<0.001**	**0.90**	**1.57**	tnfrsf1a
ereri_intref	0.00	0.00	−1.17	0.242	−0.01	0.00	tnfrsf1a
ereri_intmed	0.01	0.03	0.17	0.864	−0.06	0.07	tnfrsf1a
ereri_pie	**0.03**	**0.01**	**2.54**	**0.011**	**0.01**	**0.05**	tnfrsf1a
tereri	**1.26**	**0.17**	**7.66**	**<0.001**	**0.94**	**1.59**	plaur
ereri_cde	**1.18**	**0.17**	**6.94**	**<0.001**	**0.84**	**1.51**	plaur
ereri_intref	0.00	0.01	0.11	0.912	−0.02	0.02	plaur
ereri_intmed	0.04	0.03	1.14	0.255	−0.03	0.10	plaur
ereri_pie	**0.05**	**0.01**	**4.15**	**<0.001**	**0.03**	**0.07**	plaur
tereri	**1.27**	**0.17**	**7.55**	**<0.001**	**0.94**	**1.60**	scarb2
ereri_cde	**1.25**	**0.17**	**7.15**	**<0.001**	**0.90**	**1.59**	scarb2
ereri_intref	0.00	0.00	−1.23	0.217	−0.01	0.00	scarb2
ereri_intmed	0.00	0.03	0.01	0.992	−0.06	0.06	scarb2

(*Continued on next page*)

Table 3. *Continued*

FOURWAYDECOMP	β or %	SE	z	P	LCL	UCL	PROTEIN
ereri_pie	**0.03**	**0.01**	**2.72**	**0.007**	**0.01**	**0.05**	scarb2
tereri	**1.25**	**0.16**	**7.64**	**<0.001**	**0.93**	**1.58**	lair1
ereri_cde	**1.21**	**0.17**	**7.15**	**<0.001**	**0.88**	**1.54**	lair1
ereri_intref	0.00	0.00	−0.28	0.777	−0.01	0.01	lair1
ereri_intmed	0.02	0.03	0.59	0.558	−0.04	0.08	lair1
ereri_pie	**0.03**	**0.01**	**2.43**	**0.015**	**0.00**	**0.05**	lair1
tereri	**1.30**	**0.17**	**7.83**	**<0.001**	**0.97**	**1.62**	ckap4
ereri_cde	**1.28**	**0.17**	**7.51**	**<0.001**	**0.95**	**1.61**	ckap4
ereri_intref	0.00	0.00	−1.24	0.215	−0.01	0.00	ckap4
ereri_intmed	−0.01	0.03	−0.31	0.756	−0.07	0.05	ckap4
ereri_pie	**0.03**	**0.01**	**2.9**	**0.004**	**0.01**	**0.05**	ckap4
tereri	**1.26**	**0.16**	**7.73**	**<0.001**	**0.94**	**1.58**	tnfrsf1b
ereri_cde	**1.23**	**0.17**	**7.38**	**<0.001**	**0.90**	**1.55**	tnfrsf1b
ereri_intref	0.00	0.00	−0.83	0.407	−0.01	0.00	tnfrsf1b
ereri_intmed	0.01	0.03	0.31	0.756	−0.05	0.06	tnfrsf1b
ereri_pie	**0.03**	**0.01**	**2.77**	**0.006**	**0.01**	**0.05**	tnfrsf1b
tereri	**1.30**	**0.17**	**7.76**	**<0.001**	**0.97**	**1.63**	cd302
ereri_cde	**1.28**	**0.17**	**7.37**	**<0.001**	**0.94**	**1.62**	cd302
ereri_intref	0.00	0.00	−1.37	0.169	−0.01	0.00	cd302
ereri_intmed	−0.01	0.03	−0.23	0.816	−0.07	0.06	cd302
ereri_pie	**0.03**	**0.01**	**3.05**	**0.002**	**0.01**	**0.05**	cd302
tereri	**1.33**	**0.17**	**7.7**	**<0.001**	**0.99**	**1.67**	epha2
ereri_cde	**1.28**	**0.18**	**7.21**	**<0.001**	**0.93**	**1.63**	epha2
ereri_intref	−0.01	0.01	−1.18	0.236	−0.02	0.01	epha2
ereri_intmed	0.00	0.03	0.09	0.926	−0.06	0.07	epha2
ereri_pie	**0.05**	**0.01**	**4.53**	**<0.001**	**0.03**	**0.07**	epha2
tereri	**1.26**	**0.16**	**7.7**	**<0.001**	**0.94**	**1.58**	havcr2
ereri_cde	**1.23**	**0.17**	**7.32**	**<0.001**	**0.90**	**1.56**	havcr2
ereri_intref	0.00	0.00	−0.92	0.356	0.00	0.00	havcr2
ereri_intmed	0.01	0.03	0.22	0.829	−0.06	0.07	havcr2
ereri_pie	**0.02**	**0.01**	**2.01**	**0.045**	**0.00**	**0.04**	havcr2
tereri	**1.24**	**0.16**	**7.6**	**<0.001**	**0.92**	**1.56**	tnfrsf9
ereri_cde	**1.24**	**0.17**	**7.37**	**<0.001**	**0.91**	**1.57**	tnfrsf9
ereri_intref	0.00	0.00	−0.5	0.615	−0.01	0.01	tnfrsf9
ereri_intmed	−0.02	0.03	−0.67	0.503	−0.08	0.04	tnfrsf9
ereri_pie	**0.03**	**0.01**	**2.59**	**0.01**	**0.01**	**0.04**	tnfrsf9
tereri	**1.29**	**0.16**	**7.84**	**<0.001**	**0.97**	**1.61**	tff3
ereri_cde	**1.23**	**0.17**	**7.39**	**<0.001**	**0.90**	**1.55**	tff3
ereri_intref	0.00	0.01	−0.03	0.979	−0.02	0.01	tff3
ereri_intmed	0.03	0.03	0.89	0.375	−0.03	0.08	tff3
ereri_pie	**0.03**	**0.01**	**3.79**	**<0.001**	**0.02**	**0.05**	tff3

ereri_cde, excess relative risk due to neither mediation nor interaction or controlled direct effect; ereri_intmed, excess relative risk due to mediated interaction or mediated interaction; ereri_intref, excess relative risk due to interaction only or interaction referent; ereri_pie, excess relative risk due to mediation only or pure indirect effect; pct_cde, percent of total effect that is controlled direct effect; pct_intmed, percent of total effect that is mediated interaction; pct_intref, percent of total effect that is interaction referent; pct_pie, percent of total effect that is pure indirect effect; tereri, Total excess relative risk; UK, United Kingdom. See Table S7 for some of these protein abbreviations. Others can be found at: https://www.ncbi.nlm.nih.gov/gene. Tereri and ereri_cde are interpreted as Log_e(hazard ratios). Bold figures represent findings with $p < 0.05$.

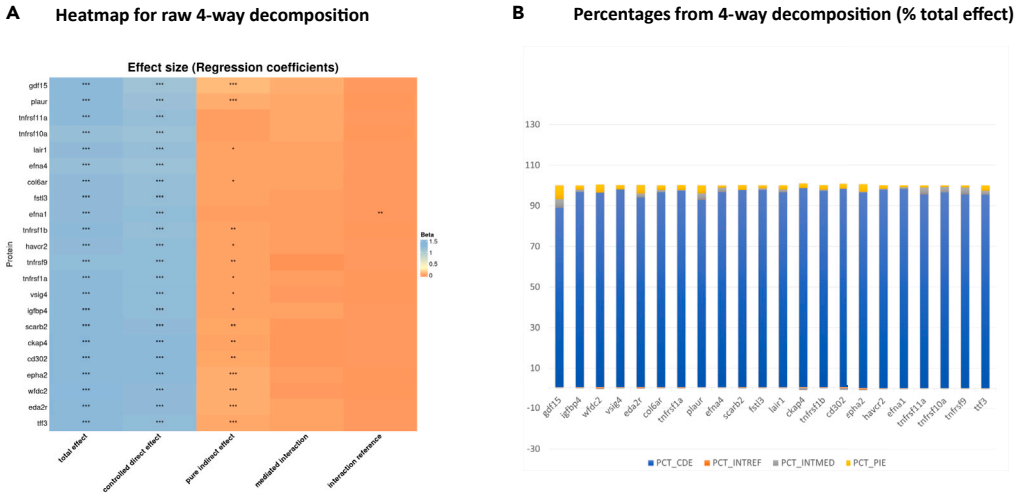


Figure 3. Four-way decomposition of the association between hospital-treated infections and incidence of all-cause dementia by the selected plasma proteomic biomarkers (k = 22): UK biobank 2006–2021

(A) Heatmap for raw 4-way decomposition. Effects could range from 0 to 1.5. *p < 0.05; **p < 0.010; ***p < 0.001.

(B) Percentages from 4-way decomposition (% of total effect). Proteins are ordered in ascending order of p value for the effect of hospital-treated infections on the protein after correction for multiple testing and extraction of 22 proteins based on effect sizes <-0.25 or >0.25 (gdf15 through ttf3). PCT_CDE (blue): percent of total effect that is a controlled direct effect; PCT_INTREF (orange): percent of total effect that is interaction referent; PCT_INTMED (gray): percent of total effect that is mediated interaction; PCT_PIE (yellow): percent of total effect that is pure indirect effect.

AD, Alzheimer's Disease; ereri_cde, excess relative risk due to neither mediation nor interaction or controlled direct effect; ereri_intmed, excess relative risk due to mediated interaction or mediated interaction; ereri_intref, excess relative risk due to interaction only or interaction referent; ereri_pie, excess relative risk due to mediation only or pure indirect effect; pct_cde, percent of total effect that is controlled direct effect; pct_intmed, percent of total effect that is mediated interaction; pct_intref, percent of total effect that is interaction referent; pct_pie, percent of total effect that is pure indirect effect; PRS, Polygenic Risk Score; tereri, Total excess relative risk; UK, United Kingdom. See Table S7 for protein abbreviations and also at https://www.ncbi.nlm.nih.gov/gene/.

Three recent studies used the UK Biobank to investigate illnesses and their links to cognitive decline and dementia. Muzambi et al. used ICD-10 codes to identify illnesses in linked primary care data over the prior 5 years, but no link was found between cognitive decline and brain volume.[4] Sipilä et al. focused on hospital-treated infections, which were operationalized across multiple dimensions such as quantity, severity, viral vs. bacterial.[3] As in our study, dementia risk increased >2-fold in the presence of hospital-treated infections.[3] Thus, such severe infections may have systematic influence on dementia.[3] Furthermore, a recent study indicated some moderation between infection burden and a measure of cardiovascular health, namely Life's Essential 8 in relation to brain imaging markers of neurodegeneration and white matter integrity.[18] Furthermore, secondary analyses showed that dominant groups of infections in the category "hospital-treated" included urinary tract infections and other genitourinary disorders, cellulitis and other bacterial infections, while predominantly "non-hospital treated" infections included vaccine-preventable infections (e.g., measles, mumps, pertussis) among others.[18] The lack of specificity for infection and dose-response effects suggests that systemic inflammation, rather than specific microorganisms, may be the major mechanism and proximal cause of dementia. Finally, Makli et al. studied illnesses identified by the UK Biobank, IgG data, and primary care linkage to determine dementia.[5] Only Herpes Simplex virus 1 was shown to be significantly associated to dementia diagnosis among the 15 illnesses studied.[5] The scientists also discovered that having HSV1, VZV1, HHV6, and HHV7 infections in contrast to having no infections was associated with dementia.[5]

Most importantly, in our recent analysis using all individuals aged 50 years or older who were dementia-free at baseline and with complete data on a similar set of covariates as for our present study, we also found that non-hospital treated infection group had a lower dementia risk compared with the "None" group, in addition to our main finding that hospital-treated infections versus all others and hospital-treated versus none were significantly associated with an increased risk for dementia.[6] This was mainly ascribed to the type of infections that were found to be most common in each of the two groups (i.e., hospital-treated vs. non-hospital treated).[6] In the larger sample, the most common infections in the hospital-treated were septicemia, staphylococcal and streptococcal infections, as well as "other bacterial infections", while in the non-hospital treated group, the most common infections were vaccine-preventable infections, most notably measles, mumps, rubella, and chicken pox (varicella), as well as other infections such as acute lower respiratory infection, acute tonsillitis, and infections of skin and subcutaneous tissue.[6] Given that this present study included a random sample of the larger study, it is expected that this ranking and classification of infections remains comparable.

Our study indicated that several elements in the proteome as well as several pathways are involved in mediating the association of hospital-treated infections with dementia. The top mediator was GDF15, a stress-responsive cytokine that is involved in the regulation of inflammation and cellular stress responses, with studies suggesting a key role played in the immune response to infection. Furthermore,

GDF15 has been associated with various disease processes, including cancer, cardiovascular disease, metabolic disorders, and more recently cognitive aging and dementia.[19] A review of the recent literature is presented in Table S6.

In short, GDF15 was upregulated in response to COVID-19, sepsis, HIV, hepatitis C, and periodontitis infections to name a few, with some limited evidence linking it to cognition, dementia and AD.[19–22] In addition to its well-studied roles in inflammation and metabolism, GDF15 has been identified in brain tissues and there is accumulating evidence that it is involved in several brain disorders. In adult rats, GDF15 is expressed in the central and peripheral nervous systems, and secreted in the cerebrospinal fluid where it can reach its target cells.[23] In mice, GDF15 is co-expressed with epidermal growth factor receptor in neural precursors to promote cell migration and proliferation and knocking out GDF15 in hippocampal precursors impairs both processes.[24] Altered neurogenesis in the hippocampus is an early and critical event in AD pathogenesis, suggesting a key role for GDF15 in AD. GDF15 in the hippocampus also plays a role in cerebral stroke and other brain disorders. Occlusion of the middle cerebral artery in a stroke model in mice, triggers an increase in GDF15 expression in the hippocampus.[25] GDF15 in cell culture studies can be both neurotrophic and neuroprotective in the midbrain, increasing survival after culture and protecting iron-intoxicated cultures of dopaminergic neurons, which degenerate during Parkinson's Disease.[23] These studies confirm the association of GDF15 with brain diseases, though also suggesting its role in regulating healing and prevention, rather than pathology.

Among top 17 mediators that were strongly related to hospital-treated infections, the elements that were previously and consistently shown to be associated with both infections and dementia, with at least 2 human studies in each category, included TFF3, TNFRSF1a, TNFRSF1b, CKAP4, PLAUR, WFDC2, LAIR1, CD302 and COL6A3. Among those, the literature-based strongest evidence, particularly in human studies, was for TFF3 (trefoil factor 3), a group of stable secretory proteins expressed in gastrointestinal mucosa. While their functions are not well-defined, these proteins are thought to protect the mucosa from insults, stabilize the mucus layer and affect healing of the epithelium. Consistent with our findings, another report found that higher plasma TFF3 levels were associated with cognitive impairment or dementia.[12] However, the role of infections was not assessed in this study. Among others, this protein was shown to be upregulated in periodontitis, *Helicobacter pylori (Hp)* infection, sepsis and various types of gastrointestinal infections.[26–28] A recent study indicated that *Hp* and periodontitis synergistically increased the risk for dementia in a national survey of US older adults.[29] Furthermore, at least 3 independent studies indicated that TFF3 was associated with dementia risk.[12,30,31] Our study indicated that 1 SD increase in TFF3 explained ~4.6% of the total effect of hospital-treated infections on dementia risk when combining both PIE and INTMED, with a statistically significant PIE explaining 2.6% of this total effect ($p < 0.001$). Furthermore, around 4% of this total effect was explained by 1 SD increase in EDA2R, a member of the type III transmembrane protein of the TNFR (tumor necrosis factor receptor) superfamily, which binds the EDA-A2 isoform of ectodysplasin, the latter playing an important role in maintenance of hair and teeth. This lends further support to the role played by periodontal disease in dementia risk. A comparable effect mediation (~3.8% PIE as % TE) with strong literature support was found for PLAUR, which was implicated in COVID-19, HIV, and other infections affecting pulmonary functions, and was upregulated with Aβ amyloid deposition (See Table S4). WFDC2, LAIR1, CD302, TFNRSF1a, TNFRSF1b, and CKAP4 were found to be important mediators between hospital-treated infections and dementia risk. These proteins were studied in relation to COVID-19 severity, pulmonary tuberculosis, fibrosis, *H. pylori* infection, autoimmunity, Hepatitis C infections, and inflammatory conditions. For all these proteins, there was at least moderately strong evidence of an association with dementia and its sub-types.

Our STRING analyses suggested that transmembrane receptor tyrosine kinase signaling pathway is the strongest and most significant pathway within the protein cluster that included GDF15 and several other top mediators of the association between hospital-treated infections and dementia risk. Other important pathways pertained to protein phosphorylation more generally. The hyperphosphorylation of serine and threonine residues in tau protein, a major factor in AD, may affect neurodegeneration.[32] Recent evidence suggests that tau phosphorylation can also occur on tyrosine residues.[32] The pathogenesis of AD tau and Aβ aggregates may be associated with the Abl family of tyrosine kinases, which regulate cytoskeleton cellular signaling cascades.[32] Proteostasis, or protein homeostasis, involves a highly complex interconnection of pathways that influence the fate of a protein from synthesis to degradation.[33] Targeting proteostasis with repurposed medications is one of the current research therapy options for treating AD.[33] Forty-seven trials using lithium, rapamycin, rifampicin, and tyrosine kinase inhibitors were included in a meta-analytic study.[33] Lithium microdosing showed a substantial advantage in both humans and animals.[33] Nevertheless, this review indicated that tyrosine kinase inhibitors were among the least studied drugs among those repurposed medications. Thus, more trials are required for this group of drugs to test their relative efficacy and safety in AD treatment. Hybrid compounds containing tacrine (a centrally acting acetylcholinesterase inhibitor and indirect cholinergic agonist), flavonoids, and medicinal plants have shown potential in AD treatment.[34] Curcumin is among traditional herbal remedies containing a substantial amount of tyrosine kinase inhibitors.[35] Curcumin has been recently shown to slow age-related cognitive decline in a meta-analysis of large controlled clinical trials, showing some protective effect on working memory domain among others.[36] Curcumin was also linked to improved brain pathologies associated with dementia.[37] Nevertheless, given the limited studies on tyrosine kinase inhibitors more generally, further trials are needed on a diverse set of drugs and nutraceuticals. Other drug targets that may block the effect of hospital-treated infections on dementia risk include, based on our findings, the regulation of leukocytes, gastrointestinal epithelial maintenance, regulation of the extracellular matrix, pyrimidine and glutathione metabolism, susceptibility to T cell mediated cytotoxicity, and positive regulation of I-kappaB kinase/NF-κB signaling among others.

In conclusion, hospital-treated infections were related to dementia incidence, an association that is weakly but partially mediated by several key protein mediators, with the primary driver being GDF15. Future studies should delve further in studying prevalent infection types formerly shown to predict dementia incidence and determine whether similar mechanisms are involved across types of infections and assess the validity of these findings in populations at greatest risk for dementia in the UK. Replication of our findings in comparable cohorts should also be coupled with *in vitro* and *in vivo* studies.

Limitations of the study

Our present study has several strengths and limitations. First, this study analyzed a large-scale proteomic analysis in a large cohort, and this is the first cohort study with enough power to investigate mediating and moderating effects of the plasma proteome in the relationship between hospital-treated infections and dementia. Second, UK Biobank investigators established the outcome variables using specific diagnosis dates obtained through record linkage. Third, the UK Biobank covers a wide range of topics, allowing for less biased estimations of exposure-outcome connections through confounder adjustment. Potential study limitations include selection bias due to missing data, measurement error caused by using self-report and ICD-10 codes, with a greater measurement bias experienced for the algorithmically defined AD diagnosis rather than the all-cause dementia diagnosis. Furthermore, the precise date of dementia incidence is uncertain. Although multiple confounders were considered, residual confounding is possible due to the observational design of this study and thus causal effects could not be inferred. Additionally, although prevalent cases of dementia were excluded up to three years after baseline assessment, reverse causality remains a possible explanation for the association between infection burden and dementia, with individuals with at least mild cognitive impairment prior to baseline assessment being at higher risk for infection. Nevertheless, on average, hospital-treated infections occurred over a decade prior to baseline assessment as shown in supplementary analyses. Furthermore, the results should be interpreted with caution in light of assumptions ascribed to four-way decomposition models.[38] Finally, the UK Biobank is not necessarily representative of the UK population and our analyses did not stratify by racial/ethnic groups due to the underpowered currently available sample for such analyses. Nevertheless, the sub-sample with proteomic data was representative of the larger sample of ~500K participants. Future studies with a larger sample UK Biobank sample should stratify the analysis by all major racial/ethnic groups.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Ethics statement
- METHOD DETAILS
  - Dementia outcomes
  - Infection burden
  - Olink proteomics
  - Covariates
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Descriptive statistics and Kaplan-Meier curves
  - Multiple regression modeling
  - Four-way decomposition models
  - Olink insight pathways and STRING analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108526.

ACKNOWLEDGMENTS

The authors would like to thank the UK biobank investigators, staff and participants, as well as the staff and analysts from the UK biobank access management system. This work was supported in part by the Intramural Research Program of the NIH, National Institute on Aging, National Institutes of Health project number AG000513. This work uses data provided by patients and collected by the NHS as part of their care and support. This research also used data assets made available by National Safe Haven as part of the Data and Connectivity National Core Study, le d by Health Data Research UK in partnership with the Office for National Statistics and funded by UK Research and Innovation (research which commenced between 1st October 2020–31st March 2021 grant ref MC_PC_20029; 1st April 2021 - 30th September 2022 grant ref MC_PC_20058). Importantly, this research has been conducted using the UK Biobank Resource under Application Number 77963.

AUTHOR CONTRIBUTIONS

M.A.B. had full access to the data used in this manuscript and completed all the statistical analyses.

M.A.B.: Study concept, plan of analysis, data management and statistical analysis, literature search and review, write-up of parts of the manuscript, revision of the manuscript.

H.A.B.: Study concept, data acquisition, plan of analysis, assistance with data management and statistical analysis, literature search and review, write-up of parts of the manuscript, revision of the manuscript.

N.N.H.: Study concept, plan of analysis, literature search and review, write-up of parts of the manuscript, review of the manuscript.

O.M.: Plan of analysis, data visualization, literature search and review, write-up of parts of the manuscript, revision of the manuscript.

Z.L.: Plan of analysis, assistance with data management, data visualization, write-up of parts of the manuscript, revision of the manuscript.

Z.W.E-H.: Literature search and review, write-up of parts of the manuscript, revision of the manuscript.

J.W.: Plan of analysis, assistance with data management, data visualization, write-up of parts of the manuscript, revision of the manuscript.

C.A.M.V.: Plan of analysis, assistance with data visualization, write-up of parts of the manuscript, revision of the manuscript.

L.J.L.: Plan of analysis, literature search and review, write-up of parts of the manuscript, revision of the manuscript.

M.K.E.: Data acquisition, plan of analysis, write-up of parts of the manuscript, revision of the manuscript.

A.B.Z.: Data acquisition, plan of analysis, write-up of parts of the manuscript, revision of the manuscript.

DECLARATION OF INTERESTS

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of AT Augusta Military Medical Center, the Defense Health Agency, Department of Defense, or US Government. Reference to any commercial products within this publication does not create or imply any endorsement by AT Augusta Military Medical Center, the Defense Health Agency, Department of Defense, or US Government. The authors declare no conflict of interest.

Received: July 21, 2023
Revised: September 18, 2023
Accepted: November 20, 2023
Published: November 23, 2023

REFERENCES

1. Butler, L., and Walker, K.A. (2021). The Role of Chronic Infection in Alzheimer's Disease: Instigators, Co-conspirators, or Bystanders? Curr. Clin. Microbiol. Rep. *8*, 199–212.
2. Ganz, T., Fainstein, N., and Ben-Hur, T. (2022). When the infectious environment meets the AD brain. Mol. Neurodegener. *17*, 53.
3. Sipilä, P.N., Heikkilä, N., Lindbohm, J.V., Hakulinen, C., Vahtera, J., Elovainio, M., Suominen, S., Väänänen, A., Koskinen, A., Nyberg, S.T., et al. (2021). Hospital-treated infectious diseases and the risk of dementia: a large, multicohort, observational study with a replication cohort. Lancet Infect. Dis. *21*, 1557–1567.
4. Muzambi, R., Bhaskaran, K., Rentsch, C.T., Smeeth, L., Brayne, C., Garfield, V., Williams, D.M., Chaturvedi, N., and Warren-Gash, C. (2022). Are infections associated with cognitive decline and neuroimaging outcomes? A historical cohort study using data from the UK Biobank study linked to electronic health records. Transl. Psychiatry *12*, 385.
5. Mekli, K., Lophatananon, A., Cant, R., Burns, A., Dobson, C.B., Itzhaki, R.F., and Muir, K.R. (2022). Investigation of the association between the antibody responses to neurotropic viruses and dementia outcomes in the UK Biobank. PLoS One *17*, e0274872.
6. Beydoun, H.A., Beydoun, M.A., Meirelles, O., Erickson, L.D., Gamaldo, A.A., Weiss, J., Launer, L.J., Evans, M.K., and Zonderman, A.B. (2023). Cardiovascular health, infection burden, and incident dementia in the UK Biobank. Alzheimers Dement. *19*, 4475–4487.
7. Coombs, K.M. (2020). Update on Proteomic approaches to uncovering virus-induced protein alterations and virus -host protein interactions during the progression of viral infection. Expert Rev. Proteomics *17*, 513–532.
8. Deutsch, E.W., Omenn, G.S., Sun, Z., Maes, M., Pernemalm, M., Palaniappan, K.K., Letunica, N., Vandenbrouck, Y., Brun, V., Tao, S.C., et al. (2021). Advances and Utility of the Human Plasma Proteome. J. Proteome Res. *20*, 5241–5263.
9. Harris, S.E., Cox, S.R., Bell, S., Marioni, R.E., Prins, B.P., Pattie, A., Corley, J., Muñoz Maniega, S., Valdés Hernández, M., Morris, Z., et al. (2020). Neurology-related protein biomarkers are associated with cognitive ability and brain volume in older age. Nat. Commun. *11*, 800.
10. Trares, K., Bhardwaj, M., Perna, L., Stocker, H., Petrera, A., Hauck, S.M., Beyreuther, K., Brenner, H., and Schöttker, B. (2022). Association of the inflammation-related proteome with dementia development at older age: results from a large, prospective, population-based cohort study. Alzheimer's Res. Ther. *14*, 128.
11. Whelan, C.D., Mattsson, N., Nagle, M.W., Vijayaraghavan, S., Hyde, C., Janelidze, S., Stomrud, E., Lee, J., Fitz, L., Samad, T.A., et al. (2019). Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. Acta Neuropathol. Commun. *7*, 169.
12. Tanaka, T., Lavery, R., Varma, V., Fantoni, G., Colpo, M., Thambisetty, M., Candia, J., Resnick, S.M., Bennett, D.A., Biancotto, A., et al. (2020). Plasma proteomic signatures predict dementia and cognitive impairment. Alzheimers Dement. *6*, e12018.
13. Lindbohm, J.V., Mars, N., Walker, K.A., Singh-Manoux, A., Livingston, G., Brunner, E.J., Sipilä, P.N., Saksela, K., Ferrie, J.E., Lovering, R.C., et al. (2022). Plasma proteins, cognitive decline, and 20-year risk of dementia in the Whitehall II and Atherosclerosis Risk in Communities studies. Alzheimers Dement. *18*, 612–624.
14. Beydoun, M.A., Beydoun, H.A., Elbejjani, M., Dore, G.A., and Zonderman, A.B. (2018). Helicobacter pylori seropositivity and its association with incident all-cause and Alzheimer's disease dementia in large national surveys. Alzheimers Dement. *14*, 1148–1158.
15. Beydoun, M.A., Beydoun, H.A., Hossain, S., El-Hajj, Z.W., Weiss, J., and Zonderman, A.B. (2020). Clinical and Bacterial Markers of Periodontitis and Their Association with Incident All-Cause and Alzheimer's Disease Dementia in a Large National Survey. J. Alzheimers Dis. *75*, 157–172.
16. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. *51*, D638–D646.
17. Honjo, K., van Reekum, R., and Verhoeff, N.P.L.G. (2009). Alzheimer's disease and infection: do infectious agents contribute to progression of Alzheimer's disease? Alzheimers Dement. *5*, 348–360.
18. Beydoun, M.A., Beydoun, H.A., Gale, S.D., Hedges, D., Weiss, J., Li, Z., Erickson, L.D., Noren Hooten, N., Launer, L.J., Evans, M.K., and Zonderman, A.B. (2023). Cardiovascular health, infection burden and their interactive association with brain volumetric and white matter integrity outcomes in the UK Biobank. Brain Behav. Immun. *113*, 91–103.
19. Jiang, W.W., Zhang, Z.Z., He, P.P., Jiang, L.P., Chen, J.Z., Zhang, X.T., Hu, M., Zhang, Y.K., and Ouyang, X.P. (2021). Emerging roles of growth differentiation factor-15 in brain disorders (Review). Exp. Ther. Med. *22*, 1270.
20. Tsai, V.W.W., Husaini, Y., Sainsbury, A., Brown, D.A., and Breit, S.N. (2018). The MIC-1/GDF15-GFRAL Pathway in Energy Homeostasis: Implications for Obesity, Cachexia, and Other Associated Diseases. Cell Metabol. *28*, 353–368.
21. Reyes, J., and Yap, G.S. (2023). Emerging Roles of Growth Differentiation Factor 15 in

Immunoregulation and Pathogenesis. J. Immunol. *210*, 5–11.
22. Beydoun, M.A., Noren Hooten, N., Weiss, J., Beydoun, H.A., Georgescu, M., Freeman, D.W., Evans, M.K., and Zonderman, A.B. (2023). GDF15 and its association with cognitive performance over time in a longitudinal study of middle-aged urban adults. Brain Behav. Immun. *108*, 340–349.
23. Strelau, J., Sullivan, A., Böttner, M., Lingor, P., Falkenstein, E., Suter-Crazzolara, C., Galter, D., Jaszai, J., Krieglstein, K., and Unsicker, K. (2000). Growth/differentiation factor-15/ macrophage inhibitory cytokine-1 is a novel trophic factor for midbrain dopaminergic neurons in vivo. J. Neurosci. *20*, 8597–8603.
24. Carrillo-García, C., Prochnow, S., Simeonova, I.K., Strelau, J., Hölzl-Wenig, G., Mandl, C., Unsicker, K., von Bohlen Und Halbach, O., and Ciccolini, F. (2014). Growth/ differentiation factor 15 promotes EGFR signalling, and regulates proliferation and migration in the hippocampus of neonatal and young adult mice. Development *141*, 773–783.
25. Schindowski, K., von Bohlen und Halbach, O., Strelau, J., Ridder, D.A., Herrmann, O., Schober, A., Schwaninger, M., and Unsicker, K. (2011). Regulation of GDF-15, a distant TGF-beta superfamily member, in a mouse model of cerebral ischemia. Cell Tissue Res. *343*, 399–409.
26. Keles Yucel, Z.P., Afacan, B., Atmaca İlhan, H., Kose, T., and Emingil, G. (2022). The trefoil factor family 1 (TFF-1) and 3 (TFF-3) are upregulated in the saliva, gingival crevicular fluid and serum of periodontitis patients. Oral Dis. *28*, 1240–1249.
27. Ramadan, R.A., Zaki, M.A., Ooda, S.A., Abo Khalifa, H.M., and Ragab, W.S. (2020). Comparison of Serum Trefoil Factor-3 to Endoscopy in Diagnosing Helicobacter Pylori Associated Gastric Ulcer. Asian Pac. J. Cancer Prev. *21*, 2149–2153.
28. Meijer, M.T., Uhel, F., Cremer, O.L., Scicluna, B.P., Schultz, M.J., and van der Poll, T.; MARS consortium (2020). Elevated trefoil factor 3 plasma levels in critically ill patients with abdominal sepsis or non-infectious abdominal illness. Cytokine *133*, 155181.
29. Beydoun, M.A., Beydoun, H.A., Weiss, J., Hossain, S., El-Hajj, Z.W., and Zonderman, A.B. (2021). Helicobacter pylori, periodontal pathogens, and their interactive association with incident all-cause and Alzheimer's disease dementia in a large national survey. Mol. Psychiatr. *26*, 6038–6053.
30. Zou, J., Chen, Z., Liang, C., Fu, Y., Wei, X., Lu, J., Pan, M., Guo, Y., Liao, X., Xie, H., et al. (2018). Trefoil Factor 3, Cholinesterase and Homocysteine: Potential Predictors for Parkinson's Disease Dementia and Vascular Parkinsonism Dementia in Advanced Stage. Aging Dis. *9*, 51–65.
31. Paterson, R.W., Bartlett, J.W., Blennow, K., Fox, N.C., Alzheimer's Disease Neuroimaging Initiative, Shaw, L.M., Trojanowski, J.Q., Zetterberg, H., and Schott, J.M. (2014). Cerebrospinal fluid markers including trefoil factor 3 are associated with neurodegeneration in amyloid-positive individuals. Transl. Psychiatry *4*, e419.
32. Lebouvier, T., Scales, T.M.E., Williamson, R., Noble, W., Duyckaerts, C., Hanger, D.P., Reynolds, C.H., Anderton, B.H., and Derkinderen, P. (2009). The microtubule-associated protein tau is also phosphorylated on tyrosine. J. Alzheimers Dis. *18*, 1–9.
33. Heard, D.S., Tuttle, C.S.L., Lautenschlager, N.T., and Maier, A.B. (2018). Repurposing Proteostasis-Modifying Drugs to Prevent or Treat Age-Related Dementia: A Systematic Review. Front. Physiol. *9*, 1520.
34. Simunkova, M., Alwasel, S.H., Alhazza, I.M., Jomova, K., Kollar, V., Rusko, M., and Valko, M. (2019). Management of oxidative stress and other pathologies in Alzheimer's disease. Arch. Toxicol. *93*, 2491–2513.
35. Golonko, A., Lewandowska, H., Świsłocka, R., Jasińska, U.T., Priebe, W., and Lewandowski, W. (2019). Curcumin as tyrosine kinase inhibitor in cancer treatment. Eur. J. Med. Chem. *181*, 111512.
36. Tsai, I.C., Hsu, C.W., Chang, C.H., Tseng, P.T., and Chang, K.V. (2021). The Effect of Curcumin Differs on Individual Cognitive Domains across Different Patient Populations: A Systematic Review and Meta-Analysis. Pharmaceuticals *14*, 1235.
37. Reddy, P.H., Manczak, M., Yin, X., Grady, M.C., Mitchell, A., Tonk, S., Kuruva, C.S., Bhatti, J.S., Kandimalla, R., Vijayan, M., et al. (2018). Protective Effects of Indian Spice Curcumin Against Amyloid-beta in Alzheimer's Disease. J. Alzheimers Dis. *61*, 843–866.
38. Discacciati, A., Bellavia, A., Lee, J.J., Mazumdar, M., and Valeri, L. (2018). Med4way: a Stata command to investigate mediating and interactive mechanisms using the four-way effect decomposition. Int. J. Epidemiol. *48*, 15–20.
39. Ho, F.K., Gray, S.R., Welsh, P., Gill, J.M.R., Sattar, N., Pell, J.P., and Celis-Morales, C. (2022). Ethnic differences in cardiovascular risk: examining differential exposure and susceptibility to risk factors. BMC Med. *20*, 149.
40. UK Biobank (2022). Algorithmically-defined Outcomes (ADOs). https://biobank.ctsu.ox.ac.uk/crystal/ukb/docs/alg_outcome_main.pdf.
41. Ronaldson, A., Arias de la Torre, J., Sima, R., Ashworth, M., Armstrong, D., Bakolis, I., Hotopf, M., and Dregan, A. (2022). Prospective associations between depression and risk of hospitalisation for infection: Findings from the UK Biobank. Brain Behav. Immun. *102*, 292–298.
42. Andersson, N.W., Goodwin, R.D., Okkels, N., Gustafsson, L.N., Taha, F., Cole, S.W., and Munk-Jørgensen, P. (2016). Depression and the risk of severe infections: prospective analyses on a nationwide representative sample. Int. J. Epidemiol. *45*, 131–139.
43. Wik, L., Nordberg, N., Broberg, J., Björkesten, J., Assarsson, E., Henriksson, S., Grundberg, I., Pettersson, E., Westerberg, C., Liljeroth, E., et al. (2021). Proximity Extension Assay in Combination with Next-Generation Sequencing for High-throughput Proteome-wide Analysis. Mol. Cell. Proteomics *20*, 100168.
44. Petrera, A., von Toerne, C., Behler, J., Huth, C., Thorand, B., Hilgendorff, A., and Hauck, S.M. (2021). Multiplatform Approach for Plasma Proteomics: Complementarity of Olink Proximity Extension Assay Technology to Mass Spectrometry-Based Protein Profiling. J. Proteome Res. *20*, 751–762.
45. Townsend P, P.P., and Beattie, A. (1987). Health and Deprivation: Inequality and the North (Croom Helm).
46. Chadeau-Hyam, M., Bodinier, B., Vermeulen, R., Karimi, M., Zuber, V., Castagné, R., Elliott, J., Muller, D., Petrovic, D., Whitaker, M., et al. (2020). Education, biological ageing, all-cause and cause-specific mortality and morbidity: UK biobank cohort study. EClinicalMedicine *29–30*, 100658.
47. Lloyd-Jones, D.M., Hong, Y., Labarthe, D., Mozaffarian, D., Appel, L.J., Van Horn, L., Greenlund, K., Daniels, S., Nichol, G., Tomaselli, G.F., et al. (2010). Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. Circulation *121*, 586–613.
48. Hayman, L.L., and Martyn-Nemeth, P. (2022). A New Metric for Promoting Cardiovascular Health: Life's Essential 8. J. Cardiovasc. Nurs. *37*, 509–511.
49. Lloyd-Jones, D.M., Allen, N.B., Anderson, C.A.M., Black, T., Brewer, L.C., Foraker, R.E., Grandner, M.A., Lavretsky, H., Perak, A.M., Sharma, G., et al. (2022). Life's Essential 8: Updating and Enhancing the American Heart Association's Construct of Cardiovascular Health: A Presidential Advisory From the American Heart Association. Circulation *146*, e18–e43.
50. Bradbury, K.E., Young, H.J., Guo, W., and Key, T.J. (2018). Dietary assessment in UK Biobank: an evaluation of the performance of the touchscreen dietary questionnaire. J. Nutr. Sci. *7*, e6.
51. Ioachimescu, O.C. (2022). From Seven Sweethearts to Life Begins at Eight Thirty: A Journey From Life's Simple 7 to Life's Essential 8 and Beyond. J. Am. Heart Assoc. *11*, e027658.
52. VanderWeele, T.J. (2014). A unification of mediation and interaction: a 4-way decomposition. Epidemiology *25*, 749–761.
53. Ikram, M.A., and VanderWeele, T.J. (2015). A proposed clinical and biological interpretation of mediated interaction. Eur. J. Epidemiol. *30*, 1115–1118.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
UK Biobank data	https://www.ukbiobank.ac.uk	N/A
Algorithmically defined dementia outcome	alg_outcome_main.pdf (ox.ac.uk)	N/A
UK Biobank Olink proteomics	https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=1839	N/A
AD PRS	https://www.medrxiv.org/content/10.1101/2022.06.16.22276246v1.supplementary-material?versioned=true	N/A
UK Biobank showcase	: Showcase Homepage (ox.ac.uk)	N/A
UK Biobank Research Analysis Platform	UK Biobank Research Analysis Platform	N/A
Software and algorithms		
Stata software release 18	http://www.stata.com	N/A
Med4way command	https://github.com/anddis/med4way	N/A
Four-way decomposition supplement	UKB_paper8_supplementarydata/FOURWAY_DECOMP at main · baydounm/UKB_paper8_supplementarydata (github.com)	N/A
Parmby command	http://www.stata.com	N/A
Multproc command	http://www.stata.com	N/A
R version 4.3.2	http://www.r-project.org	N/A
STRING database	https://string-db.org	N/A
OLINK insight pathway	https://insight.olink.com	N/A
Github repository for this paper	https://github.com/baydounm/UKB-paper8-supplementarydata	N/A

RESOURCE AVAILABILITY

Lead contact

The lead contact is the corresponding author, Dr. May A. Beydoun who can be contacted via e-mail at baydounm@mail.nih.gov for inquires related to type of data and the methodology used in this work.

Materials availability

No materials were produced in this study.

Data and code availability

- Data: The data analyzed in this study is subject to the following licenses/restrictions: UK Biobank is a large-scale biomedical database and research resource, containing in-depth genetic and health information from half a million United Kingdom participants. The database is regularly augmented with additional data and is globally accessible to approved researchers undertaking vital research into the most common and life-threatening diseases. Requests to access these datasets should be directed to https://www.ukbiobank.ac.uk/. Therefore, data used in this study cannot be directly shared with other researchers.
- Code: Statistical code and other relevant resources to this current work can be accessed via GitHub repository at: https://github.com/baydounm/UKB-paper8-supplementarydata. Additional inquiries can be directed to the corresponding author.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Ethics statement

The studies involving human participants were reviewed and approved by the UK Biobank which has approval from the Institutional Review Boards, namely, the North West Multi-center Research Ethics Committee for the United Kingdom, from the National Information Governance Board for Health and Social Care for England and Wales, and from the Community Health Index Advisory Group for Scotland. All participants gave informed consent for the study via a touch-screen interface that required agreement for all individual statements on the consent form as well as the participant's signature on an electronic pad. Written informed consent for participation was not required for this study in accordance with the National Legislation and the Institutional Requirements.

The UK Biobank is a prospective cohort study involving >500,000 people in the UK between the ages of 37 and 73 who were recruited between 2006 and 2010.[39] Study purpose and design are described elsewhere.[39] Participants were examined at 22 assessment centers in England, Scotland, or Wales and completed a questionnaire and a face-to-face interview.[39] Participants were assessed by trained personnel who took phenotypic measurements and biological samples.[39] Sequential survey administration occurred during an assessment center visit using touch-screen self-completed questionnaire followed by a computer-assisted personal interview which was interviewer administered if needed. The UK Biobank was approved by the Northwest Multi-Centre Research Ethics Committee. Approval of this project was obtained from the UK Biobank access management team (application #77963), and by the Institutional Review Board of the National Institutes of Health.

Of the initial larger sample of 502,366 participants, 384,607 were aged ≥50 years at recruitment, while 347,250 had complete covariates of interest. Proteomic data was available for up to 37,316 participants, and the largest analytic sample consisted of 37,269 who were dementia-free at baseline (Figure 1).

METHOD DETAILS

Dementia outcomes

We removed participants whose age of occurrence was smaller than their baseline age from the algorithmically generated dementia outcomes (fields 42018 and 42020).[40] For incident AD diagnosis, the method employed ICD-10 codes F00 or G30, although numerous codes were used for all-cause dementia, including vascular dementia (F01, I67.3), notably A81.0, F00, F01, F02, F03, F05, G30, G31.0, G31.1, G31.8, and I67.3. The date of the first incidence of all-cause dementia was determined by selecting the smallest of numerous date variables/fields available for this outcome.[40] In a sensitivity analysis, subjects with follow-up durations not exceeding 3 years were excluded to eliminate possible inclusion of prevalent dementia cases.

Infection burden

Earliest date of occurrence of infections were used to obtain an infection burden index, that was computed for overall hospital-treated infections, as was done in a previous study.[41] The latter study provided a list of viral and bacterial infections included in the index. It is worth noting that the list of ICD-10 codes included can further be sub-categorized as central nervous system (CNS) infections (A17, A80-A81, A85-A89, B00.3-B00.4, B01.0-B01.1, B02.0-B02.2, B05.0-B05.1, B06.0, B2.61-B26.2, G00-G01, G02.0, G03, G04.2, G05.0-G05.1,e.g., meningitis, viral encephalitis), gastrointestinal infections (A00-A05, A08, e.g., salmonella, shigellosis), liver infections (B15-B19, e.g., hepatitis A), respiratory infections (A15-A16, A36-A38, J00-J06, J09-J18, J20-J22, e.g., pneumonia, laryngitis), sepsis (A40-A41, e.g., streptococcal sepsis), skin infections (A46, B00-B09, L00-L05, L08, e.g., cellulitis, measles), urogenital infections (N30.0, N39.0, N41.0–41.1, N71-N72, e.g., cystitis, prostatitis), and other infections (A18-A19, A31-A32, A39, A42-A44, A48-A49, B25-B27, B30, B33-B34, B95-B98, H62.0-H62.1, H67.0-H67.1, M00, M01.0-M01.5, N61, e.g., bone infection, mastitis).[41,42]

In summary, a primary or secondary infection diagnosis was determined using linked hospital admission records. The ICD-10 codes are listed in Tables S1 and S2. A count of infection types was used to reflect prevalent infection burden, total and hospital-treated, with dates of occurrence restricted to being prior to start dates at baseline and data sources subdivided into hospital-treated and non-hospital treated (i.e., from other sources, mainly self-report and/or primary care). A three-level exposure: 0 = None, 1 = Non-hospital treated and 2 = hospital-treated infections was created to group UK Biobank participants according to prevalent infection status. Nevertheless, the main exposure was 0 = None or non-hospital-treated vs. 1 = Hospital-treated infection, (UKB showcase: https://biobank.ndph.ox.ac.uk/showcase/search.cgi).

Olink proteomics

As part of the UK Biobank Pharma Proteomics Project (UKB-PPP), proteomic analysis was performed on 54,306 plasma samples from unique UK Biobank participant-visits using Olink Explore 1536 Proteomics platform, an approximately 12% random sample of the full UK Biobank study.[18] This platform uses Proximity Extension Assay (PEA) technology and 1,472 protein analytes were quantified, which correspond to 1,463 unique proteins from inflammation, oncology, cardiometabolic and neurological panels.[43] Details can be found in Sun et al.[18] and described in detail below. A few previous studies have also been conducted to assess reproducibility between Olink and other platforms (e.g.,[44]).

The assay

Plasma samples were serially diluted to 1:10, 1:100, 1:1000 and assayed in four 384-well plates, which are made up of four abundance blocks for each of the four different panels per 96 samples. After overnight incubation at 4°C of plasma samples with proximity probes, the respective oligonucleotides that were in close proximity were extended and amplified using DNA polymerase. This step generates a DNA sequence which was amplified by polymerase chain reaction (PCR 1), creating amplicons that contain protein assay information. The total amplicons generated for each sample from the four abundance groups per panel are merged, the end result being that there is one well of amplicons for each sample per panel. Index plates specific for each panel were added to sample plates, and subsequently a second PCR step (PCR 2) was performed that allows for samples in each plate to be pooled into one library per panel. The libraries then undergo bead purification and quality control was analyzed using a Bioanalyzer. Samples were sequenced on a Novaseq600 using S4 flow cells v1.5 (35 cycles) and sequence counts were translated into Normalized Protein eXpression (NPX) values within Olink's MyData Cloud Software.

Olink quality controls

Olink has quality controls built-in throughout the workflow. These include 3 spike-in engineered internal controls for every sample, and each abundance block. One control is the incubation control (Inc Ctrl), a green fluorescent protein (GFP), which was utilized for data QC. Extension controls were used for data normalization. Amplification controls (Amp Ctrl), that utilize a synthetic double stranded DNA template, were used for monitoring and QC for the PCR portion of the procedure. External controls are run on each plate as well. Triplicate negative controls were used on each plate to calculate the assay limit of detection (LOD), a pooled plasma sample was run also in triplicate as a plate control sample. In addition, a pooled sample control was run in duplicate for estimation of run precision. Three different proteins, IL6, IL8 (CXCL8), and TNF, were included in each of the 4 panels for quality control (QC) purposes and to analyze the correlation. The QC assessment was analyzed both during run QC and sample QC. Olink's standard procedures were used to generate data into NPX, Olink's relative quantification unit, which is on a log-2 scale. This procedure includes normalization of matched counts of an assay to the extension control which has been spiked into every sample, the values are then log-2 transformed, and the level was adjusted using the plate control.

NPX calculation and normalization

Samples were run from two different sets: Set 1 – UKB; and ii) Set 2 – COVID and within each set samples were randomized on plates. NPX was calculated by first taking the log2 ratio of counts of each assay of each sample to the extension control counts. Next, the plate control assay-specific median value was subtracted. This was used as the plate normalized NPX values for both sets. For set 1 samples, the assay-specific plate median NPX value was subtracted, and the batch-specific median NPX value of each assay was added to adjust for within batch effects. At this part of the procedure the data was normalized within each batch. Adjustment factors were then computed from the difference of the assay specific median NPX value of each batch to the reference batch (which was batch 1). Additional adjustment factors were then added to the NPX values of each batch of set 1. Thus, normalization of set 1 was both to within-batch and across-batches intensity normalization. Set 2 samples were normalized using reference samples that were shared between the two different sets. The plates with one sample at least of set 2 were allocated a selected set 1 sample randomly chosen. Samples (n = 93) from batches 1–6 of set 1 were chosen with missing frequency <10% and representative of the dynamic range of NPX values. These specific samples were then run on empty wells of the 93 plates from batch 7 of set 2. Specific adjustments were computed then from the assay-specific median of the pairwise differences between set 1 and set 2. Additional factors were used for adjustment of set 2 NPX values. The final set of NPX values consisted of set 1 intensity normalized NPX values and reference normalized NPX values for set 2.

Data pre-processing and quality checking

The initial UKB-Olink dataset consisted of 58,699 samples from 54,309 individuals. Participant samples were then excluded if they had withdrawn from the study or the samples were unprocessed, 58,362 samples and 54,306 individuals remained. Samples were then excluded due to QC failures (including missing NPX values) leaving a total of 58,360 samples from 54,304 individuals.

Two approaches were used to identify outliers: principal component analysis (PCA) and analyzing the median and IQR of NPX values across proteins by sample. Data points were moved if either the PC1 or PC2 values were >5 standard deviations (SD) from the mean or if a median NPX was >5 SDs from the mean of the median, or an IQR of NPX was >5 SD from the mean of the IQR.

Subsequently, after the removal of outliers, data points with a QC or assay warning were excluded. Therefore, there were 58,240 samples and 54,189 individuals that were included in the dataset.

Duplicate samples were run to calculate intra-individual coefficient of variation for each protein and ranged from 2.4% to 25%. Three different proteins (CXCL8, IL6, TNF) were run on all four panels (Cardiometabolic, Inflammation, Neurology and Oncology). The mean correlations of these proteins across all four panels were r = 0.96 for CXCL8 (range: 0.95–155 0.98), r = 0.92 for IL6 (range: 0.88–0.95) and r = 0.81 for TNF (range: 0.79–0.84). Batch and plate effects were also analyzed but there was no detectable evidence for either.

Covariates

Socio-demographic and socio-economic factors

Directed Acyclic graphs were used to determine whether several covariates should be selected among potential confounders. Age, sex, race/ethnicity (White, Black, South Asian, and Others), and household size were potential socio-demographic confounders. Three indicators of socio-economic status were educational achievement, household income and Townsend deprivation index (TDI).[45] A touch-screen questionnaire collected baseline information on educational attainment, which was re-grouped based on previous research[46] as 0 = Low, combining "None", "CSEs/Equivalent", "NVQ/HND/HNC/Equivalent" and "Other professional qual"; 1 = Intermediate, combining "O Levels/GCSEs/Equivalent" and "A/AS Levels Equivalent"; 2 = Higher level or "College/University." Total household income before tax was categorized into 1 = "< £18,000", 2 = "£18,000–£29,999", 3 = "£30,000–£51,999", 4 = "£52,000–£100,000" and 5 = ">£100,000". Based on national census data, TDI ratings were calculated, considering residential postcode-level car ownership, home overcrowding, owner occupation, and unemployment. Higher TDI scores suggest greater socioeconomic deprivation.[45] TDI was therefore multiplied by −1 to reflect higher SES and combined with z-scores of educational attainment and household income into one SES summary score.

Life's essential 8

In 2010, the American Heart Association (AHA) widened its scope of interest by prioritizing wellness over illness, by defining a new measure of cardiovascular health (CVH) aiming at individual and population-level health promotion.[47,48] CVH was defined with 7 potentially modifiable biological and lifestyle factors, that when at optimal levels, would result in greater cardiovascular disease (CVD)-free survival, longevity, and better quality of life. This measure of CVH was labeled "Life's Simple 7" (LS7), with its 7 components including indicators of diet quality, greater physical activity, reduced cigarette smoking, lower body mass index, total cholesterol, fasting blood glucose, and optimal blood pressure levels. Using clinical thresholds that were accepted for both children and adults, each metric was categorized as poor (0), intermediate (1), or ideal (2). The overall summary score of LS7 could range from 0 (all metrics at poor levels) to 14 (all 7 metrics at ideal levels).[47,48] Since 2010 AHA statement was published, CVH was re-evaluated and an AHA Presidential Advisory proposed an enhanced version of CVH, reflecting advances made over a decade of research, while LS7's methodological limitation have been remedied.[48,49] This new measure was labeled "Life's Essential 8" (LE8), retaining all 7 components of LS7 with major modifications to definitions and scales (described below). Sleep health was added to generate LE8 given its known influence on CVH across the life span,[48,49] (Table S3, for detailed algorithm used to generate LE8 total score).

Going further in detail regarding the dietary component of LE8, the touchscreen questionnaire of the UKB main study included twenty-nine questions regarding diet and eighteen questions related to alcohol. The touchscreen questionnaire inquired about food consumption frequency and nature, over the past year of the following food groups: cooked vegetables, salad/raw vegetables, fresh fruit, dried fruit, oily fish, other fish, processed meats, poultry, beef, lamb, pork, cheese, salt added to food, tea, water, as well as questions on the type of milk most commonly consumed, type of spread most commonly consumed, number of slices and type of bread most commonly consumed, number of bowls and type of breakfast cereal most commonly consumed, cups of coffee and type most commonly consumed, as well as questions on the avoidance of specific foods and food groups (eggs, dairy products, wheat, sugar), age last ate meat (for participants who reported never consuming processed meats, poultry, beef, lamb or pork), temperature preference of hot drinks, changes in diet in the past 5 years, and variation in diet. Four of the dietary questions originally utilized in the pilot trial were slightly altered for the main assessment phase: these were the items related to avoiding specific foods and food groups; spread type; bread type; and variation in diet.

The Healthy Diet Index (HDI) score combined several food groups in terms of quantity and frequency of consumption per week, when available to reflect the guidelines listed in Table S4. However, those criteria were modified to fit the availability of data in the UK biobank. Table S5 represents the food groups that were selected, their respective coding scheme and the scoring system to reflect better diet quality, approximating the criteria in Table S4. The touchscreen questionnaire was later validated against the 24-h recall that was administered over time to UK biobank participants and has shown adequate agreement in terms of ranking for each food group of interest.[50]

In summary, a composite measure of cardiovascular health known as Life's Essential 8 (LE8)[51] by the American Heart Association was included among potential confounders (supplementary method 2 and Tables S3–S5).

AD polygenic risk score

PRS scores were created and applied to meta-analyzed (and, whenever possible, ancestry specific) GWAS summary statistics that were either completely extracted from external GWAS data (the Standard PRS set) or from an amalgamation of external and internal UK Biobank data (the Enhanced PRS set) using a Bayesian approach. The Standard PRS Set (also referred to as the "UKB-Free" set), which consists of 28 diseases and 8 quantitative traits, was created using external GWAS data; the method was described in the supplementary material for the main paper by Thompson et al. in 2022 (https://www.medrxiv.org/content/10.1101/2022.06.16.22276246v1.supplementary-material?versioned=true). We choose AD PRS, which was originally found in the PGS catalog (https://www.pgscatalog.org), from the set of standard PRS. A version of the AD PRS with APOE was chosen by us. In summary, AD PRS was partly included as potential effect modifier.

QUANTIFICATION AND STATISTICAL ANALYSIS

All analyses were conducted using Stata 18.0 (StataCorp, College Station, TX).

Descriptive statistics and Kaplan-Meier curves

As a first step, descriptives including means ± standard errors (SE) and proportions of study sample characteristics, including key exogenous covariates, effect modifiers such as AD PRS, the main exposure of interest, and incidence proportions of all-cause and AD dementia, were calculated overall and stratified by sex. We further compared groups using linear, logistic, and multinomial logit models, comparing means and proportions of these key variables by sex. As a second stage, time-to-event was defined from age at entry ≥50years (i.e., delayed entry) until age of exit, defined by either age at event, or age at censoring (death or end of follow-up on October 31st, 2021). Age at baseline was estimated to the nearest month and year of birth and time to event was expressed as years elapsed since baseline assessment. Using this survival time setup, Kaplan-Meier survival rates for dementia-free survival across age, were estimated and compared across a three-level exposure of interest: 0 = None; 1 = non-hospital-treated infections; 2 = Hospital-treated infection, as an attempt to replicate previous findings from a larger UK Biobank cohort analysis.

Multiple regression modeling

Third, we constructed Cox proportional hazards (PH) models after evaluating the proportionality assumption, to test associations of hospital-treated infection burden with all-cause dementia, and adjusting for baseline age, sex, race/ethnicity, household size, SES *Z* score and LE8 total score. This part of the analysis was carried out overall, by sex and stratified by AD PRS, overall and among men and women, separately. Fourth, hospital-treated infections were entered into a series of separate multiple linear regression models with outcomes being the 1,463 plasma proteomic biomarkers, while adjusting for the same covariates as above (Stata *parmby* command). A volcano plot was constructed using R *ggplot* command, to illustrate the p values and effect sizes of each of these 1,463 equations and display the ones that passed multiple testing using Bonferroni correction (*parmby* and *multproc* commands in Stata). Of those, the top hits with >0.25 effect size in absolute value (corresponding to $^1/_4$ SD higher plasma protein value among the hospital-treated infection group vs. all others) were selected among those who passed Bonferroni correction. A weaker effect size threshold was chosen if needed, corresponding the top 98th percentile (or the 1st and 99th percentile on either side), with the goal of reducing the number of selected proteins to <30.

Four-way decomposition models

Fifth, those select biomarkers were then entered into a 4-way decomposition model with main exposure being hospital-treated infections, the main outcome being time to dementia using a Cox PH model for the final equation, and the mediator being one of each selected proteomic markers. Those makers were also allowed to be potential moderators.[52] Thus, in addition to the total effect, tereri, with standard error (SE) and p value, interpreted as Log_e(Hazard Ratios), four parameters decomposing the total effect according to mediation and interaction were also estimated with their SE, 95% CI and p value, namely ereri_cde (controlled direct effect: neither mediation nor interaction), ereri_interef (interaction referent: interaction only), ereri_intmed (mediated interaction: both mediation and interaction), ereri_pie (pure indirect effect: mediation only). The mediated interaction refers to when the exposure affects the outcome in the presence of the mediator and the presence of the exposure is required for the mediator to be present. The pure indirect effect (PIE) is similar but has one critical difference, which is that the mediator has an effect on the outcome in the absence of the exposure.[53] The sum of the mediated interaction and PIE yields the total indirect effect, which captures the interaction between the exposure and mediator on the outcome. The exposure-mediator model was based on a multivariable adjusted linear regression model whereas the exposure/mediator-outcome was a series of Cox PH models adjusted for the same covariates. Details related to *Med4way* command [https://github.com/anddis/med4way] and related methods are provided elsewhere.[38] We refer the readers to https://github.com/baydounm/UKB-paper8-supplementarydata under FOURWAY_DECOMP folder for more details regarding estimation of parameters. Findings are presented in tabular format and illustrated using a heatmap. Top selected mediators in the infection-dementia relationship are highlighted in terms of their function and connection to infection in previous studies.

Sixth, among the selected plasma proteins previously found to be strongly related to hospital-treated infections, a sub-set was entered into a principal components analysis (PCA) in order to reduce the data to a few components consisting of correlated plasma proteins, if they were statistically significant mediators with a positive or inverse PIE, at type I error of 0.05. The purpose of this step was to determine which clusters of significantly mediating proteins had the largest percent mediation in the infection-dementia association or if singular, determine that overall percentage among correlated plasma proteins. The number of components extracted was determined using the Kaiser rule (eigenvalue >1) and when multiple components were extracted, they underwent orthogonal rotation using varimax, for ease of interpretation. Using the regression method, those PCA scores (z-scores) were predicted and entered into another set of 4-way decomposition models to assess the extent of mediation and/or moderation by these components in the total effect of hospital-treated infections on incident dementia. In this analysis, as in earlier stage of the analytic plan, sex was the main stratifying variable and regression models were conducted overall and among men and women, separately. AD PRS was used as a secondary stratifying variable to assess mediating/moderating effects of PCA scores on the infection-dementia association across AD polygenic risk tertiles.

Olink insight pathways and STRING analyses

Finally, findings for significant pure indirect effects (PIE) across the entire proteome (i.e., k = 1,463 proteins) is also presented in supplementary materials and entered into Olink insight pathways browser to determine the most common pathways involved among those mediators (https://github.com/baydounm/UKB-paper8-supplementarydata) and visualized as a set of independent and connected pathways. A supplementary datasheet lists those detailed pathways. Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database analysis (STRING: functional protein association networks (string-db.org)) was also carried out in order to visualize all key protein mediators in our four-way decomposition analysis and their relationships with each other.[16] These proteomic mediators were also clustered into 10 k-means clusters in order to visualize their relationship in a more meaningful way. The cluster containing the top mediator was visualized more closely, where applicable, and the association of the various proteins in that cluster with exposure and outcome were examined closely using results from the four-way decomposition models, adjusting for all exogenous variables in those models. Furthermore, where possible, gene ontology (GO) enrichment analysis (biological processes or cellular component if former not represented) was also conducted and explored within each cluster, with detailed findings shown for the top cluster.[16] When GO was not available, it was replaced by the local network cluster (STRING) term (CL).[16]