Causal Role of IL-33/ST2 in Dementia: A Mendelian Randomisation Study

2/05/2024

TRANSLATIONAL DATA SCIENCE

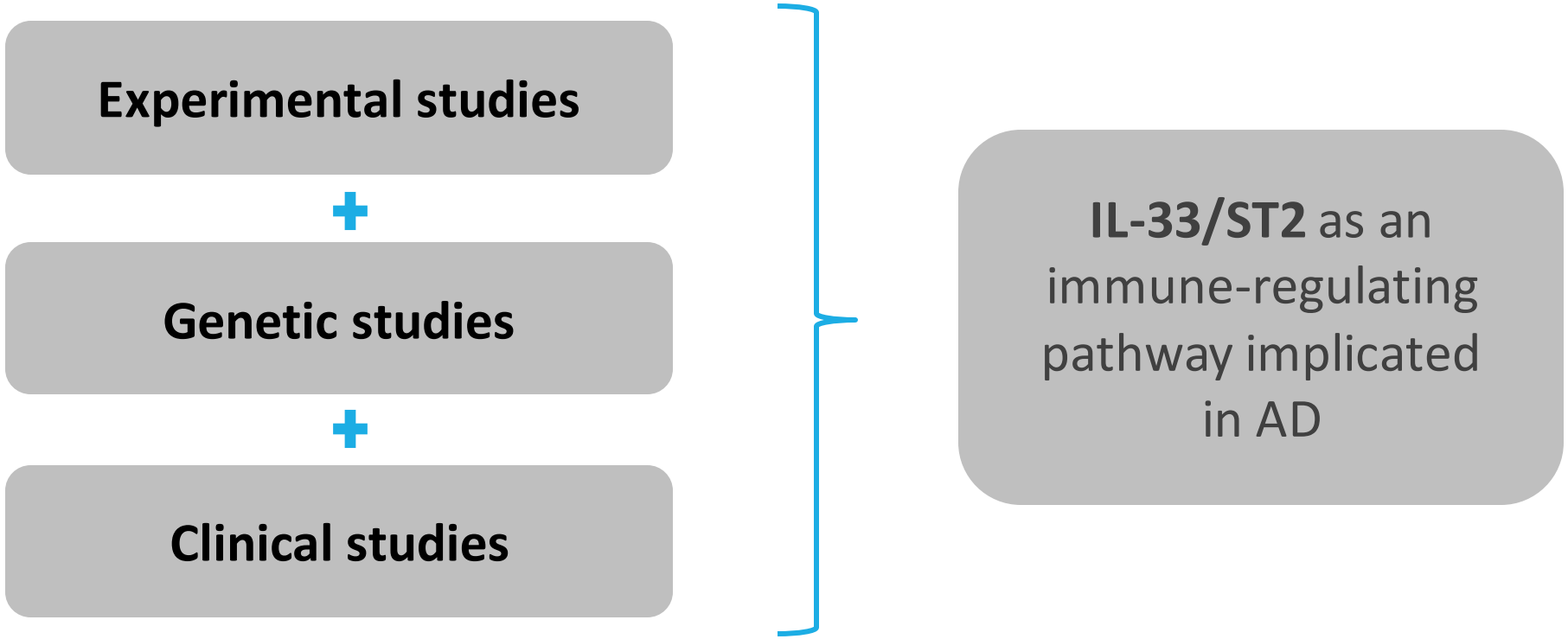
GROUP 1

Background

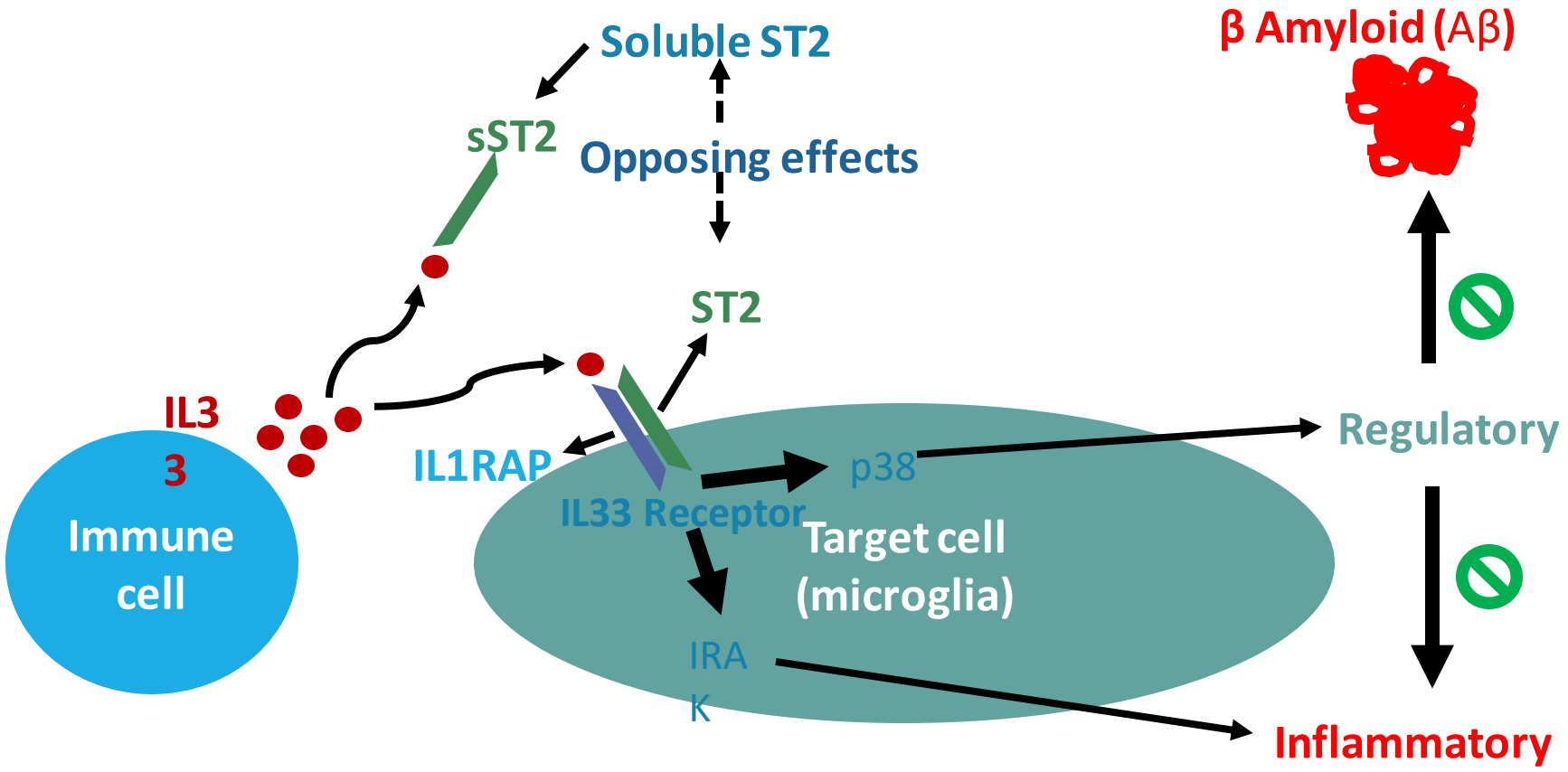
Dementia

- A degenerative and progressive brain pathologies characterized by cognitive and memory deterioration
- 2 major types:
  - Alzheimer's disease (AD)
  - Vascular Dementia (VD)
- 47.5 million worldwide with dementia – 850,000 in the UK
- Highly **polygenic**
- Core mechanistic pathways not fully understood but **immune system and inflammatory processes important roles in AD pathology**

Identification of IL-33/ST2 as risk factors for Alzheimer's disease



Overview of suggested IL-33/ST2 axis in AD



Research Question

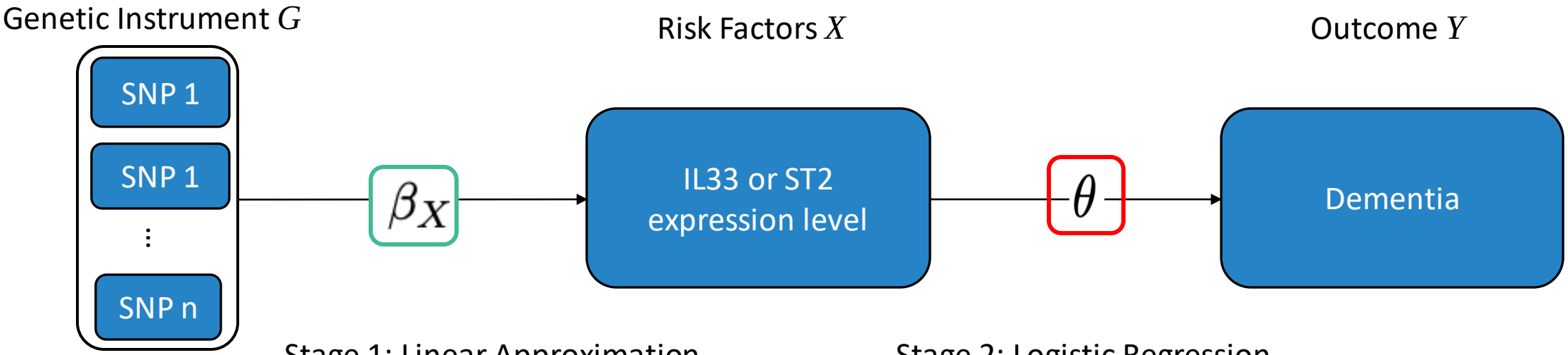
Is the **association** between IL-33/ST2 pathway and dementia **causal**?

Aim to answer this question using **Mendelian Randomization** methods

Methods Overview

Individual-level MR

Two-Stage Least Square Method using PRS



Stage 1: Linear Approximation

$$X = f(G) = \beta_X G + \epsilon_X$$

$$PRS \propto \hat{X} = \beta_X G$$

Polygenic Risk Score as a proxy

Stage 2: Logistic Regression

$$Y = h(\hat{X})$$

$$= logit(p_i)$$

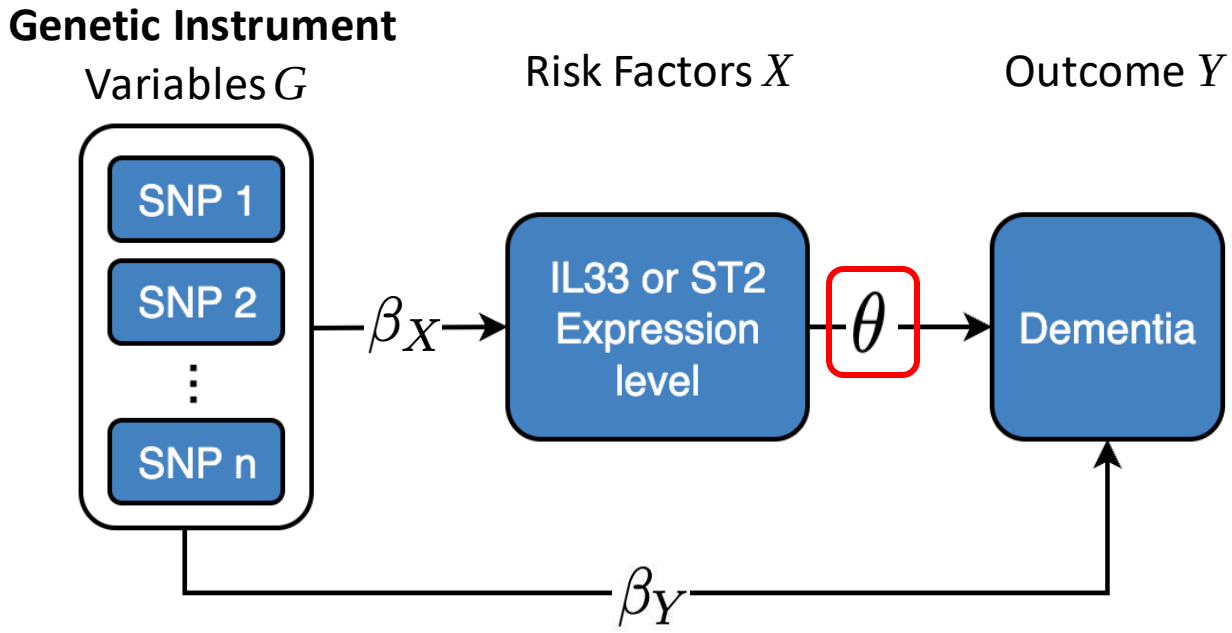
$$= \hat{\theta}_{2SLS} \cdot PRS + \epsilon_Y$$

Estimated causal effect

- β_X is obtained from GWAS of IL-33 or ST2
- G and Y are obtained from UK Biobank

Summary-level MR

Two-Sample



- β_X is obtained from GWAS of IL-33 or ST2
- β_Y is obtained from GWAS of Alzheimer's Disease

- MR-Inverse variance weighting (IVW)

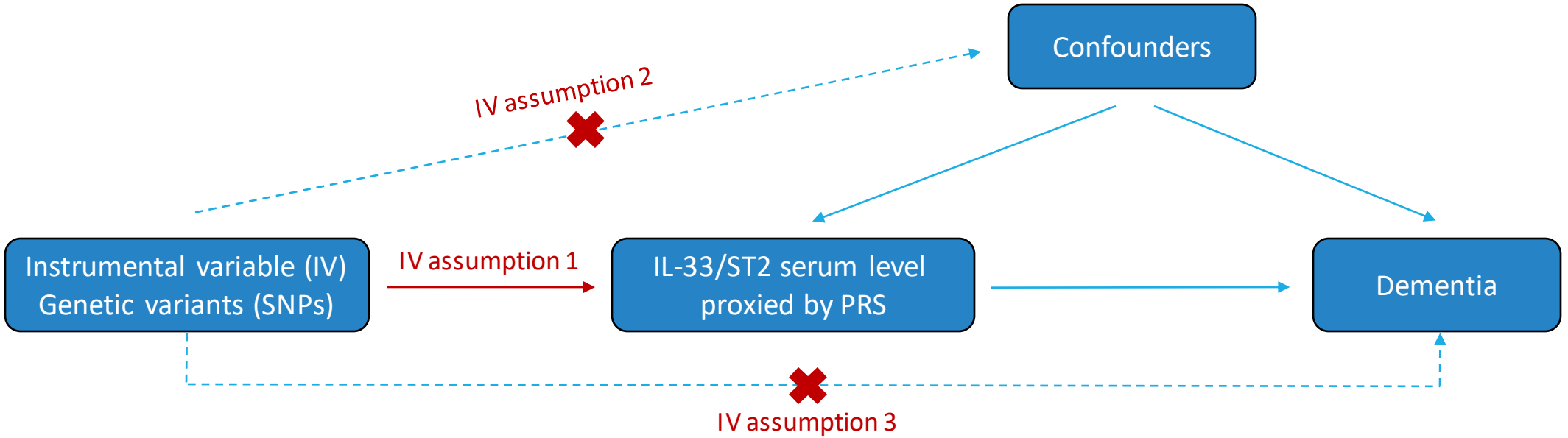
$$\beta_Y = \hat{\theta}_{IVW} \cdot \beta_X + \epsilon$$

- MR-Egger

$$\beta_Y = \theta_0 + \hat{\theta}_{MR\text{-}Egger} \cdot \beta_X + \epsilon$$

- MR-Simple Median
- MR-Weighted Median

MR Assumptions



1. **Relevance**: Examine Exposure ~ PRS regression, F-statistics > 10
2. **Exchangeability**: Hard to verify; violated by population stratification
3. **Exclusion restriction**: Hard to verify; violated by pleiotropy, linkage disequilibrium

Model Overview

Step	Analysis performed	Formula
1. Observational study	Logistic regression	Dementia ~ Risk factors
2. PRS strength – relevance assumption	Linear regression	Risk factors ~ PRS
3. Positive control	Logistic regression	Asthma ~ PRS
4. Negative control	Logistic regression	HotDrink ~ PRS
5. Individual-level MR	Logistic regression	Dementia ~ PRS
	Logistic regression (conditioned)	Dementia ~ PRS ST2 + PRS IL33
	Logistic regression stratified by gender	Dementia ~ PRS
6. Summary-level MR	IVW, MR-Egger, Simple median, Weighted median	

- Risk factors: IL-33/ST2 serum protein measurements

Analysis plan

Selection of genetic IV for IL-33 and ST2

Protein	Population and size	Variants selection		After clumping R = 0.01	Selected SNPs
		Threshold	*# SNPs*	*# SNPs*	
IL - 33	11,793 European *(Sweden, Denmark, U.K., Germany, Estonia, Croatia)*	p-value < 10^{-6}	10	**2**	“rs8042883”, “rs10415966”
ST2	21,758 European	p-value < 5×10^{-8}	390	**2**	"rs1468789”, "rs13029918”
		p-value < 10^{-6}	1208	**26**	"rs115540952”, "rs11465602”, etc.

Definition of data extraction

- **Exclusion/inclusion criteria**:
  - Excluded self-reported outcomes
  - Excluded non white ethnicity (White = White, British, Irish, Any other white background)
- **Outcome**:
  - All cases included (prevalent and incident)
  - Three types of dementia (Alzheimer's Disease, Vascular Dementia, Other Dementia)
- **Genotype data:**
- Under 2 variants: only individuals with both SNPs missing were excluded
  - Under 26 variants: only individuals with all 26 SNPs missing were excluded
- **Protein data**:
  - KNN to impute the 12% missing for ST2
  - Quantile regression to imputed the 88% missing for IL-33
- **Positive and negative controls**:
  - Asthma (+) and hot drink (-): binary variable

Table 1.

Variable	All types			Alzheimer's Disease		
	Case N = 7140 (1.56%)[1]	Control N = 451726 (98.4%)[1]	p-value[2]	Case N = 3027 (0.66%)[1]	Control N = 455233 (99.3%)[1]	p-value[2]
Current Age (Years)	80.31 (4.85)	72.58 (8.02)	<0.001	80.72 (4.31)	72.64 (8.03)	<0.001
Gender			<0.001			0.005
Female	3,396 (47.6%)	245,697 (54.4%)		1,567 (51.8%)	247,279 (54.3%)	
Male	3,744 (52.4%)	206,029 (45.6%)		1,460 (48.2%)	207,954 (45.7%)	

Variable	Vascular Dementia			Other Dementia		
	Case N = 1589 (0.35%)[1]	Control N = 455911 (99.7%)[1]	p-value[2]	Case N = 5081 (1.11%)[1]	Control N = 453674 (98.9%)[1]	p-value[2]
Current Age (Years)	81.02 (4.16)	72.65 (8.03)	<0.001	80.24 (4.97)	72.62 (8.02)	<0.001
Gender			<0.001			<0.001
Female	656 (41.3%)	247,728 (54.3%)		2,367 (46.6%)	246,680 (54.4%)	
Male	933 (58.7%)	208,183 (45.7%)		2,714 (53.4%)	206,994 (45.6%)	

PRS Construction

$$PRS = \frac{\sum_{i=1}^{N} \beta_i G_i}{N}$$

$G_i \in \{0, 1, 2\}$ is the number of effect allele for the i-th SNP

$\beta_i \in \mathbb{R}$ is the effect size of the i-th SNP

N is the number of non-missing SNP

Results

1. Observational Study

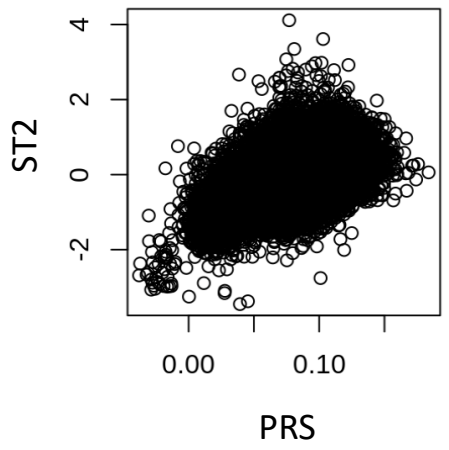
Logistic regression: Outcome ~ Risk factor

outcome	# cases	Risk factor: ST2		Risk factor: IL33	
		OR (95% CI)	P-value	OR (95% CI)	P-value
All types of dementia	7140	1.563 (1.410, 1.732)	1.64E-17	0.983 (0.928, 1.041)	0.571
Alzheimer's Disease	3027	1.507 (1.298, 1.748)	6.53E-08	0.996 (0.916, 1.082)	0.931
Vascular Dementia	1589	1.738 (1.383, 2.179)	1.92E-06	0.970 (0.853, 1.103)	0.643
Other Dementia	5081	1.500 (1.330, 1.692)	3.85E-11	1.004 (0.939, 1.074)	0.898

2. PRS Strength – addressing MR assumption I (relevance)

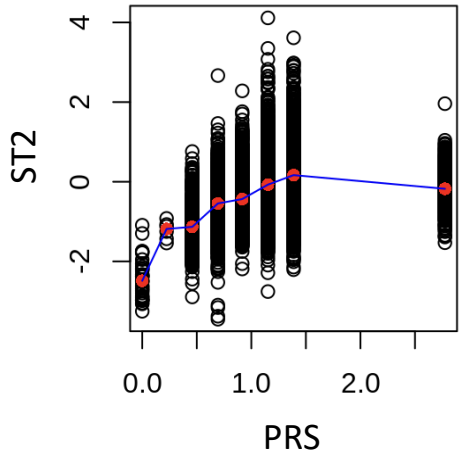
Linear regression: IL-33/ST2 ~ PRS

- **ST2** – 26IV



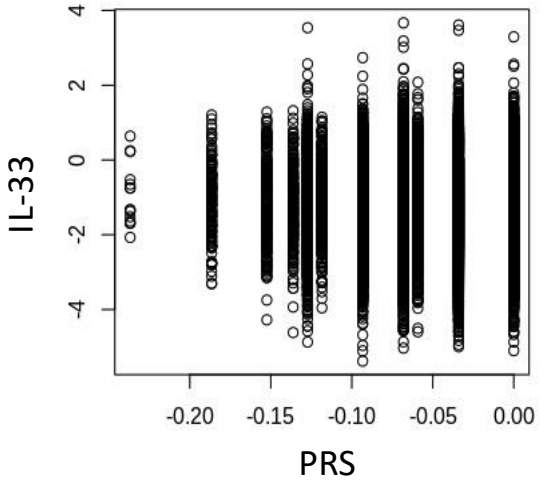
F-statistic: 16970
R^2: 0.2582
P-value: <2.2e-16

- **ST2** – 2IV



F statistic : 8636
R^2: 0.164
P-value: <2e-16

- **IL-33** – 2IV



F-statistic: 0.3735
R^2: -1.286e-05
P-value: 0.541

3. Positive control: asthma

Logistic regression: Asthma ~ PRS

PRS	OR (95% CI)	P-value
ST2 - 26IV	0.181 (0.127, 0.259)	**<2e-16**
ST2 - 2IV	0.840 (0.811, 0.871)	**<2e-16**
IL-33 - 2IV	0.944 (0.752, 1.186)	**0.622**

cases: 65046

4. Negative control: PRS against a dummy risk factor, hot drink temperature in UKB

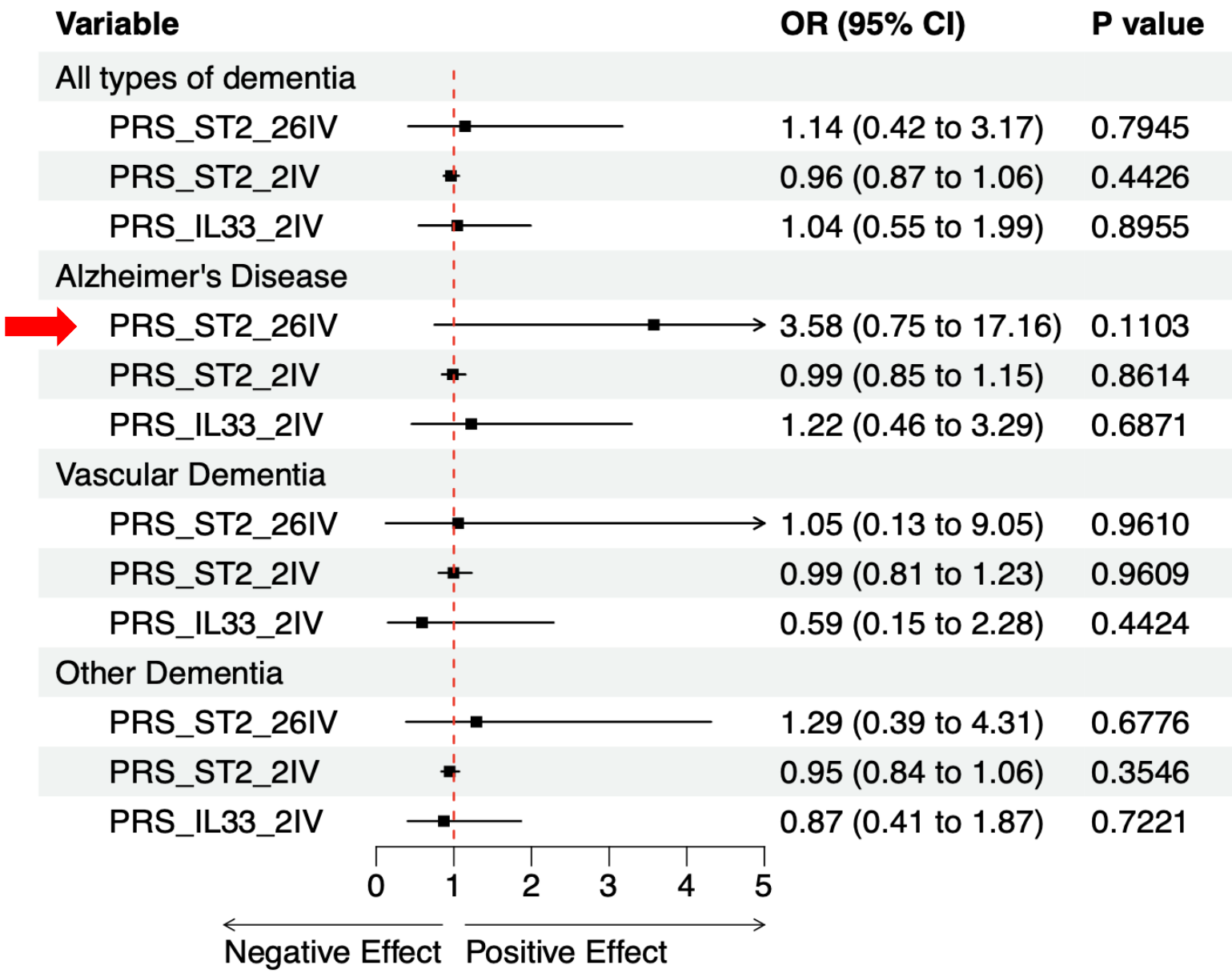
Logistic regression: HotDrink ~ PRS

PRS	OR (95% CI)	P-value
ST2 - 26IV	0.460 (0.132, 1.615)	**0.225**
ST2 - 2IV	0.959 (0.848, 1.084)	**0.508**
IL33 - 2IV	0.574 (0.261, 1.267)	**0.168**

cases: 4637

5. Individual-level MR

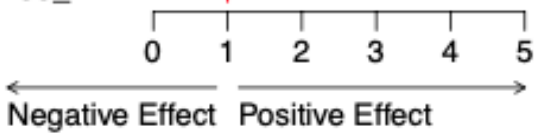
Logistic regression: Dementia ~ PRS



5. Individual-level MR – conditioned

Logistic regression:
Model 1 = Dementia ~ PRS IL33 + PRS ST2 26 IV
Model 2 = Dementia ~ PRS IL33 + PRS ST2 2 IV

Model	Variable	OR (95% CI)	P value
All types of dementia			
Model 1	PRS_ST2_26IV	1.14 (0.42 to 3.17)	0.7944
	PRS_IL33_2IV	1.04 (0.55 to 1.99)	0.8953
Model 2	PRS_ST2_2IV	0.96 (0.87 to 1.06)	0.4427
	PRS_IL33_2IV	1.04 (0.55 to 1.99)	0.8963
Alzheimer's Disease			
Model 1	PRS_ST2_26IV	3.58 (0.75 to 17.16)	0.1102
	PRS_IL33_2IV	1.23 (0.46 to 3.29)	0.6861
Model 2	PRS_ST2_2IV	0.99 (0.85 to 1.15)	0.8618
	PRS_IL33_2IV	1.22 (0.46 to 3.29)	0.6873
Vascular Dementia			
Model 1	PRS_ST2_26IV	1.05 (0.13 to 9.04)	0.9614
	PRS_IL33_2IV	0.59 (0.15 to 2.28)	0.4425
Model 2	PRS_ST2_2IV	0.99 (0.81 to 1.23)	0.9602
	PRS_IL33_2IV	0.59 (0.15 to 2.28)	0.4424
Other Dementia			
Model 1	PRS_ST2_26IV	1.29 (0.39 to 4.31)	0.6778
	PRS_IL33_2IV	0.87 (0.41 to 1.87)	0.7224
Model 2	PRS_ST2_2IV	0.95 (0.84 to 1.06)	0.3544
	PRS_IL33_2IV	0.87 (0.41 to 1.86)	0.7213

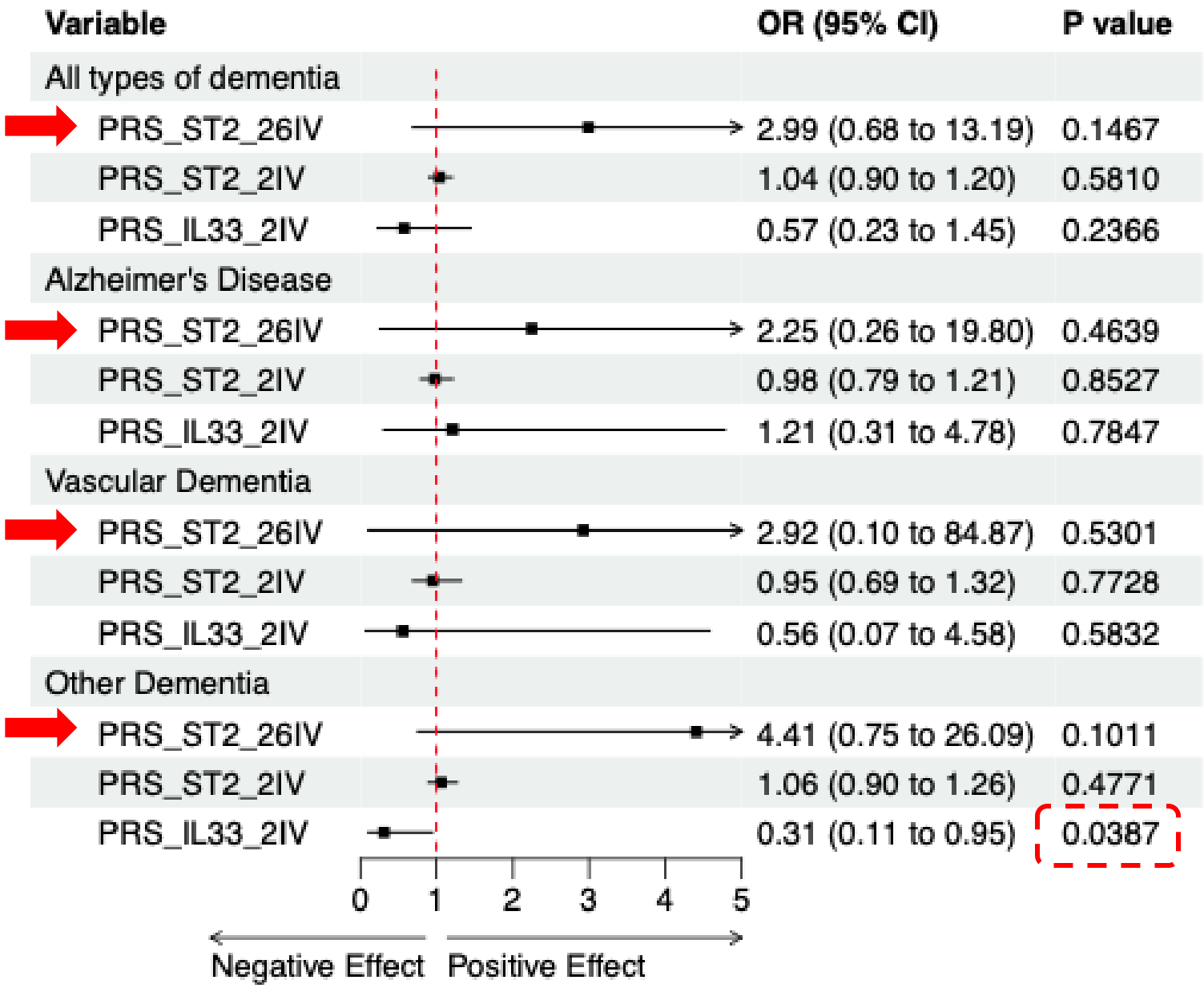


5. Individual-level MR – stratified by gender

Female Stratum

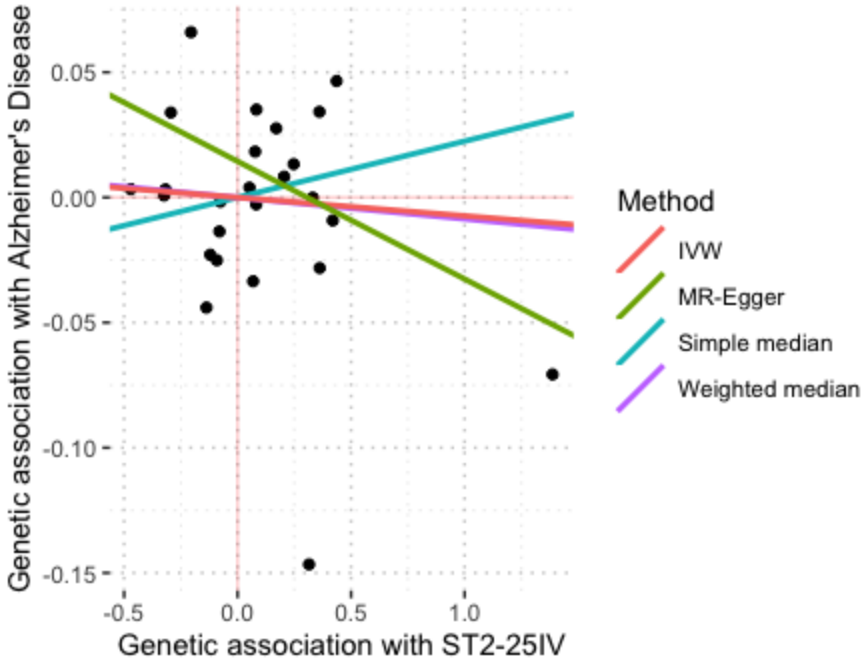
No significance in male stratum

Logistic regression: Dementia ~ PRS



6. Summary-level MR (sensitivity analysis)

Instrument variables	Instrument Strength (F)	Method	P-value
IL33 – 2IV	25.15	IVW	0.709
ST2 – 2IV	562.33	IVW	0.281
ST2 – 25IV	1504.52	IVW	0.742
		MR-Egger	0.106
		Simple Median	0.599
		Weighted Median	0.764



- Summary-level MR uses one less IV compared to individual-level MR – missing from the AD GWAS
- AD GWAS with total 21,982 cases obtained from Kunkle et al. (2019)

Discussion

Conclusions

Observational study

1. Observational study: ST2 is significant with all types of dementia, IL-33 non-significant

Mendelian randomisation

1. PRS associates with ST2 serum levels (strong instrument), but not with IL-33
2. Method validated by positive and negative control
3. Individual level MR: ST2 and IL-33 not significantly associated with dementia
4. Summary level MR: replicates previous results (non-significance)

Limitations

1. **Previous MR study** established causal link between ST2 and AD among female APOE4+ population – not attempted given the project scope
2. **2 sample MR:** source and target population were the same (limited data available)
3. **Number of cases** may have led to **power issues** → summary level MR
4. **IV selection:** relaxed p-value threshold for IL-33 from $p=5 \times 10^{-8}$ to $p=10^{-6}$
5. **Difficult to address all MR assumptions:** addressed relevance, and partially independence (focusing on Caucasian), and exclusion/pleiotropy (biology, larger GWAS/MR Egger, etc.)
6. **Protein data:** IL-33 measured from blood serum, 85% not detectable – tissue-specificity may be more relevant
7. **No multiple testing correction**

Future perspectives

1. **Stratification by APOE-ε4 status**
2. **Non-linear MR** to understand non-linear relationships between IL-33/ST2 & AD
3. **Tissue-specific protein data** may allow for more accurate IL-33 measurements and reduce the proportion of missing values
4. **Population-specific studies** by measuring ethnic-specific effects to understand the global applicability of the findings
5. **UKB value** will increase as the cohort continues to age, offering more insights into the role of the IL-33/ST2 axis on dementia
6. **Exploration of additional biological pathways** and dementia using MR analysis

References

- Kunkle et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nature Genetics*. 2021 Mar; 51(3): 414-430. doi: 10.1038/s41588-019-0358-2.
- Jiang et al. An IL1RL1 genetic variant lowers soluble ST2 levels and the risk effects of APOE-ε4 in female patients with Alzheimer's disease. *Nature Aging*. 2022 Jul; 2: 616-624. doi: 10.1038/s43587-022-00241-9.
- Folkersen et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nature Metabolism*. 2020 Oct; 2(10): 1135-1148. doi: 10.1038/s42255-020-00287-2.
- Zhao et al. Genetics of circulating inflammatory proteins identifies drivers of immune-mediated disease risk and therapeutic targets. *Nature Immunology.* 2023; 24(9): 1540-1551. doi:10.1038/s41590-023-01588-w.
- World Health Organisation. Dementia. https://www.who.int/news-room/ fact-sheets/detail/dementia. 2023 Mar. [Accessed: 13 March 2024].

Appendix

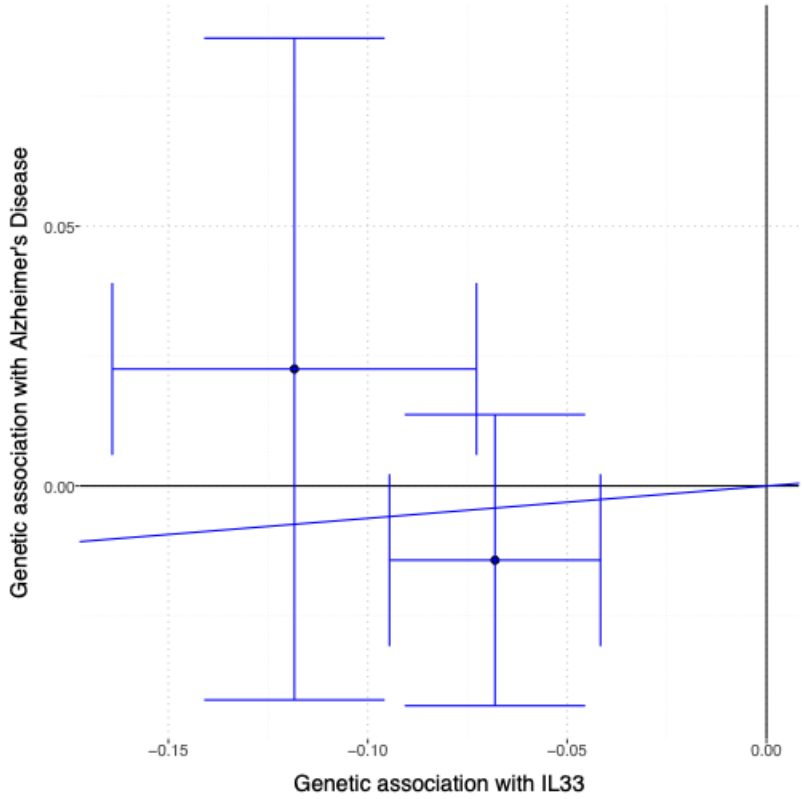
Summary-level MR – IL33 (2 IVs)

Inverse-variance weighted method
(variants uncorrelated, fixed-effect model)
Number of Variants : 2

Method Estimate Std Error 95% CI p-value
IVW 0.062 0.167 -0.265, 0.389 0.709

Residual standard error = 1.157
Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.

Heterogeneity test statistic (Cochran's Q) = 1.3397 on 1 degrees of freedom, (p-value = 0.2471)
I^2 = 25.4%.
F statistic = 25.6.



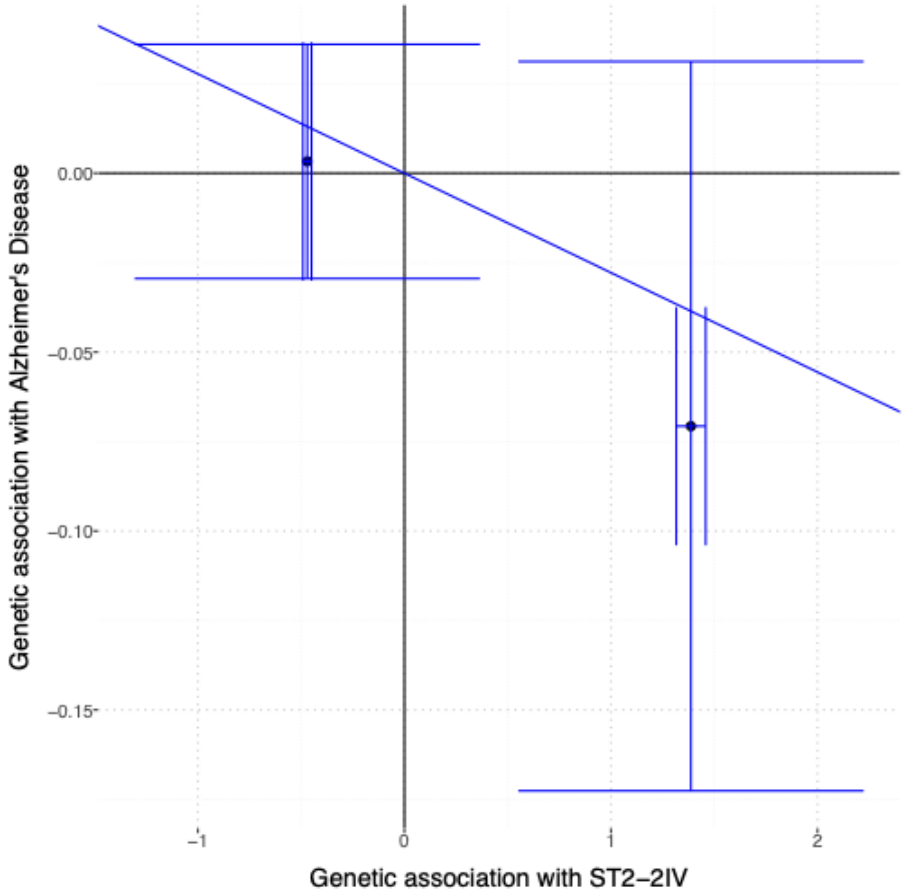
Summary-level MR – ST2 (2 IVs)

Inverse-variance weighted method
(variants uncorrelated, fixed-effect model)
Number of Variants : 2

Method Estimate Std Error 95% CI p-value
IVW -0.028 0.026 -0.078, 0.023 0.281

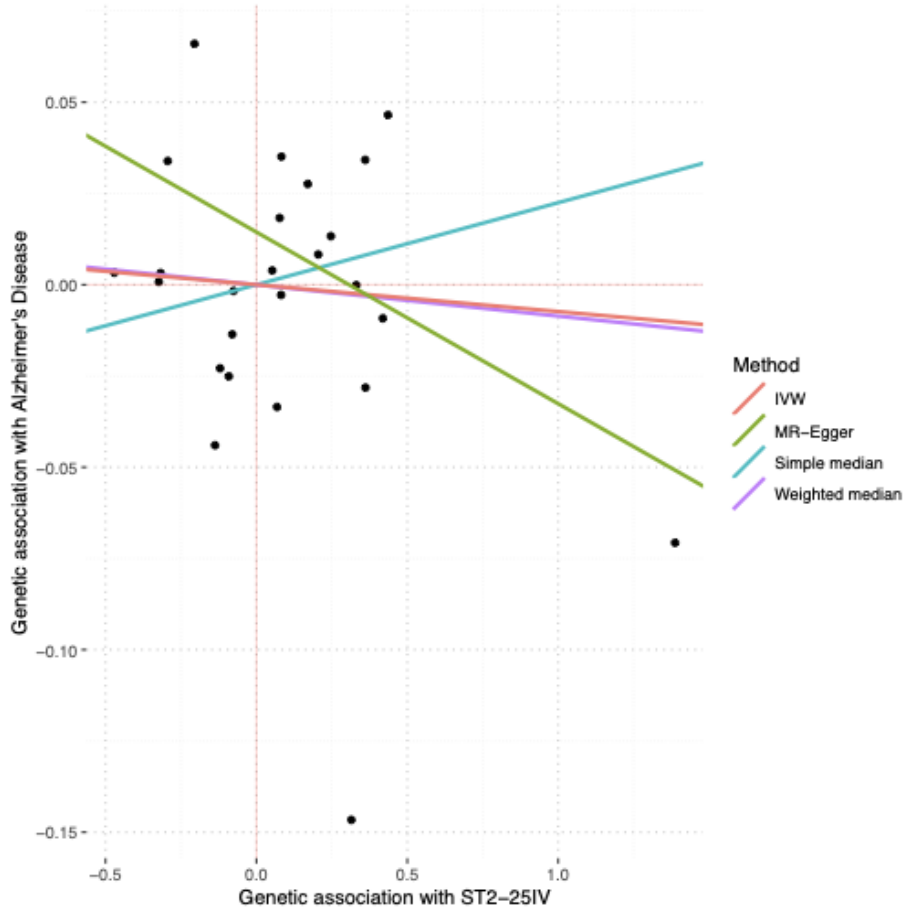
Residual standard error = 0.851
Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
Residual standard error is set to 1 in calculation of confidence interval when its estimate is less than 1.

Heterogeneity test statistic (Cochran's Q) = 0.7248 on 1 degrees of freedom, (p-value = 0.3946).
I^2 = 0.0%.
F statistic = 1730.2.



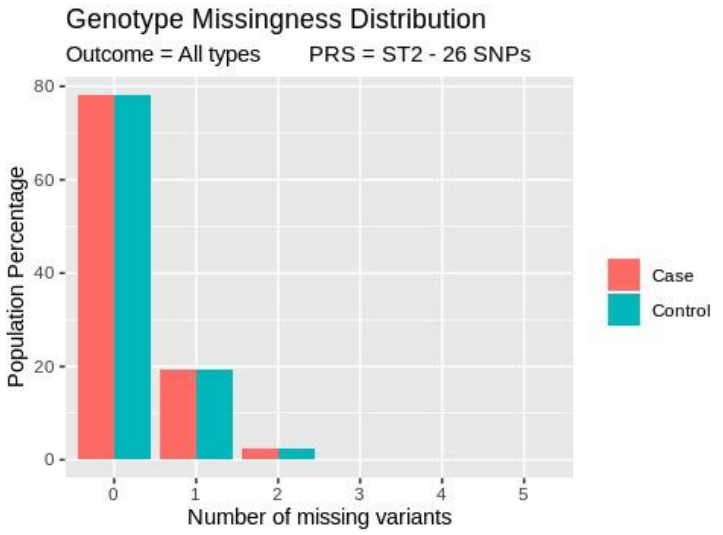
Summary-level MR – ST2 (25 IVs)

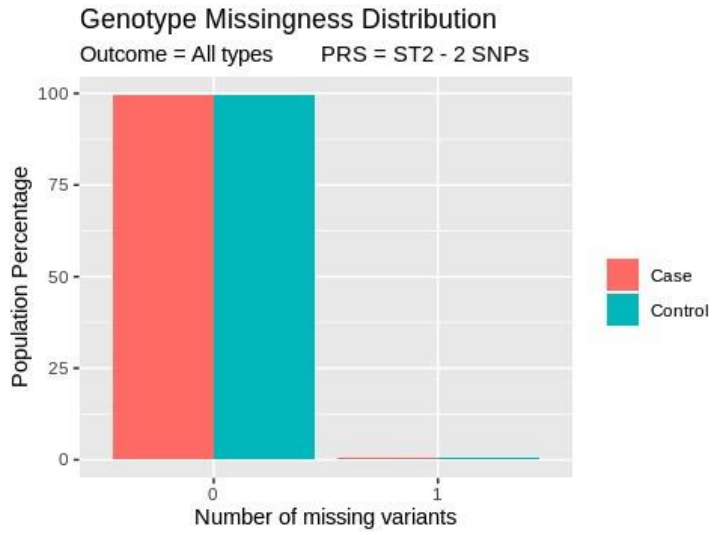
Method	Estimate	Std Error	95% CI		P-value
Simple median	0.022	0.043	-0.061	0.106	0.599
Weighted median	-0.009	0.029	-0.065	0.048	0.764
Penalized weighted median	-0.008	0.029	-0.065	0.048	0.768
IVW	-0.007	0.022	-0.051	0.036	0.742
Penalized IVW	-0.003	0.021	-0.043	0.038	0.888
Robust IVW	-0.005	0.020	-0.043	0.034	0.812
Penalized robust IVW	-0.003	0.019	-0.041	0.035	0.887
MR-Egger	-0.047	0.029	-0.104	0.010	0.106
(intercept)	0.014	0.007	0.000	0.029	0.048
Penalized MR-Egger	-0.044	0.028	-0.100	0.011	0.118
(intercept)	0.015	0.007	0.001	0.029	0.040
Robust MR-Egger	-0.049	0.020	-0.087	-0.010	0.013
(intercept)	0.018	0.009	0.000	0.035	0.045
Penalized robust MR-Egger	-0.049	0.020	-0.089	-0.009	0.015
(intercept)	0.018	0.009	0.001	0.034	0.042

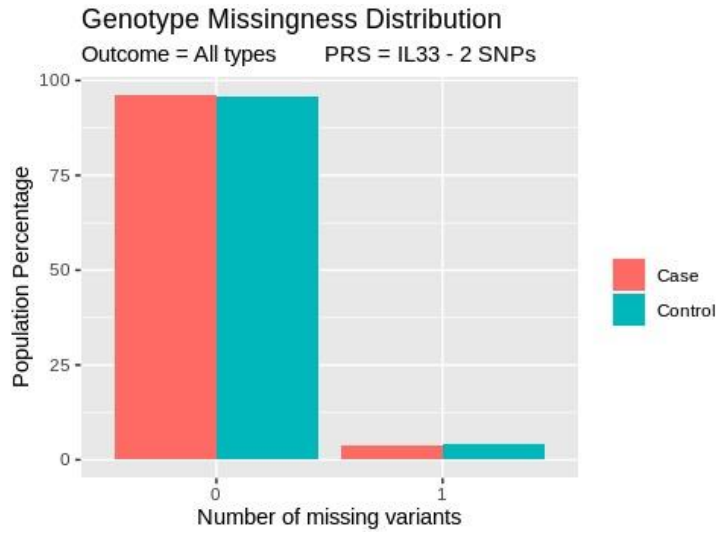


Genotype Missingness Distribution
All types

- Under 2 variants → Only individuals with both SNPs missing were excluded
- Under 26 variants → Only individuals with all 26 SNPs missing were excluded

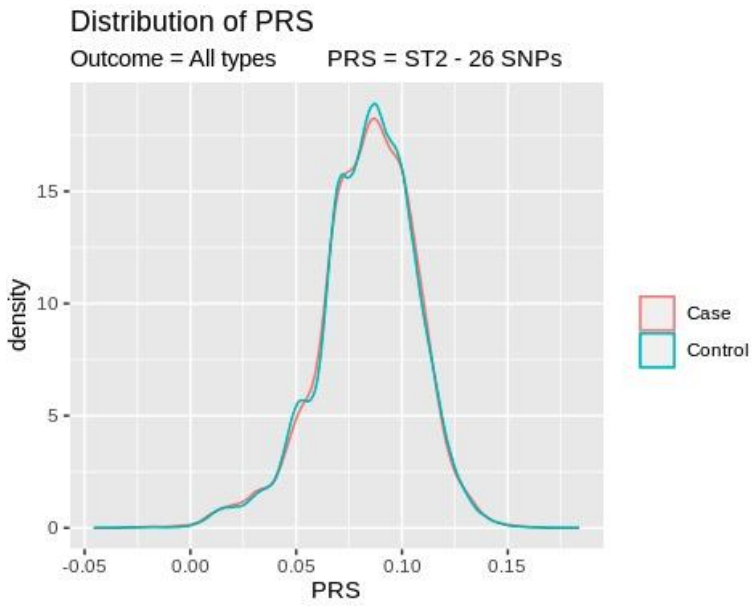


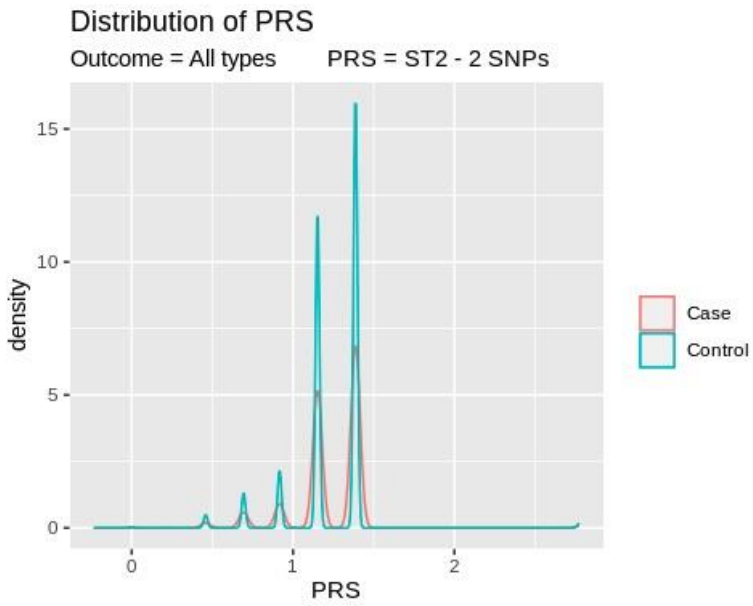


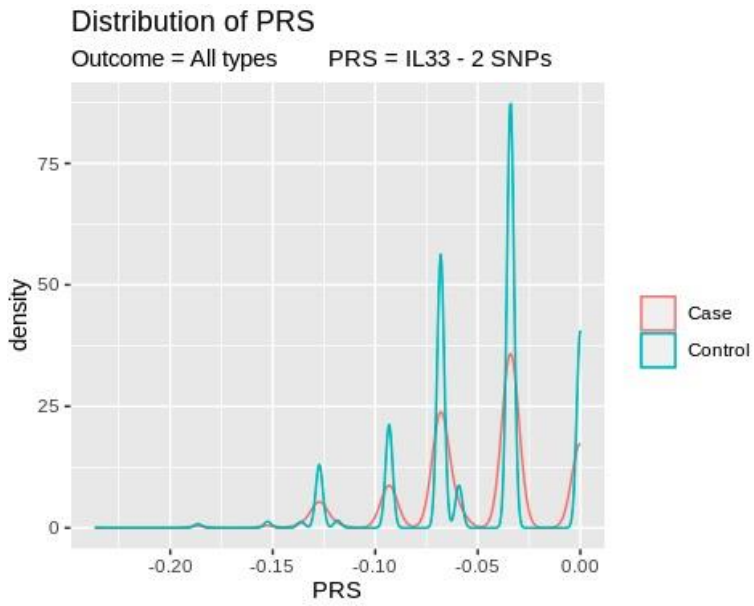


cases: 7140

PRS Distribution
All types

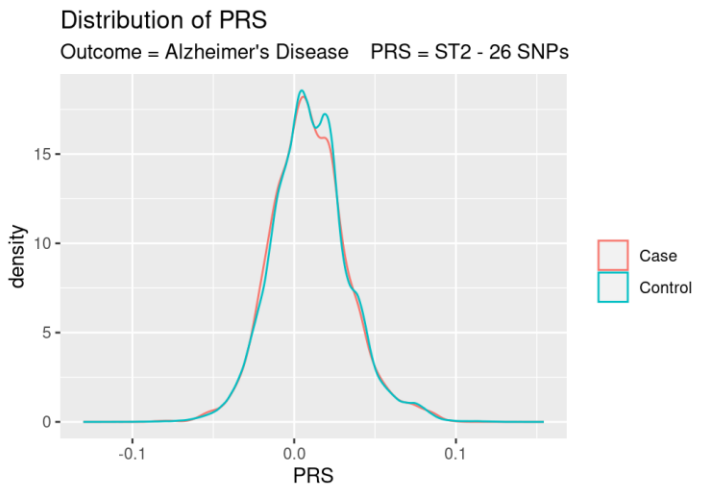


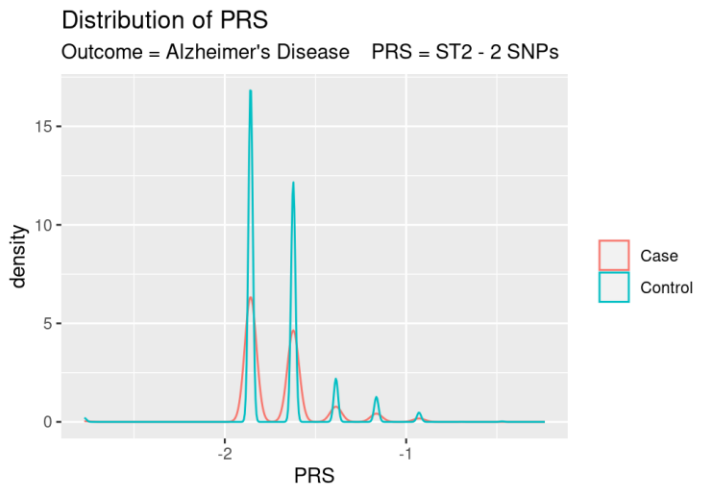


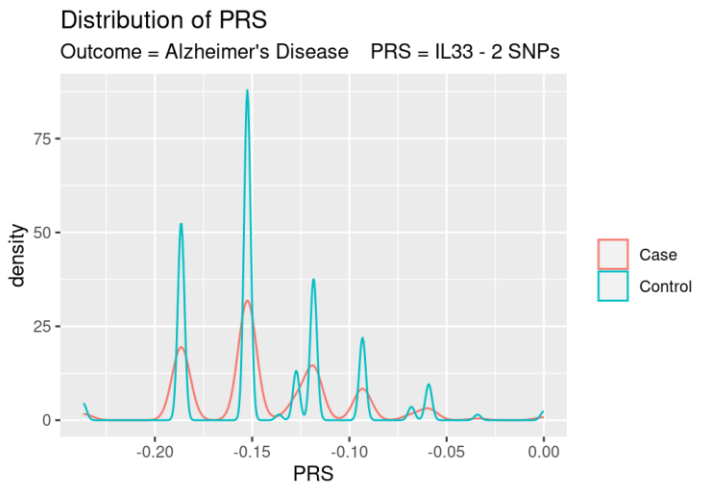


PRS Distribution

Alzhiemer's disease



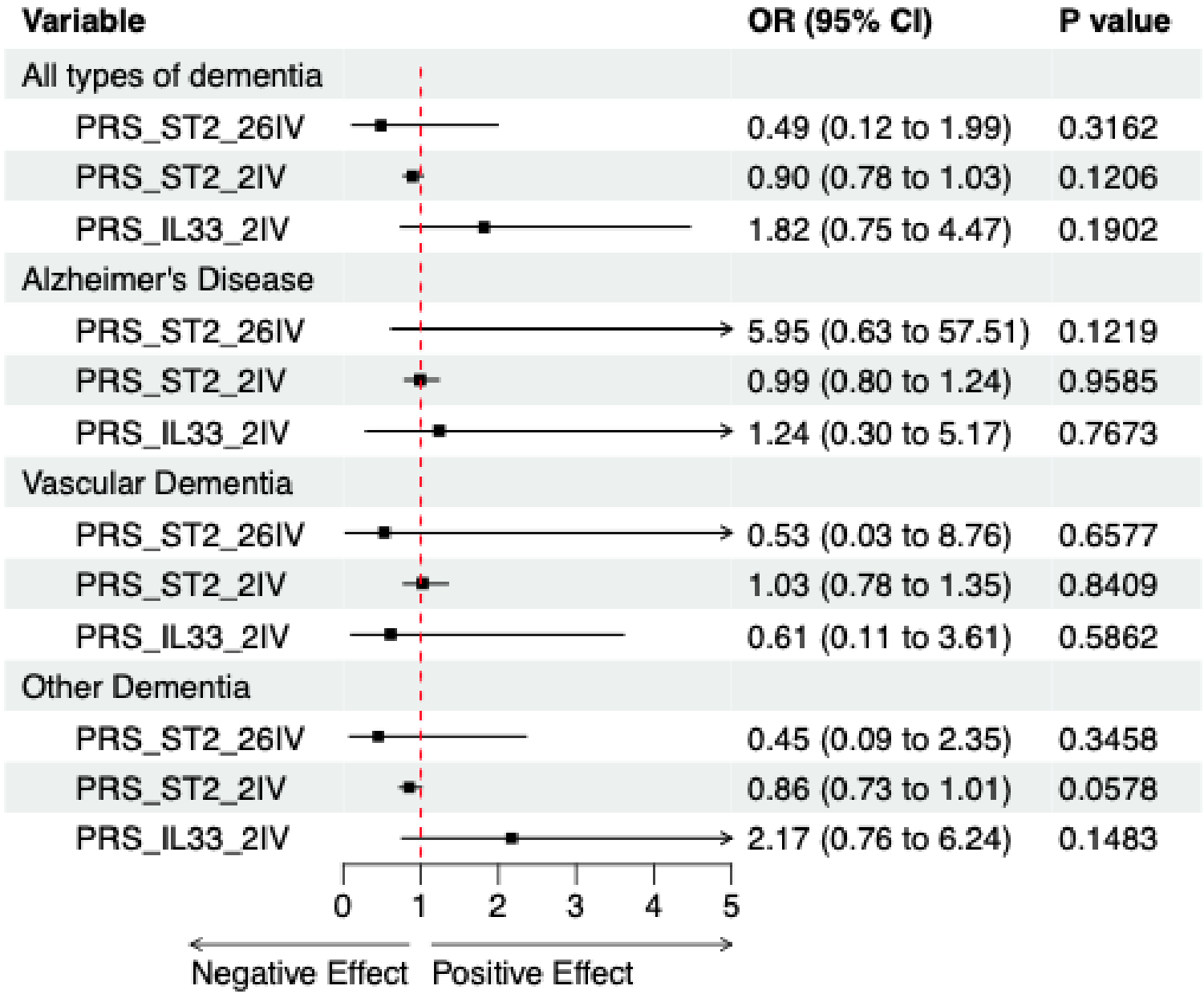




5. Individual-level MR – stratified by gender

Male Strata

Logistic regression: Dementia ~ PRS



IL-33 instrument – limitations

- The assessment of IL-33 PRS were hindered with several factors
  - The IL33 protein level in the UKBB samples were mostly undetectable (~ 85%)
  - Serum level of IL33 is not an appropriate marker and other studies use either tissue expression (mRNA) or surrogate markers such as Eosinophil count
- Asthma, as a top candidate for IL-33 does not seem to be correlated to IL-33 PRS, necessitating further work on the instrument which is beyond the scope of this project

Identification of ST2 and IL-33 as risk factors

- Dementia are a group of degenerative and progressive brain pathologies characterized by cognitive and memory deterioration
- The major types include Alzheimer's disease (AD) and Vascular Dementia (VD)
- Core mechanistic pathways that drive AD are still not fully understood
- Immune system and inflammatory processes play important roles in AD pathology
- IL33/ST2 axis is an immune-regulating pathway implicated in AD, impacting Aβ clearance
- We aimed to use MR to investigate possible links between IL33/ST2 and AD in the UKB