Inflammatory diseases, inflammatory biomarkers, and Alzheimer's disease: an observational analysis and Mendelian randomisation study

Supplement

eFigure 1. Alzheimer's disease (AD) cases diagnosis algorithm.

eFigure 2. Study design.

eFigure 3. Risk differences curves for AD onset and death, estimated from the cause-specific hazards using Clinical Practice Research Datalink.

eFigure 4. Cumulative incidence function (CIF) curves for dementia onset and for death using Clinical Practice Research Datalink.

eFigure 5. Risk differences curves for dementia onset and death, estimated from the cause-specific hazards using Clinical Practice Research Datalink.

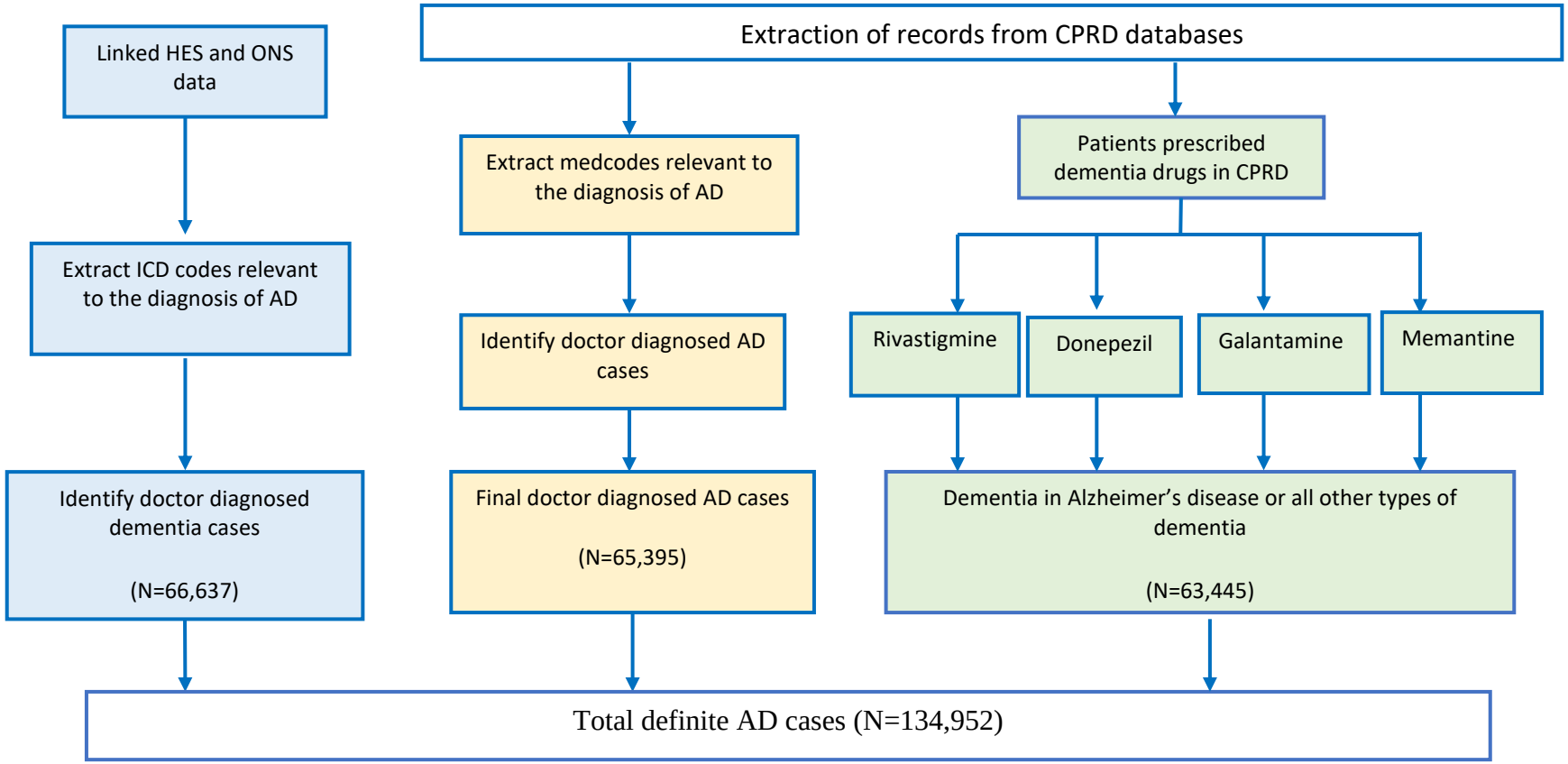
eFigure 6. Associations between circulating MIP1b level and rheumatoid arthritis (RA) using two-sample Mendelian randomisation (MR). (a) MIP1b→RA; (b) MIP1b→RA, excluding one genetic instrument nominally associated with RA ($P<0.05$); (c) RA→MIP1b; (d) RA→MIP1b, excluding two genetic instruments nominally associated with MIP1b ($P<0.05$).

eFigure 7. Genetic association of Alzheimer's disease (AD) and Crohn's disease (Crohn) within the genomic region +/- 50kb from the MIG coding gene. Dashed line indicates P-value of 10^{-4}.

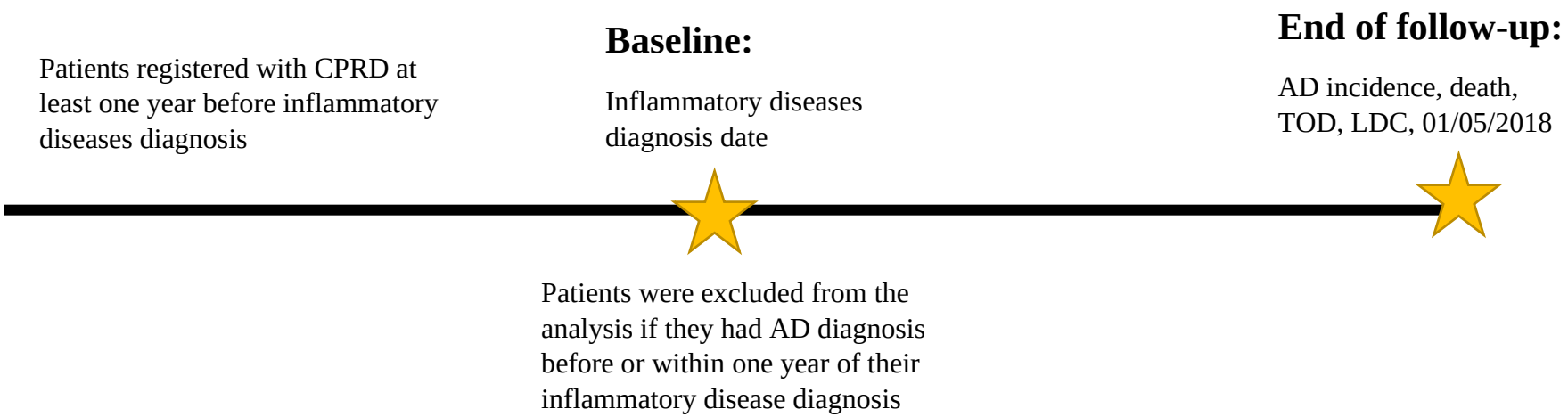
eFigure 8. Genetic associations of Alzheimer's disease (AD) with MIG and of Crohn's disease (Crohn) with MIG, within the genomic region +/- 50kb from the MIG coding gene. Dashed line indicates P-value of 10^{-4}.

eFigure 9. Genetic associations of Alzheimer's disease (AD) with IP10 and of Crohn's disease (Crohn) with IP10, within the genomic region +/- 50kb from the IP10 coding gene. Dashed line indicates P-value of 10^{-4}.

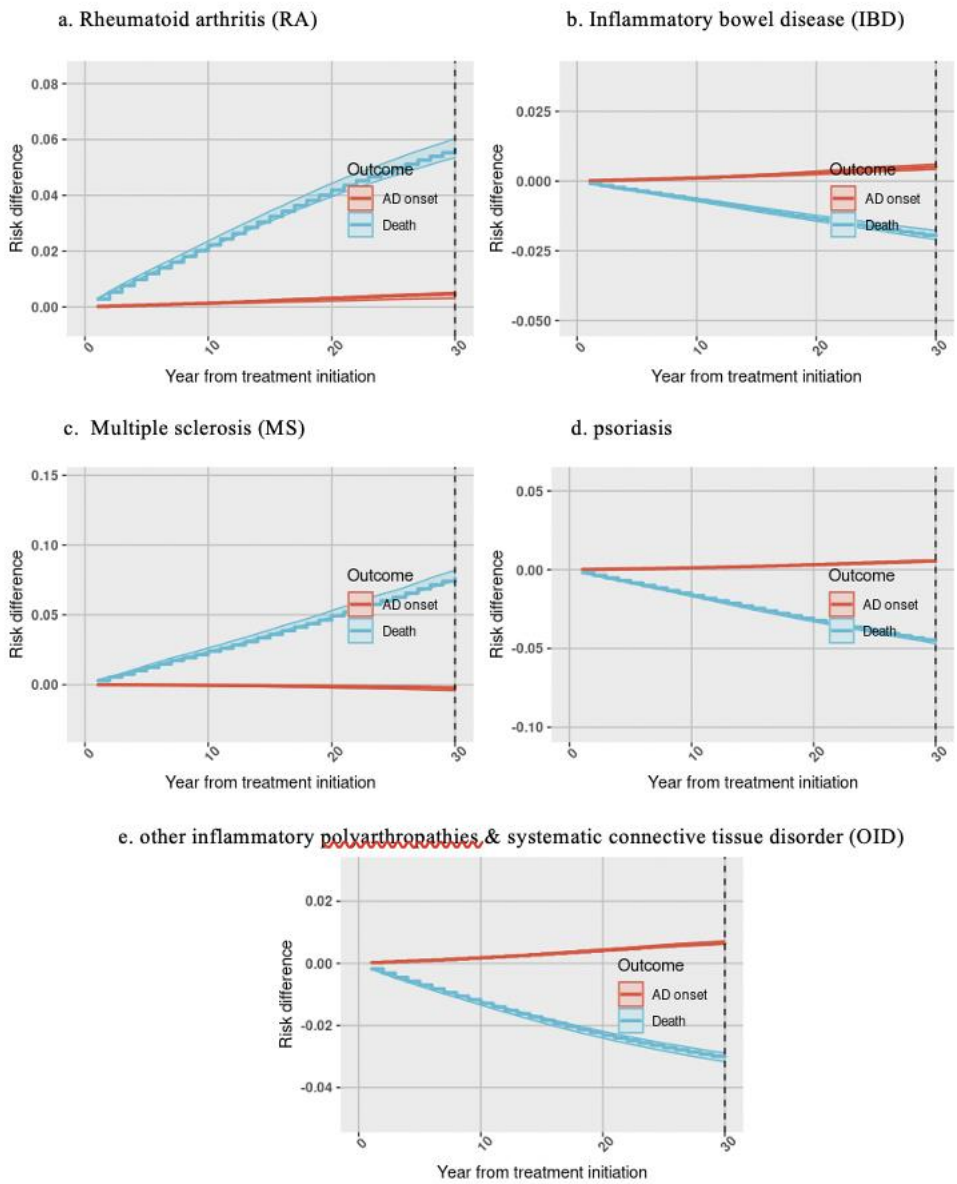
eFigure 2. Alzheimer's disease (AD) cases diagnosis algorithm



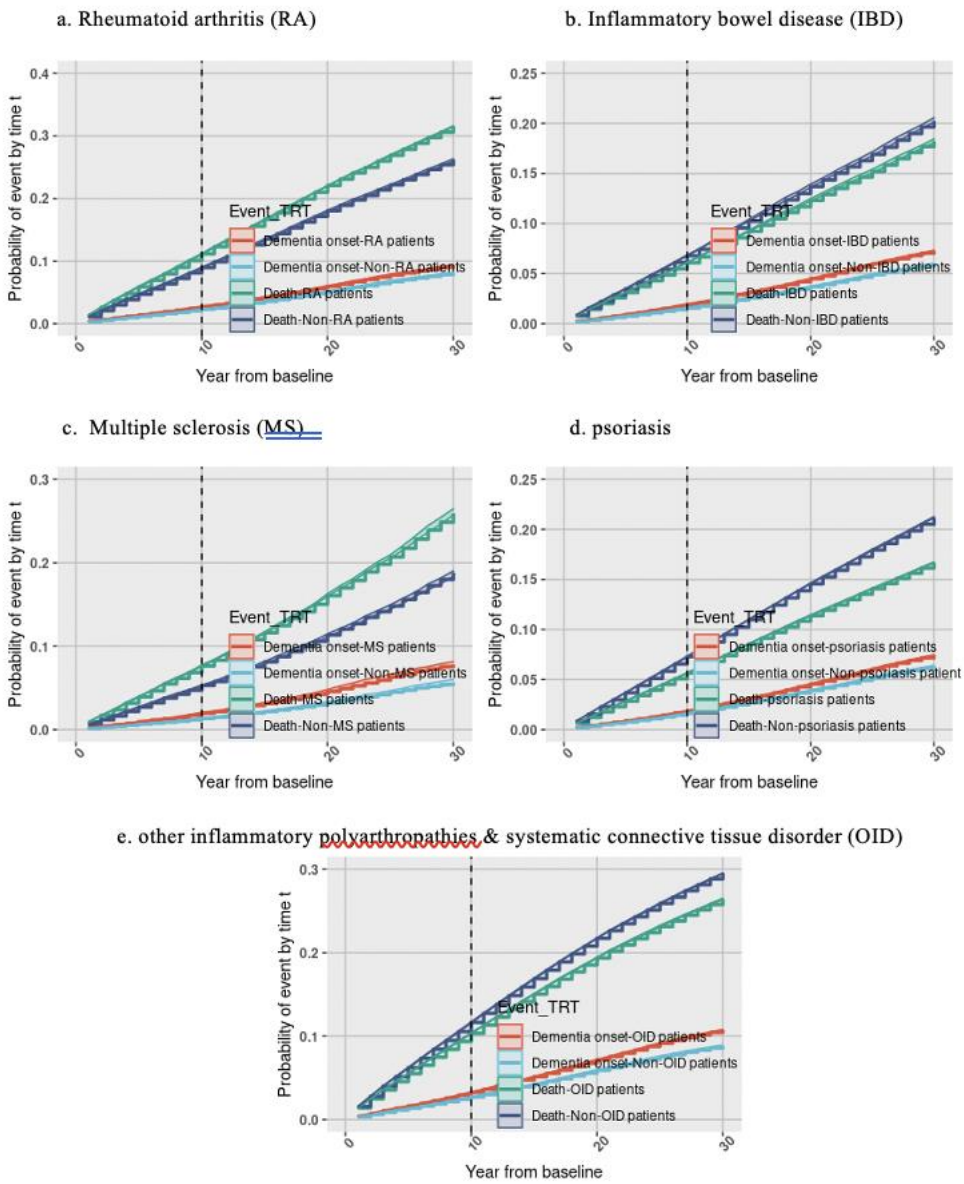
eFigure 2. Study design for CPRD analysis



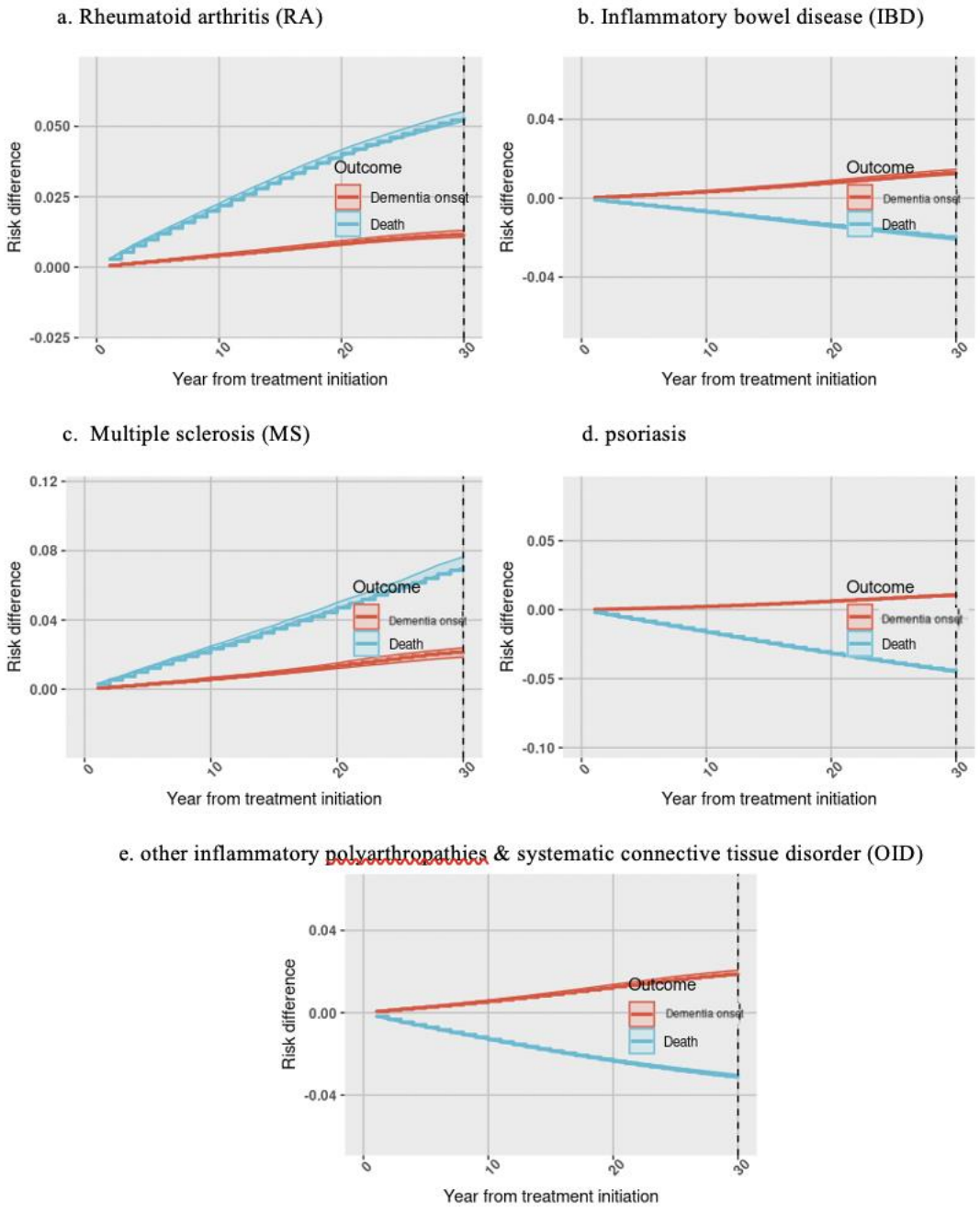
eFigure 3. Risk differences curves for AD onset and death, estimated from the cause-specific hazards using Clinical Practice Research Datalink.



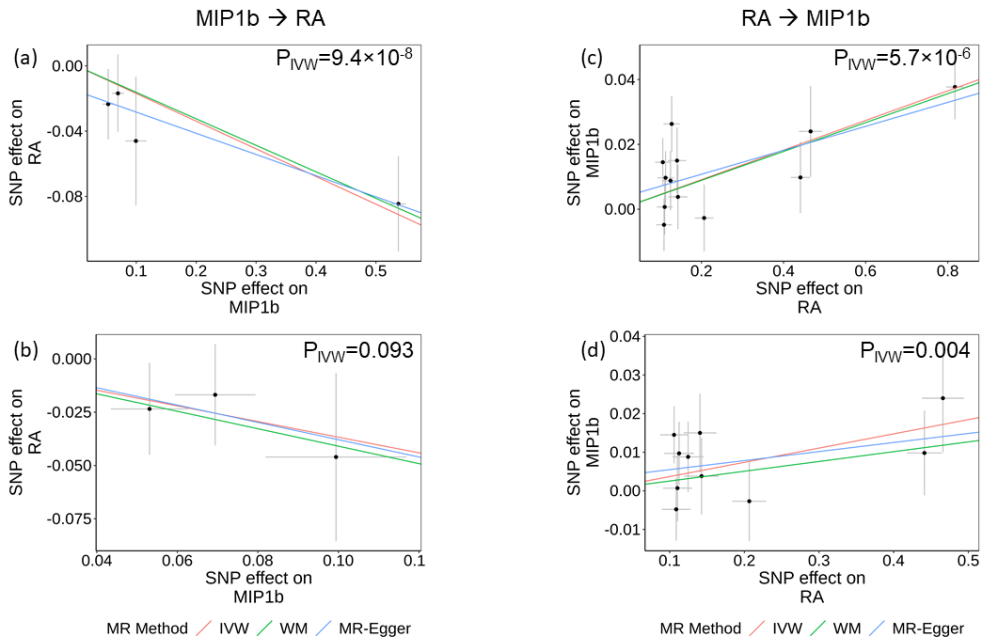
eFigure 4. Cumulative incidence function (CIF) curves for dementia onset and for death using Clinical Practice Research Datalink.



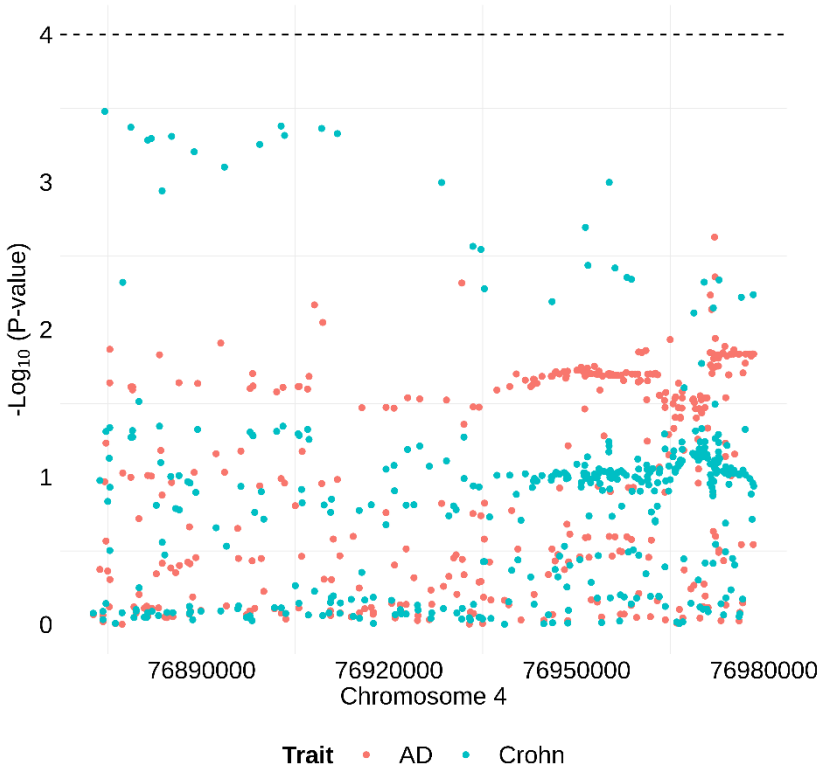
eFigure 5. Risk differences curves for dementia onset and death, estimated from the cause-specific hazards using Clinical Practice Research Datalink.



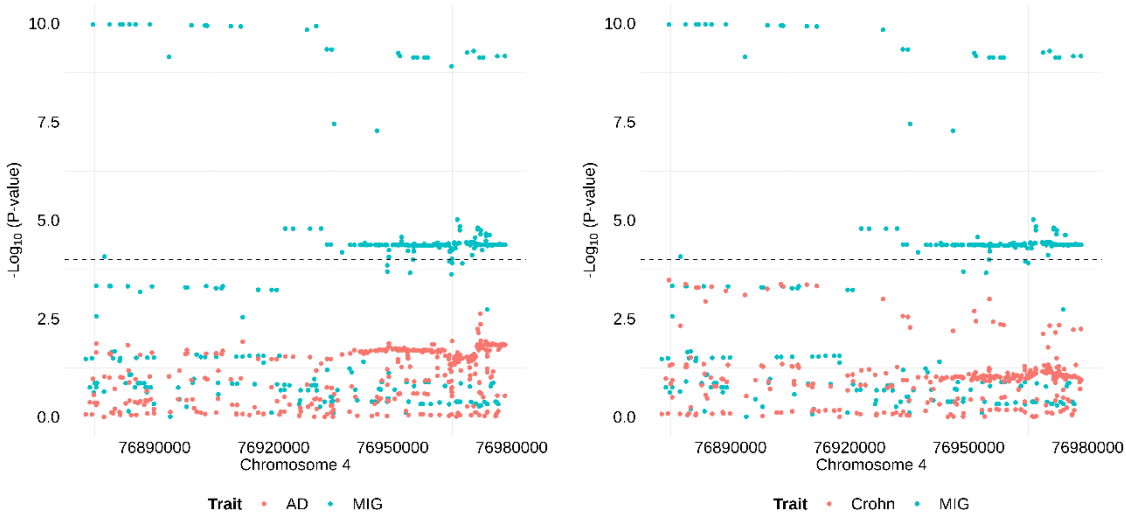
eFigure 6. Associations between circulating MIP1b level and rheumatoid arthritis (RA) using two-sample Mendelian randomisation (MR). (a) MIP1b→RA; (b) MIP1b→RA, excluding one genetic instrument nominally associated with RA (P<0.05); (c) RA→MIP1b; (d) RA→MIP1b, excluding two genetic instruments nominally associated with MIP1b (P<0.05).



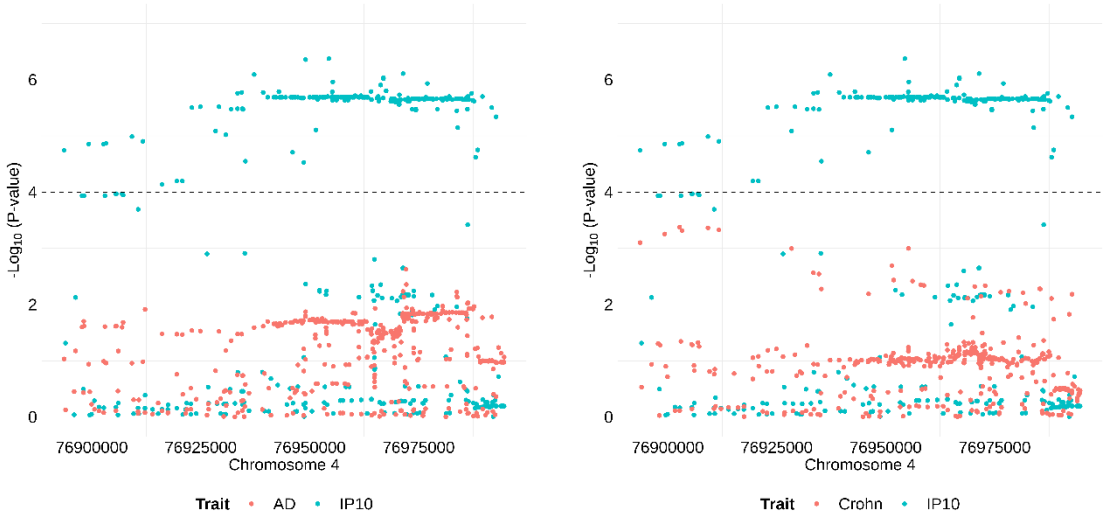
eFigure 7. Genetic association of Alzheimer's disease (AD) and Crohn's disease (Crohn) within the genomic region +/- 50kb from the MIG coding gene. Dashed line indicates P-value of 10^{-4}.



eFigure 8. Genetic associations of Alzheimer's disease (AD) with MIG and of Crohn's disease (Crohn) with MIG, within the genomic region +/- 50kb from the MIG coding gene. Dashed line indicates P-value of 10^{-4}.



eFigure 9. Genetic associations of Alzheimer's disease (AD) with IP10 and of Crohn's disease (Crohn) with IP10, within the genomic region +/- 50kb from the IP10 coding gene. Dashed line indicates P-value of 10^{-4}.



eMethods

Clinical Practice Research Datalink (CPRD) observational analysis

Exposure assessment

We analysed six inflammatory diseases: 1) rheumatoid arthritis (RA), 2) inflammatory bowel disease (IBD), 3) multiple sclerosis (MS), 4) psoriasis, and 5) other inflammatory polyarthropathies & systematic connective tissue disorders (OID). The inflammatory diseases definitions and diagnosis code lists were based on CALIBER which is a research platform sharing 'research ready' variables extracted from linked electronic health records (EHR) including the Clinical Practice Research Datalink (CPRD), Hospital Episode Statistics (HES) and national registry social deprivation and mortality data from the Office for National Statistics (ONS)[1, 2]. The validity of the diagnosis code lists was determined by comparing with other published studies accompanying with a full set of clinical diagnosis codes [3-6]. See Supplementary Tables S1 to S5 for the full code lists.

Outcome assessment

We developed an algorithm integrating diagnoses from both primary care (at least one diagnosis code for AD in CPRD) and secondary care (at least one diagnosis code for AD in HES) and treatment of dementia (at least one drug prescription code for dementia in CPRD) to improve the AD case ascertainment and decrease the possibility of a false-positive diagnosis of AD cases. We identified a total number of 134,952 AD cases with at least of one 1) AD diagnosis record in CPRD, 2) AD diagnosis record in HES or ONS, or 3) drug prescription record for dementia in CPRD in our CPRD cohort. Among them, 22,872 CPRD patients had both the diagnosis code for AD in CPRD and HES/ONS and one dementia drug prescription.

Statistical analyses

Estimation of effect of the presence of inflammatory diseases on all-cause mortality and competing death

For a single time-to-death outcome, we estimated the Cox proportional hazards model and the nonparametric Kaplan-Meier survival curves for both inverse-probability-of-treatment-weighted arms. The latter model allows to estimate robustly the time-varying causal survival curves, while the former model provides a one-number summary of the treatment effect through a fixed hazard ratio. Technically, to estimate the effect of the presence of

inflammatory diseases on all-cause mortality, we considered the same causal framework as for the competing risks but only for a single outcome, time-to-death. This means that both our analyses rely on the same causal assumptions and the assumption of independent censoring. Practically, for both cases we used our R package, *causalCmprsk*, to estimate the causal survival curves.

Mendelian randomization

Genetic associations of circulating inflammatory biomarkers

We selected circulating inflammatory biomarkers of which meta-analyses of genome-wide association study (GWAS) were available based on three Finnish cohorts, namely Northern Finland Birth Cohort 1966 (NFBC1966), the Cardiovascular Risk in Young Finns (YFS), and FINRISK.[7] To increase the power of our analysis, we further incorporated the Finnish meta-analysis with summary statistics from GWAS on proteins in the INTERVAL study [8] and SCALLOP Consortium [9]. For each biomarker, we regressed the genetic association for suggestive and independent single-nucleotide polymorphisms (SNPs with $P<1\times10^{-5}$ and $R^2<0.1$) in Finnish GWAS with those from the INTERVAL GWAS and SCALLOP GWAS. When a correlation ($P\leq0.05$) was observed across the studies, we meta-analysed the GWAS. In such instances, we calibrated the SNP-biomarker estimates of the INTERVAL (3 biomarkers) and SCALLOP (8 biomarkers) consortium using the slope of the regression line to match with the SNP-biomarkers estimates from the Finnish GWAS. For the remaining biomarkers, we used summary statistics from the GWAS with the largest sample size (Finnish GWAS (26 biomarkers), the INTERVAL study (2 biomarkers), and SCALLOP consortium (7 biomarkers). **eTable 7 in the Supplement** shows studies that were eventually used for each biomarker.

Genetic associations of inflammatory diseases

We obtained genetic associations for eight inflammatory diseases from the largest GWAS, including psoriasis, rheumatoid arthritis (RA), multiple sclerosis (MS), inflammatory bowel disease (IBD), and the two subtypes of IBD (Crohn's disease and ulcerative colitis). Case ascertainment of inflammatory diseases was based on clinical diagnosis recorded in the hospital or self-reported records. Sample size of these GWAS ranged from 33,394 (10,588 cases and 22,806 controls) for psoriasis to 47,580 (13,838 cases and 33,742 controls) for RA. (**eTable 8 in the Supplement**).

Genetic associations of Alzheimer's disease and dementia

Genetic associations of risk of late-onset AD were obtained from the GWAS meta-analysis by the International Genomics of Alzheimer's Project (IGAP).[10] The discovery stage of the IGAP GWAS meta-analysis included 21,982 cases (clinically diagnosed or autopsy) and 41,944 cognitively normal controls of European ancestry.[10] The IGAP GWAS released in 2019 is the largest GWAS on late-onset AD based on clinical or post-mortem diagnosis.

As a sensitivity analysis, we also used genetic associations of dementia obtained from a meta-analysis of the IGAP, GR@ACE, and UK Biobank (N=409,435 individuals of European ancestry).[11] IGAP and GR@ACE included a total of 36,675 AD cases and 58,482 cognitively normal control. Case ascertainment was based on clinical diagnosis or autopsy. Self-reported parental history of dementia was used as a proxy phenotype in the UK Biobank GWAS. GWAS on maternal dementia (27,696 cases and 260,980 controls) and paternal dementia (14,338 cases and 245,941 controls) were performed separately and then meta-analysed.

eReferences

1. Kuan, V., et al., *A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service.* The Lancet Digital Health, 2019. **1**(2): p. e63-e77.
2. Denaxas, S., et al., *UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER.* Journal of the American Medical Informatics Association, 2019. **26**(12): p. 1545-1559.
3. Pujades-Rodriguez, M., et al., *Rheumatoid arthritis and incidence of twelve initial presentations of cardiovascular disease: a population record-linkage cohort study in England.* PloS one, 2016. **11**(3): p. e0151245.
4. Parisi, R., et al., *Psoriasis and the risk of major cardiovascular events: cohort study using the clinical practice research datalink.* Journal of Investigative Dermatology, 2015. **135**(9): p. 2189-2197.
5. Muller, S., et al., *An algorithm to identify rheumatoid arthritis in primary care: a Clinical Practice Research Datalink study.* BMJ open, 2015. **5**(12): p. e009309.
6. Kontopantelis, E., et al., *Glucose, blood pressure and cholesterol levels and their relationships to clinical outcomes in type 2 diabetes: a retrospective cohort study.* Diabetologia, 2015. **58**(3): p. 505-518.
7. Karhunen, V., et al., *Genetic study of circulating cytokines offers insight into the determinants, cascades and effects of systemic inflammation.* medRxiv, 2020: p. 2020.10.26.20219477.
8. Sun, B.B., et al., *Genomic atlas of the human plasma proteome.* Nature, 2018. **558**(7708): p. 73-79.
9. Folkersen, L., et al., *Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals.* Nat Metab, 2020. **2**(10): p. 1135-1148.
10. Kunkle, B.W., et al., *Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing.* Nat Genet, 2019. **51**(3): p. 414-430.
11. de Rojas, I., et al., *Common variants in Alzheimer’s disease and risk stratification by polygenic risk scores.* Nature Communications, 2021. **12**(1): p. 3417.