Online Data Supplement

Contents

Appendix 1. Details of our exposure GWAS conducted in UK BioBank

UKBiobank population and recruitment has been described in detail elsewhere.[2, 3]

We used all individuals with valid genotype and results for the traits of interest. We did not look for or exclude those with specific underlying diseases. Interstitial lung disease is very unlikely to affect these results. Interstitial lung diseases are rare affecting <0.01% of UK adults (https://statistics.blf.org.uk/lung-disease-uk-big-picture). Within UKBiobank only 108 people are known to have idiopathic pulmonary fibrosis, of which 61 are receiving treatment. 1,768 people in UKBiobank report doctor diagnosed COPD, of which 1,277 are on treatment. As our sample size was >300,000 such small numbers will not skew the results.

Although not specifically tested, we do not believe there is any cross over between our exposure and outcome populations. All analysis was performed using BOLT LMM using the IEU GWAS pipeline. This uses a linear mixed model (LMM) to account for both relatedness and population stratification, therefore allowing a wider range of individuals to be included. Only participants of European ancestry are used. A subset of 143,006 SNPs included in the model are directly genotyped. SNPs included are all meet the criteria:

Mean Allele Frequency >0.01
Genotyping rate >0.015
Hardy-Weinberg equilibrium p-value<0.0001
R^2 threshold of 0.1

For full details please see references.[4-6]

Forced Expiratory Volume in one second (FEV_1)

GWAS performed on 345,590 participants.
Quantitative trait that was measured as litres to three decimal places.
Mean FEV_1 = 2.853 (std = 0.780)
Estimated proportion of variance explained using inf model: 0.036

12,321,875 imputed SNPs in GWAS
44,522 SNPs reached significance at threshold of P_BOLT_LMM_INF $<5x10^{-8}$
360 SNPs remained after LD-clumping

When using for CAD;
41 proxies used
1 SNP removed due to incompatible alleles
7 SNPs removed as palindromic with intermediate allele frequencies
22 SNPs removed due to Steiger filtering
Leaving 307 SNPs available for analysis

When using for ischaemic stroke;
1 SNP removed for incompatible alleles
5 SNPs removed for being palindromic with intermediate allele frequencies
22 SNPs removed due Steiger filtering
Leaving 297 SNPs available for analysis

Forced Vital Capacity (FVC)

GWAS performed on 345,590 participants
Quantitative trait that was measured as litres to three decimal places.
Mean FVC = 3.782 (std = 0.985)
Estimated proportion of variance explained using inf model: 0.048

12,321,875 SNPs imputed SNPs in GWAS
58,873 SNPs reached significance at threshold of P_BOLT_LMM_INF $<5x10^{-8}$
464 SNPs remained after LD-clumping

When using for CAD;
60 Proxy SNPs found
15 SNPs removed for being palindromic with intermediate allele frequencies
14 SNPs removed due Steiger filtering
Leaving 406 SNPs available for analysis

When using for ischaemic stroke;
12 SNPs removed for being palindromic with intermediate allele frequencies
11 SNPs removed due Steiger filtering
Leaving 396 SNPs available for analysis

FEV_1/FVC <0.7

Binary trait that was measured the result of the ratio of FEV_1 and FVC. Recorded to two decimal places.

GWAS performed on 55,907 cases of FEV_1/FVC <0.7 with 297,408 controls (FEV_1/FVC ≥0.7)
Mean FEV_1/FVC 0.64 of cases. (SD 0.065)
Controls had mean ratio 0.77 (SD 0.03)
Estimated proportion of pseudo-variance explained using inf model: 0.009

12,321,875 SNPs imputed in GWAS
16,036 SNPs reached significance at threshold of P_BOLT_LMM_INF $<5x10^{-8}$
154 SNPs remained after LD-clumping
Range beta: -0.0188677 to 0.0270580

When using for CAD;
18 proxies found
4 SNPs removed as palindromic with intermediate allele frequencies
88 removed due Steiger filtering
Leaving 52 SNPs available for analysis

When using for ischaemic stroke;
3 SNPs removed for being palindromic with intermediate allele frequencies
56 removed due Steiger filtering
Leaving 40 SNPs available for analysis

Appendix 2. Details of the UK BioBank GWAS we used for covariates

All covariate GWAS were conducted by our colleagues at the IEU prior to this analysis being conducted. All are freely available on MRBase and at the IEU repository.[7, 8] Although not specifically tested, we do not believe there is any cross over between our exposure and outcome populations.

Height

Standing height of 461950 UKBiobank participants was used. The GWAS was performed in 2018.[8]
Quantitative trait recorded as centimetres.
9851866 SNPs imputed in GWAS
241226 reached significance at threshold of P_BOLT_LMM_INF $<5x10^{-8}$
990 SNPs remained after LD-clumping and removing a duplicate

When using for outcome CAD;
26 SNP proxies were found.
20 SNPs were removed for being palindromic with intermediate allele frequencies.
45 SNPs were removed due Steiger filtering for CAD.
908 SNPs available for analysis
per 1 SD decrease in height IVW OR: 1·19; 95% CI: 1·11-1·28, of CAD.

When using for outcome ischaemic stroke;
17 SNPs removed for being palindromic with intermediate allele frequencies.
62 SNPs were removed due Steiger filtering for ischaemic stroke.
892 SNPS available for analysis
per 1 SD decrease in height IVW OR: 1·02; 95% CI: 0·94-1·09, of ischaemic stroke.

BMI

BMI of 461460 UKBiobank participants was used. The GWAS was performed in 2018.[8]
Quantitative trait recorded as Kg/m^2
9851866 SNPs imputed in GWAS
68945 SNPs reached significance at threshold of P_BOLT_LMM_INF $<5x10^{-8}$
799 SNPs remained after LD-clumping

When using for outcome CAD;
32 SNPs were removed for being palindromic with intermediate allele frequencies.
8 SNPs were removed due Steiger filtering.
760 SNPs available for analysis
per 1 SD decrease in BMI IVW OR: 0·66; 95% CI: 0·62-0·71, risk CAD.

When using for outcome ischaemic stroke;
28 SNPs removed for being palindromic with intermediate allele frequencies.
6 SNPs were removed due Steiger filtering.
770 SNPS available for analysis
per 1 SD decrease in BMI IVW OR: 0·83; 95% CI: 0·78-0·89, risk ischaemic stroke.

Current Smoking

Current smoking of 462434 UKBiobank participants was used. The GWAS was performed in 2018.[8]
Ordered categorical trait. We do not know exactly how many cases and controls this involved, but more recent UKBIOBANK figures show that there are 55666 current smokers, 197787 previous smokers, and 317645 never smokers. The GWAS we used is highly likely to reflect very similar proportions.

9851867 SNPs imputed in GWAS
1949 SNPs reached significance at threshold of P_BOLT_LMM_INF $<5x10^{-8}$
37 SNPs remained after LD-clumping

When using for CAD;
1 SNP removed for being palindromic with intermediate allele frequencies.
18 SNPs were removed due Steiger filtering.
17 SNPs available for analysis
per 1 SD decrease in current smoking IVW OR: 0·61; 95% CI: 0·30-1·23, risk CAD.

When using for ischaemic stroke;
1 SNP removed for being palindromic with intermediate allele frequencies
22 SNPs were removed due Steiger filtering.

13 SNPs available for analysis
per 1 SD decrease in current smoking IVW OR: 1·00; 95% CI: 0·42-2·39, risk ischaemic stroke.

Appendix 3. Details of methods; F-statistic calculation, Steiger filtering, Palindromic SNPs and harmonisation. MR assumptions and explanation of sensitivity test

Details of method

For all exposure traits Cragg-Donald overall F-statistic was calculated.[9] The higher the F-statistic the lower the chance of weak instrument bias.[9] SNPs in close proximity may represent the same signal. Therefore, Linkage Disequilibrium-clumping (LD-clumping) was performed on all exposure SNPs. This retains the SNP with the most significant association at each locus (kb = 10000, r^2 0.001). Steiger filtering was performed to remove variants that caused more variance of the outcome than the exposure. [10] Steiger filtering estimates each SNP's rsq.exposure and rsq.outcome in the outcome population.[10] Those SNPs that explain more variance in the outcome than exposure are excluded, as they could led to a reverse causal relationship. SNPs were removed if they explained more variance of the outcome than the exposure.

For MVMR the LD-clumped and Steiger filtered SNPs for each lung function trait were combined with the LD-clumped and Steiger filtered SNPs for each covariate. The combined SNPs were then LD-clumped and duplicates removed. All SNPs had their beta effect extracted from both the lung function GWAS and the covariate GWAS to enable covariate conditioning.

For both the Shrine et al analysis and MVMR palindromic SNPs (i.e. A/T and C/G SNPs) with intermediate allele frequencies were excluded. The remaining SNPs were harmonised so that SNP-exposure and SNP-outcome effects corresponded to the same allele.[11] Proxies were identified for SNPs not found in outcome GWAS ($r^2 = 0.8$) for CAD, although this was not possible for stroke.

We assume that our instrumental variables (IV's) are associated with the exposure of interest. This assumption can be tested. We believe our IV's are strongly associated with the exposure given the stringent p-value threshold for significance. The Shrine et al [1] SNPs have gone through further testing in replication populations and have been tested via a polygenic risk score in multiple ancestry groups. The GWAS's for lung function traits that we performed ourselves showed they are responsible for a reasonable variation in the population. Our covariate GWAS's were performed to stringent p-value thresholds. F statistics for all exposures were above 10, reducing the chance of weak instrument bias.

Our second assumption is that our IV's influence our outcome only through the exposure. Our third assumption is that the IV's must not associate with measured or unmeasured confounding. These two hypotheses cannot be directly tested. However, we performed a number of tests in the 2SMR and MVMR models to reduce the risk the assumptions are violated. To account for the possibility of horizontal pleiotropy (IVs influence exposure and outcome through independent pathways) in our 2S-MR analysis, we performed MR Egger. MR-Egger is similar to IVW except the y intercept is unconstrained. If the y intercept of the MR-Egger is not equal to zero then either there is unbalanced horizontal pleiotropy (the average pleiotropic effect differs from zero) or the pleiotropic effects are independent from the genetic association with the risk factor, or both.[12] Although power is lower compared

to IVW, the gradient of the MR-Egger gives a causal estimate of the dose–response relationship between the genetic associations with the risk factor and those with the outcome, providing additional evidence for causal affect. MR-Egger method assumes that the Instrument Strength is Independent of the Direct Effect (InSIDE assumption), meaning that the SNPs pleiotropic effects are independent of their phenotypic effects.[13] For these reasons MR Egger is used as a sensitivity test rather than the main analysis. We used weighted median and mode MR methods to minimise the effect of unbalanced instruments on an overall estimate of the mean. A weighted median MR gives a consistent estimate of the causal effect when at least 50% of the weight comes from valid IVs, giving a greater robustness with strongly outlying causal estimates.[14] A weighted mode MR calculates an estimate based on the set of SNPs that form the largest homogenous cluster, which attempts to avoid the impact of invalid instruments.[15] As discussed in main paper results were consistent across sensitivity analyses demonstrating there is unlikely to be violation of assumptions.

We used a funnel plot to assess for horizontal pleiotropy by plotting the effect against its precision (beta against standard error).[7] A leave-one-out analysis was performed to ensure the results were not due to outliers with a large effect, by re-estimating the total effect after sequentially excluding one SNP at a time. Additionally, we performed a single-SNP analysis, where the effect of each SNP was individually assessed via IVW analysis and represented in a forest plot.

Heterogeneity (the variability in causal estimates obtained for each SNP) is an indication of potential violation of assumptions. This was calculated and assessed with a Q statistic.

Appendix 4. MVMR, conditioning for all covariates

We modelled effects of FEV_1 and FVC conditioning on all covariates. LD-clumped and Steiger filtered SNPs for every exposure trait were combined. The beta effect for each SNP was extracted from all exposure GWAS. SNPs not found in any of the exposure GWAS were removed. SNPs were LD-clumped (kb = 10000, r^2 0.001) and duplicates were removed. SNPs were extracted from outcome GWAS (proxies, $r^2 = 0.8$, used if not found for CAD). SNPs were harmonised and IVW effect estimate determined. As can be seen in Table S1, conditioning on all covariates shows similar results for FEV1 and FVC effect on risk of ischaemic stroke. The estimated effect for FVC on risk of CAD (OR: 1.44 per SD; 95% CI: 1.18-1·76) and FEV_1 on CAD (OR: 1·28 per SD; 95% CI: 1.02-1·61) are in the same direction but higher than the estimate when just conditioning with height. This is likely due to weak covariate instruments for two reasons. Firstly, the covariates are conditioned by each other in the model. Secondly, there are less covariate SNPs available for analysis due to removal of duplicates, LD-clumping and SNPs not being found in other covariate GWAS.

Table E1. Multivariable MR results of and FEV_1 and FVC on Coronary Artery Disease and Ischaemic Stroke conditioning with all covariates

Lung function trait	Condition	No. SNPs (LF/Hight/BMI/ Smoking)	OR (95% CI)* for Coronary Artery Disease	No. SNPs (LF/condition)	OR (95% CI)* for Ischaemic Stroke

FEV_1	Height/BMI/Smoking	80/432/391/4	1·28 (1·02, 1·61)	85/440/413/3	1·18 (0·94, 1·48)
FVC	Height/BMI/Smoking	105/406/388/4	1·44 (1·18, 1·76)	102/408/399/3	1·05 (0·86, 1·29)

Appendix 5. Results from 2S-MR using Shrine et al[1]

Shrine et al analysis

F-statistic for Shrine et al [1] exposures were; All traits=111, FEV_1=69, FVC=70, FEV_1/FVC=148, making weak instrument bias unlikely. Results for effects on CAD are reported in **Table 1**. Effects are per SD decrease in lung function trait.

Shrine et al lung function SNPs as exposure, CAD as outcome

IVW showed no consistent effect of decreasing lung function traits on risk of CAD. As can be seen in **Table 1** the direction of effect differs between the lung function traits assessed and the evidence is weak with almost all confidence intervals crossing one. IVW showed decreasing FEV_1/FVC had a protective effect on CAD (OR: 0·90 per SD; 95% CI: 0·82-0·99), although evidence was weaker in the sensitivity analysis. This may indicate the importance of decreased FVC compared FEV_1 on risk of CAD. There was strong evidence of heterogeneity of effects based on the Q p-value, however visual inspection of graphs used for sensitivity testing did not show that the effect was driven by any outliers. See supplementary information, appendix 6.

Table E1. Two-sample MR results of lung function traits on coronary artery disease using Shrine et al [1] and CARDIOGRAMplusC4D et al [16]

		Lung Function Trait (Exposure) effect on Coronary Artery Disease			
		FEV_1, FVC, FEV_1/FVC, PEF	FEV_1	FVC	FEV_1/FVC
No. SNPs used		173	60	67	93
IVW	OR per SD (95% CI)	0·95 (0·88 – 1·04)	1·14 (0·94 – 1·37)	1·01 (0·86 – 1·18)	0·90 (0·82 – 0·99)
	Q_p-value*	3.7×10^{-17}	3.07×10^{-8}	$3{\cdot}12\times10^{-6}$	$1{\cdot}1\times10^{-6}$
Weighted Median	OR per SD (95% CI)	0·95 (0·86 – 1·05)	1·24 (1·01 – 1·53)	0·98 (0·82 – 1·16)	0·96 (0·85 – 1·07)
Weighted Mode	OR per SD (95% CI)	0·92 (0·78 – 1·08)	0.80 (0·43 – 1.05)	0·93 (0·66 – 1·31)	0·96 (0·81 – 1·15)
MR-Egger	OR per SD (95% CI)	0·90 (0·72 – 1·12)	1·06 (0·57 - 1·96)	0·94 (0·51 – 1·74)	1.10 (0·86 – 1·42)

* Heterogenity test. If <0.05 it would suggest heterogenity

OR – Odds ratio; CI – Confidence Interval; IVW – Inverse Variance Weighting

Shrine et al lung function SNPs as exposure, ischaemic stroke as outcome

Examining the effect of the Shrine [1] et al. SNPs on the risk of ischaemic stroke showed a consistent direction of effect using IVW, weighted median and weighted mode, with all traits showing decreased lung function increases risk of ischaemic stroke as shown in **Table 2**. However the evidence is weak given the confidence intervals. There was evidence of heterogenity shown by the Q_p-values but plot sensitivity anlysis did not suggest there were any outliers requiring removal from analysis. Sensitivity anlysis using MR-Egger showed that for FEV_1 the direction of effect was showing a protective effect of reduced lung function (OR: 0·91 per SD; 95% CI: 0·55-1·49), however this test is of a lower statistical power as evidenced by the wide confidence interval.

Table E2. Two-sample MR results of lung function traits on ischaemic stroke using Shrine et al [1] and MEGASTROKE [16]

		Lung Function Trait (Exposure) effect on Ischaemic Stroke			
		FEV_1, FVC, FEV_1/FVC, PEF	FEV_1	FVC	FEV_1/FVC
No. SNPs used		171	58	77	93
IVW	OR per SD (95% CI)	1·05 (0·98 – 1·13)	1·01 (0·87 – 1·18)	1·04 (0·91 – 1·19)	1·04 (0·95 – 1·13)
	Q_p-value	2.2×10^{-5}	0.005	0.04	0.01
Weighted Median	OR per decrease SD (95% CI)	1·05 (0.95 – 1·64)	1·05 (0·87 – 1·28)	1·11 (0·93 – 1·32)	1·02 (0·91 – 1·14)
Weighted Mode	OR per SD (95% CI)	1·08 (0·93 – 1·26)	1·07 (0·77 - 1·51)	1·06 (0·77 – 1·46)	1·02 (0·85 – 1·12)
MR-Egger	OR per SD (95% CI)	1·13 (0·93 – 1·36)	0·91 (0·55 - 1·49)	1·47 (0·87 – 2·48)	1·05 (0·84 – 1·31)

OR – Odds ratio; CI – Confidence Interval; IVW – Inverse Variance Weighting

Appendix 6. Figures from 2SMR using Shrine et al [1]

Please note, our results in Tables 1-4 in the main article show effect per *decrease* in FVC. These figures are plotted as per *increase* in FVC.

Figure E1. Funnel plot of heterogeneity of causal effects of FVC using Shrine et al [1] on coronary artery disease

Each point is a SNP with its beta plotted against its inverse standard error. As the graph is funnel shaped, it indicates no heterogeneity.

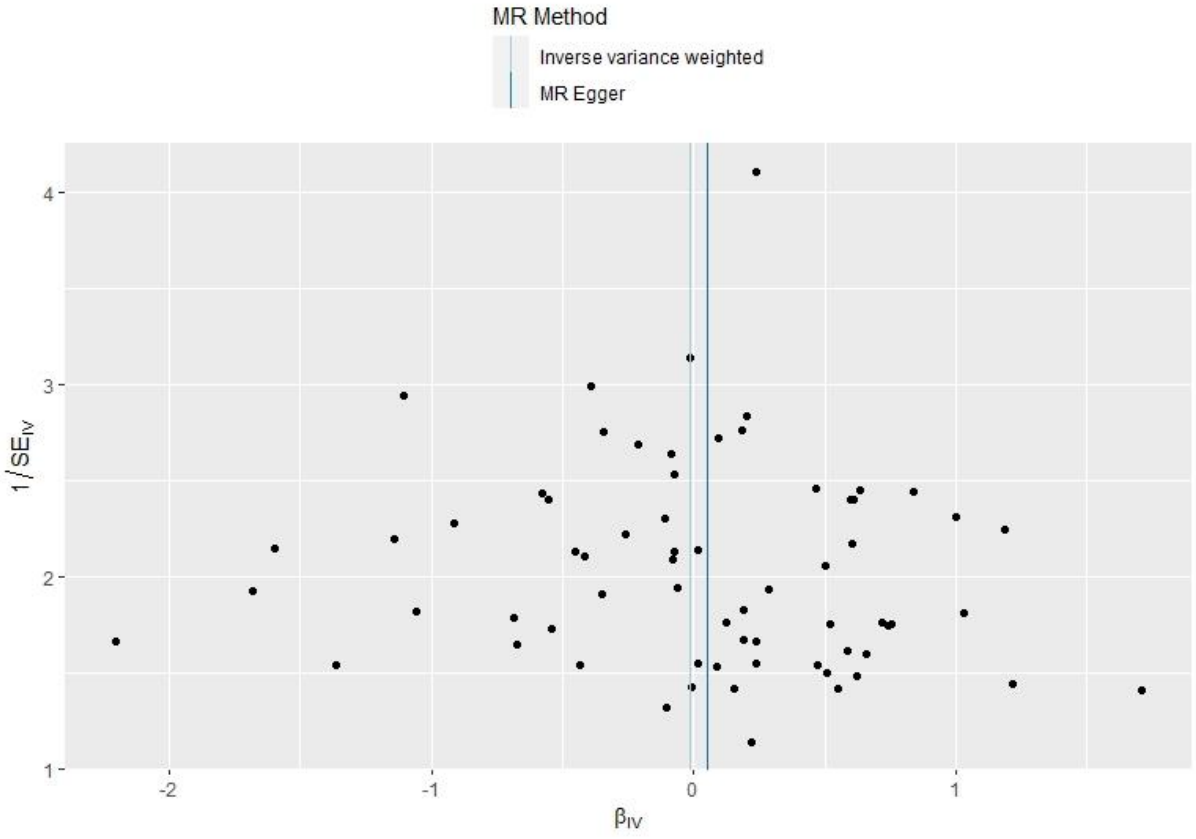


Figure E2. Scatter plot of the SNP-effect on FVC using Shrine et al [1] and SNP-effect on coronary artery disease

Each point on the graph represents the SNP-outcome association plotted against the SNP-exposure association. Bars indicate 95% confidence intervals. Coloured lines represent analysis method used. This shows no effect of FVC on coronary artery disease. MR Egger intercept is close to zero indicating no unbalanced directional pleiotropy.

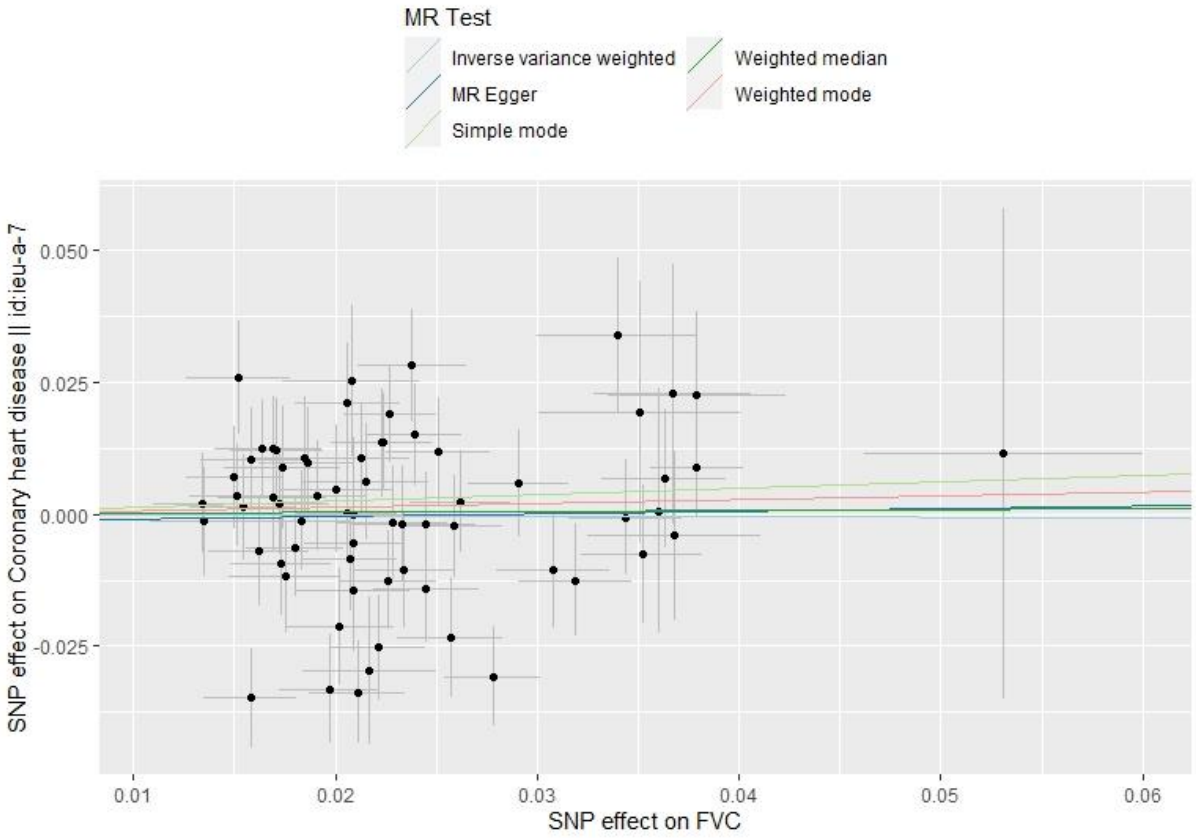


Figure E3. Leave-one-out analysis of FVC using Shrine et al [1] on coronary artery disease

Each point represents the IVW estimate if the SNP on the y axis was left out of total analysis. Bars indicate 95% confidence intervals, demostrating that no individual SNP is driving the causal effect estimate.

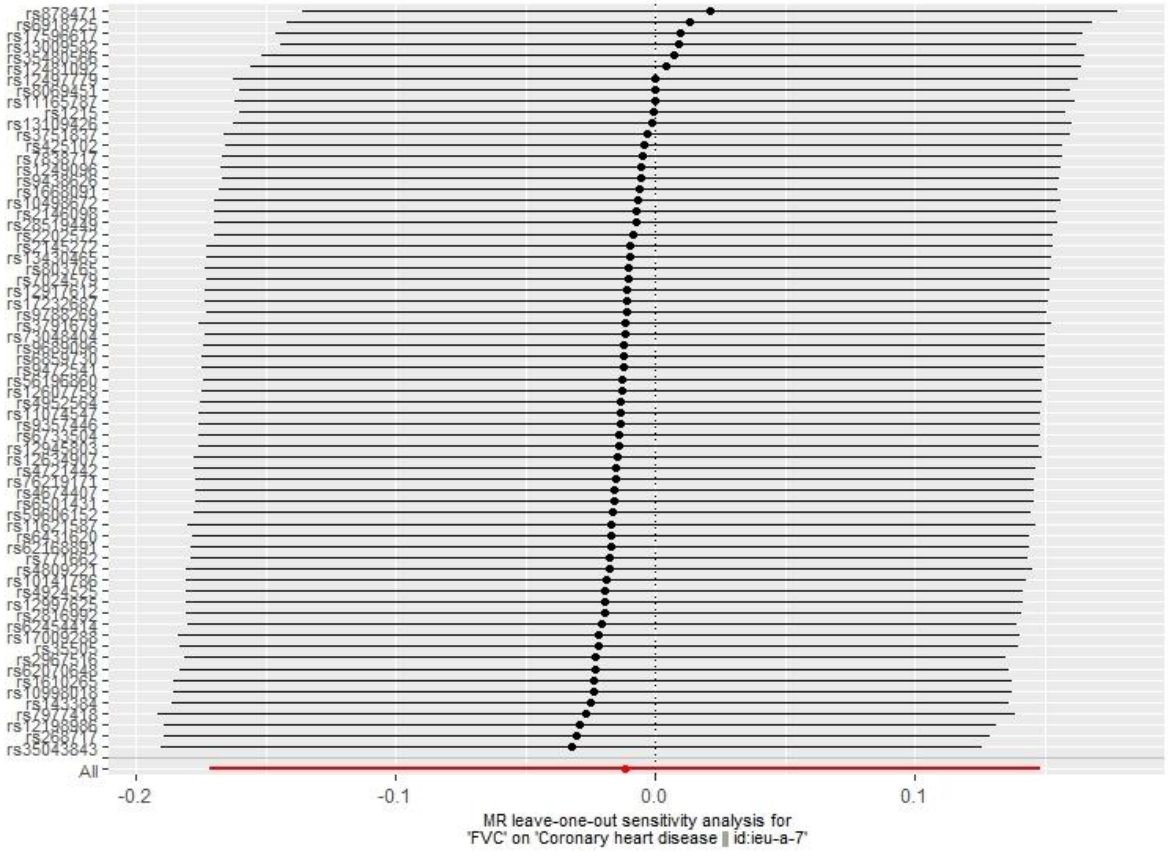


Figure E4. Single SNP analysis of FVC using Shrine et al [1] on coronary artery disease

Each point represents individual SNP calculated effect size for FVC on the odds of coronary artery disease. Bars indicate 95% CI.

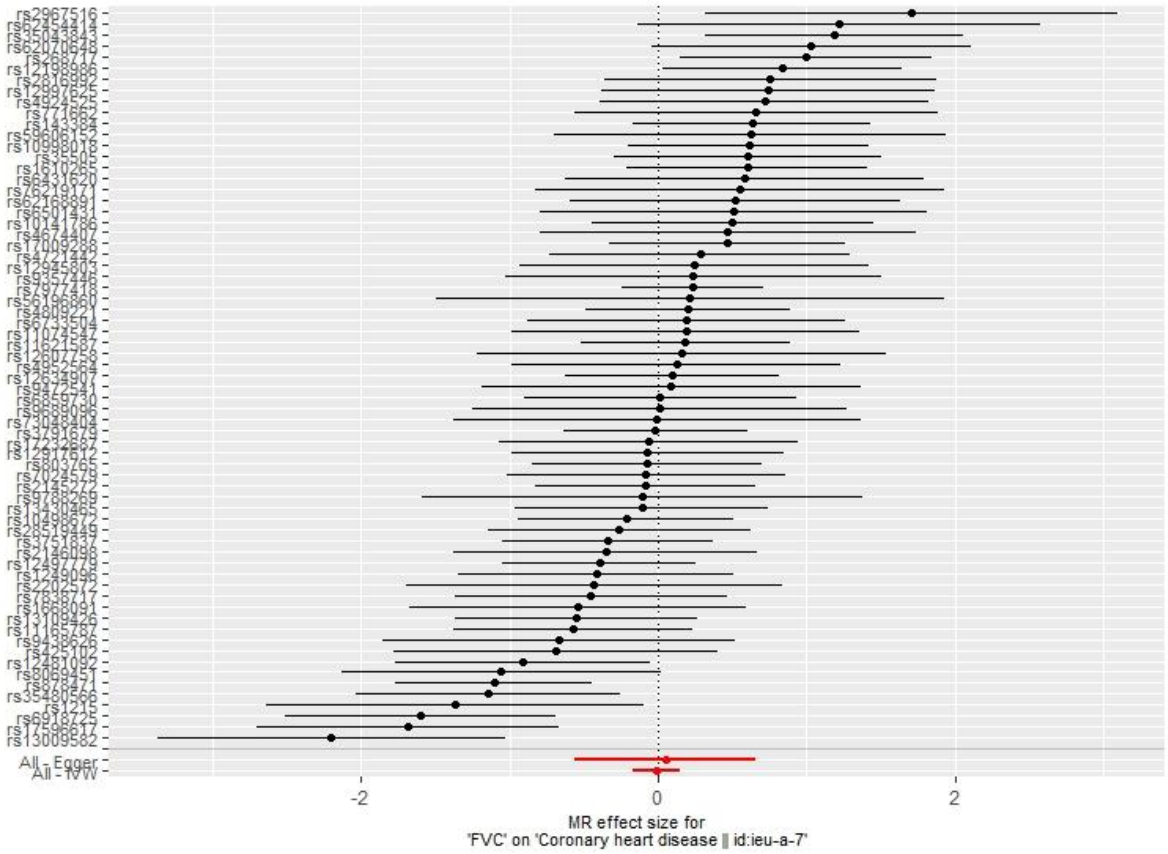
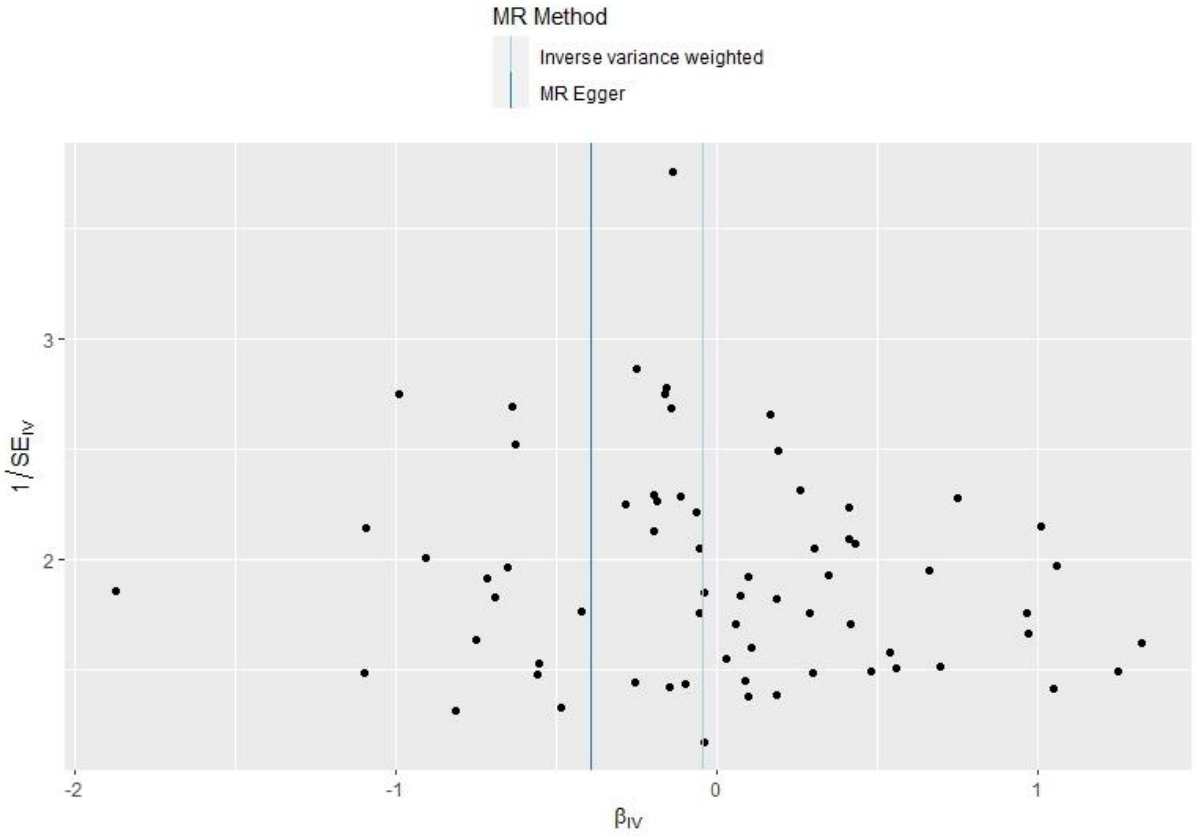


Figure E5. Funnel plot of heterogeneity of causal effects of FVC using Shrine et al [1] on ischaemic stroke



No evidence heterogeneity shown

Figure E6. Scatter plot of the SNP-effect on FVC using Shrine et al [1] and SNP-effect on ischaemic stroke

MR Egger line via close to zero, indicating no directional pleiotropy

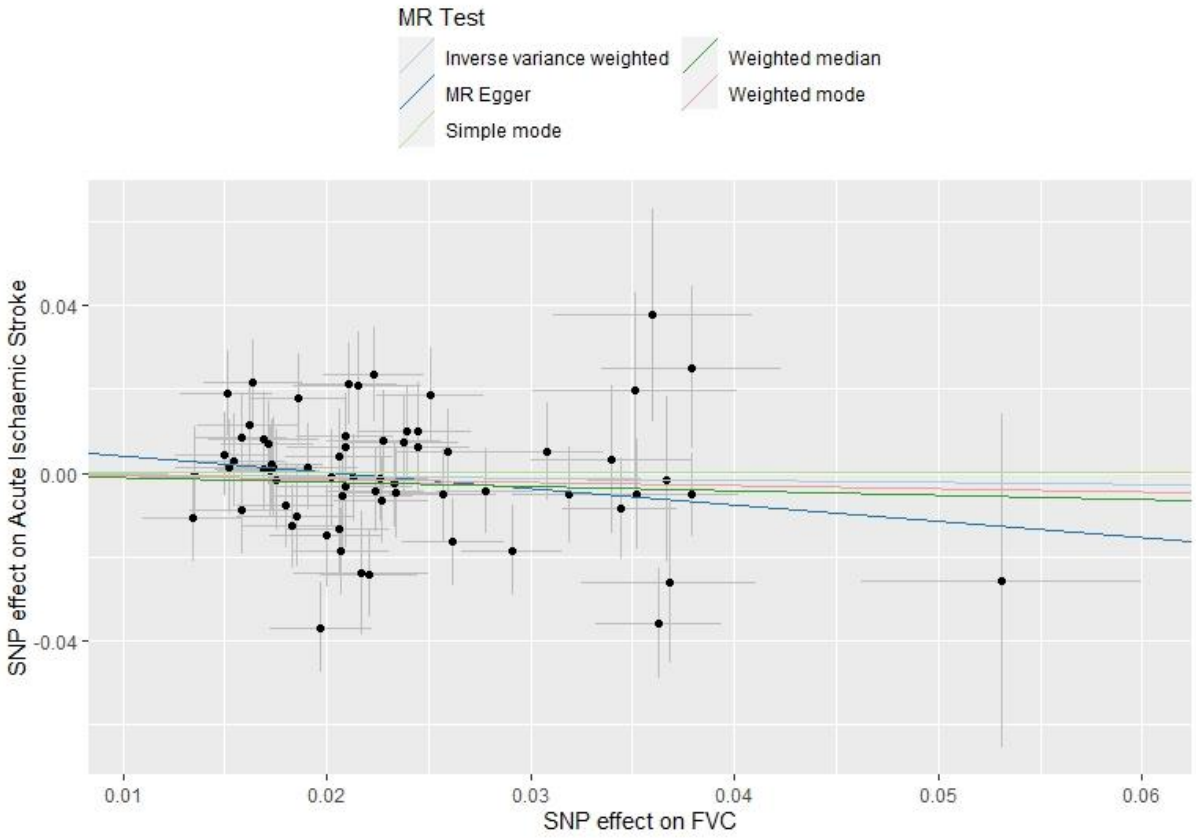


Figure E7. Leave-one-out analysis of FVC using Shrine et al [1] on ischaemic stroke

No individual SNP driving causal estimate

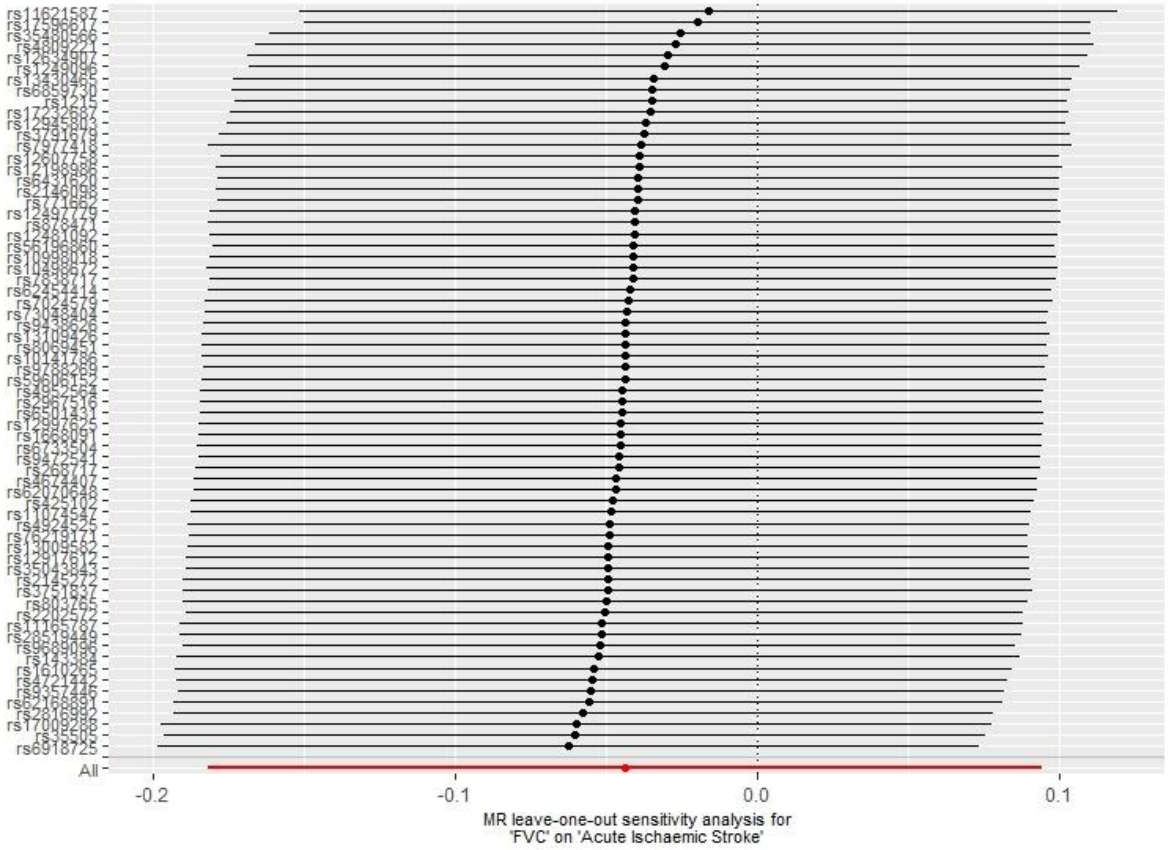
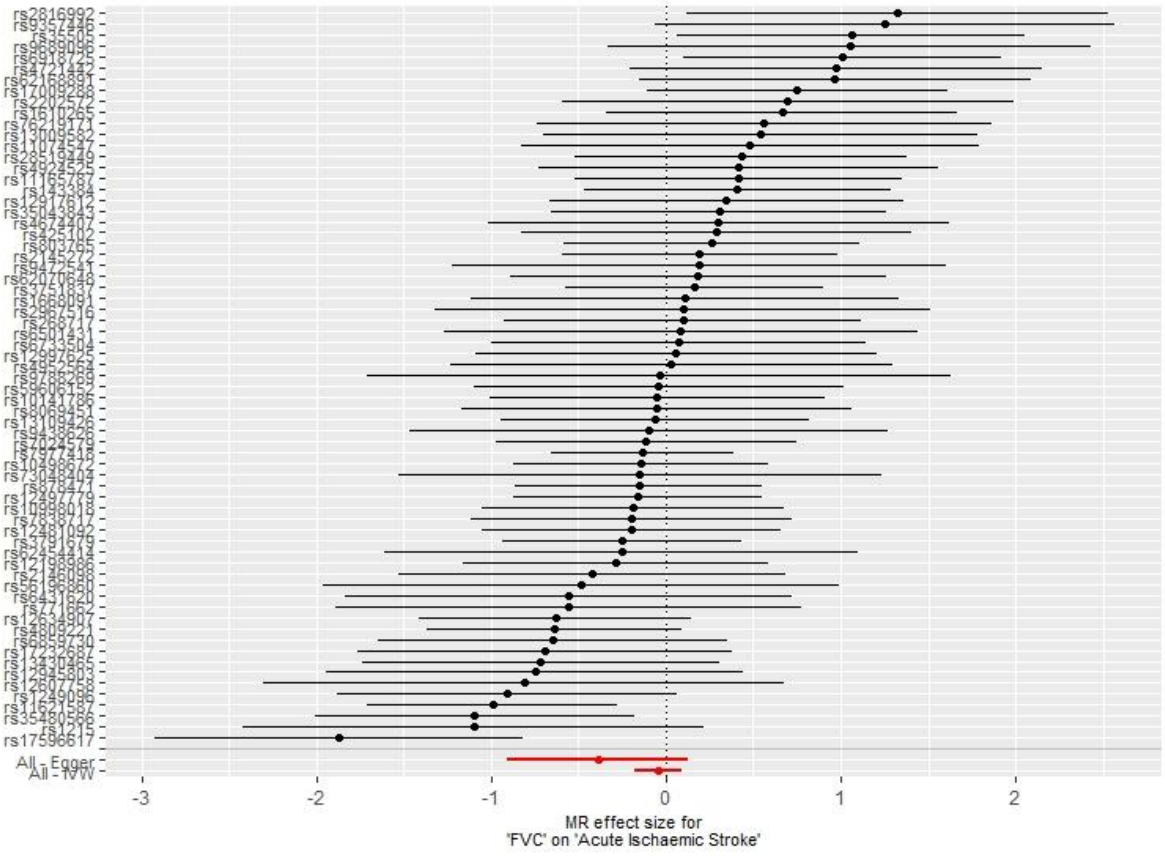


Figure E8. Single SNP analysis of FVC using Shrine et al [1] on ischaemic stroke



Appendix 7. Data sharing

Shrine et all SNPs are available from the supplementary information of the reference.[1]

Summary statistics from our lung function GWAS performed in UKBIOBANK will be available on MRBase and IEU repository within 1 month of publication of this paper.[7, 8]

Summary statistics for the covariate GWAS are currently available on MRBase and IEU repository.[7, 8]

Summary statistics for CAD outcome are available on MRBase and from authors.[7, 16]

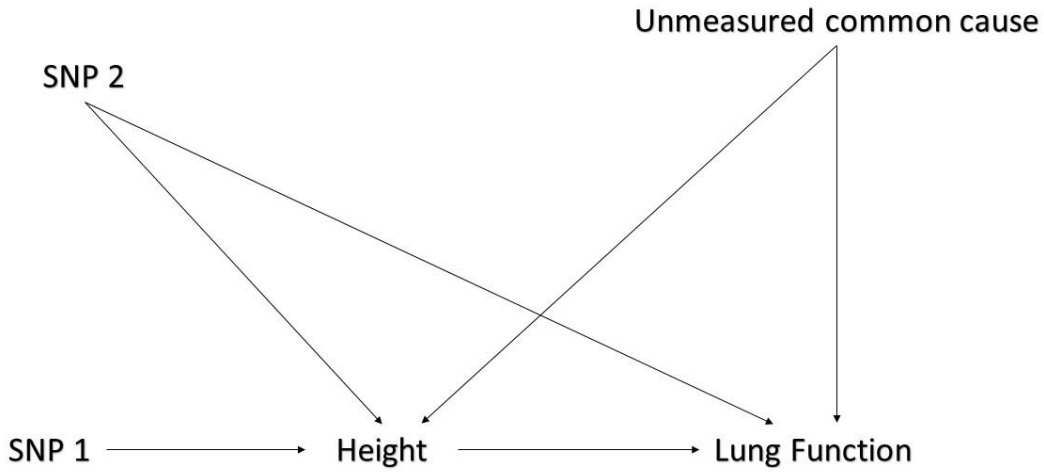
Summary statistics for ischaemic stroke are available for download here.[17]

Please contact the corresponding author for requests for code.

Appendix 8. Bias due to covariate adjustment in GWAS

Adjustment of covariates in GWAS has been shown to affect the SNP-exposure estimate leading to bias in MR studies. Figure E9 and legend explain some possible pathways. However, if there is residual confounding, covariate adjustment will bias the MR estimate for many different structures between exposure-covariate relationship. This reference explains in further detail.[18]

Figure E9. Directed acyclic graph demonstrating possible pathways leading to bias



Height is used in this example, but it would be true of other covariates, such as smoking.

SNP 1 does not have a direct effect on lung function. However, height and lung function have unmeasured common causes. Therefore, height is a collider of the path of SNP 1 and lung function (SNP 1 → Height ← Unmeasured common causes → Lung function). Adjusting on a collider opens the path on the collider, which means that adjusting for height will wrongly identify SNP 1 as having a direct effect on lung function.

SNP 2 has a direct effect on lung function, but also an indirect effect via height. Adjusting for height will cause a biased estimate for the direct effect of SNP 2 on lung function (as it would be combination of the true direct effect and the bias from the collider adjustment).

To eliminate the collider bias induced by the covariable adjustment, the GWAS would need to also be adjusted for all unmeasured common causes, but this is impossible as they are unmeasured.

Appendix 9. Extended Acknowledgements

We would like to acknowledge the work down by the authors of Shrine et al.[1] We would like to acknowledge all those at the IEU that worked on the IEU pipeline and performed previous GWAS.[4, 8] We would like to thank all participants and researchers of the UKBiobank.[2] We acknowledge the work done to perform both the outcome GWAS.[16, 19]

The MEGASTROKE project received funding from sources specified at http://www.megastroke.org/acknowledgments.html

As per request by the authors of the MEGASTROKE project we list all of their names here. To view their affliations please see link http://megastroke.org/authors.html.

Rainer Malik [1], Ganesh Chauhan [2], Matthew Traylor [3], Muralidharan Sargurupremraj [4,5], Yukinori Okada [6,7,8], Aniket Mishra [4,5], Loes Rutten-Jacobs [3], Anne-Katrin Giese [9], Sander W van der Laan [10], Solveig Gretarsdottir [11], Christopher D Anderson [12,13,14,14], Michael Chong [15], Hieab HH Adams [16,17], Tetsuro Ago [18], Peter Almgren [19], Philippe Amouyel [20,21], Hakan Ay [22,13], Traci M Bartz [23], Oscar R Benavente [24], Steve Bevan [25], Giorgio B Boncoraglio [26], Robert D Brown, Jr. [27], Adam S Butterworth [28,29], Caty Carrera [30,31], Cara L Carty [32,33], Daniel I Chasman [34,35], Wei-Min Chen [36], John W Cole [37], Adolfo Correa [38], Ioana Cotlarciuc [39], Carlos Cruchaga [40,41], John Danesh [28,42,43,44], Paul IW de Bakker [45,46], Anita L DeStefano [47,48], Marcel den Hoed [49], Qing Duan [50], Stefan T Engelter [51,52], Guido J Falcone [53,54], Rebecca F Gottesman [55], Raji P Grewal [56], Vilmundur Gudnason [57,58], Stefan Gustafsson [59], Jeffrey Haessler [60], Tamara B Harris [61], Ahamad Hassan [62], Aki S Havulinna [63,64], Susan R Heckbert [65], Elizabeth G Holliday [66,67], George Howard [68], Fang-Chi Hsu [69], Hyacinth I Hyacinth [70], M Arfan Ikram [16], Erik Ingelsson [71,72], Marguerite R Irvin [73], Xueqiu Jian [74], Jordi Jiménez-Conde [75], Julie A Johnson [76,77], J Wouter Jukema [78], Masahiro Kanai [6,7,79], Keith L Keene [80,81], Brett M Kissela [82], Dawn O Kleindorfer [82], Charles Kooperberg [60], Michiaki Kubo [83], Leslie A Lange [84], Carl D Langefeld [85], Claudia Langenberg [86], Lenore J Launer [87], Jin-Moo Lee [88], Robin Lemmens [89,90], Didier Leys [91], Cathryn M Lewis [92,93], Wei-Yu Lin [28,94], Arne G Lindgren [95,96], Erik Lorentzen [97], Patrik K Magnusson [98], Jane Maguire [99], Ani Manichaikul [36], Patrick F McArdle [100], James F Meschia [101], Braxton D Mitchell [100,102], Thomas H Mosley [103,104], Michael A Nalls [105,106], Toshiharu Ninomiya [107], Martin J O'Donnell [15,108], Bruce M Psaty [109,110,111,112], Sara L Pulit [113,45], Kristiina Rannikmäe [114,115], Alexander P Reiner [65,116], Kathryn M Rexrode [117], Kenneth Rice [118], Stephen S Rich [36], Paul M Ridker [34,35], Natalia S Rost [9,13], Peter M Rothwell [119], Jerome I Rotter [120,121], Tatjana Rundek [122], Ralph L Sacco [122], Saori Sakaue [7,123], Michele M Sale [124], Veikko Salomaa [63], Bishwa R Sapkota [125], Reinhold Schmidt [126], Carsten O Schmidt [127], Ulf Schminke [128], Pankaj Sharma [39], Agnieszka Slowik [129], Cathie LM Sudlow [114,115], Christian Tanislav [130], Turgut Tatlisumak [131,132], Kent D Taylor [120,121], Vincent NS Thijs [133,134], Gudmar Thorleifsson [11], Unnur Thorsteinsdottir [11], Steffen Tiedt [1], Stella Trompet [135], Christophe Tzourio [5,136,137], Cornelia M van Duijn [138,139], Matthew Walters [140], Nicholas J Wareham [86], Sylvia Wassertheil-Smoller [141], James G Wilson [142], Kerri L Wiggins [109], Qiong Yang [47], Salim Yusuf [15], Najaf Amin [16], Hugo S Aparicio [185,48], Donna K Arnett [186], John Attia [187], Alexa S Beiser [47,48], Claudine Berr [188], Julie E Buring [34,35], Mariana Bustamante [189], Valeria Caso [190], Yu-Ching Cheng [191], Seung Hoan Choi [192,48], Ayesha Chowhan [185,48], Natalia Cullell [31], Jean-François Dartigues [193,194], Hossein Delavaran [95,96], Pilar Delgado [195], Marcus Dörr [196,197], Gunnar Engström [19], Ian Ford [198], Wander S Gurpreet [199], Anders Hamsten [200,201], Laura Heitsch [202], Atsushi Hozawa [203], Laura Ibanez [204], Andreea Ilinca [95,96], Martin Ingelsson [205], Motoki Iwasaki [206], Rebecca D Jackson [207], Katarina Jood [208], Pekka Jousilahti [63], Sara Kaffashian [4,5], Lalit Kalra [209], Masahiro Kamouchi [210], Takanari Kitazono [211], Olafur Kjartansson [212], Manja Kloss [213], Peter J Koudstaal [214], Jerzy Krupinski [215], Daniel L Labovitz [216], Cathy C Laurie [118], Christopher R Levi [217], Linxin Li [218], Lars Lind [219], Cecilia M Lindgren [220,221], Vasileios Lioutas [222,48], Yong Mei Liu [223], Oscar L Lopez [224], Hirata Makoto [225], Nicolas Martinez-Majander [172], Koichi Matsuda [225], Naoko Minegishi [203], Joan Montaner [226], Andrew P Morris [227,228], Elena Muiño [31], Martina Müller-Nurasyid [229,230,231], Bo Norrving [95,96], Soichi Ogishima [203], Eugenio A Parati [232], Leema Reddy Peddareddygari [56], Nancy L Pedersen [98,233], Joanna Pera [129], Markus Perola [63,234], Alessandro Pezzini [235], Silvana Pileggi [236], Raquel Rabionet [237], Iolanda Riba-Llena [30], Marta Ribasés [238], Jose R Romero [185,48], Jaume Roquer [239,240], Anthony G Rudd [241,242], Antti-Pekka Sarin [243,244], Ralhan Sarju [199], Chloe Sarnowski [47,48], Makoto Sasaki [245], Claudia L Satizabal [185,48], Mamoru Satoh [245], Naveed Sattar [246], Norie Sawada [206], Gerli Sibolt [172], Ásgeir Sigurdsson [247], Albert

Smith [248], Kenji Sobue [245], Carolina Soriano-Tárraga [240], Tara Stanne [249], O Colin Stine [250], David J Stott [251], Konstantin Strauch [229,252], Takako Takai [203], Hideo Tanaka [253,254], Kozo Tanno [245], Alexander Teumer [255], Liisa Tomppo [172], Nuria P Torres-Aguila [31], Emmanuel Touze [256,257], Shoichiro Tsugane [206], Andre G Uitterlinden [258], Einar M Valdimarsson [259], Sven J van der Lee [16], Henry Völzke [255], Kenji Wakai [253], David Weir [260], Stephen R Williams [261], Charles DA Wolfe [241,242], Quenna Wong [118], Huichun Xu [191], Taiki Yamaji [206], Dharambir K Sanghera [125,169,170], Olle Melander [19], Christina Jern [171], Daniel Strbian [172,173], Israel Fernandez-Cadenas [31,30], W T Longstreth, Jr [174,65], Arndt Rolfs [175], Jun Hata [107], Daniel Woo [82], Jonathan Rosand [12,13,14], Guillaume Pare [15], Jemma C Hopewell [176], Danish Saleheen [177], Kari Stefansson [11,178], Bradford B Worrall [179], Steven J Kittner [37], Sudha Seshadri [180,48], Myriam Fornage [74,181], Hugh S Markus [3], Joanna MM Howson [28], Yoichiro Kamatani [6,182], Stephanie Debette [4,5], Martin Dichgans [1,183,184]

References

1. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, Batini C, Fawcett KA, Song K, Sakornsakolpat P, Li X, Boxall R, Reeve NF, Obeidat Me, Zhao JH, Wielscher M, Weiss S, Kentistou KA, Cook JP, Sun BB, Zhou J, Hui J, Karrasch S, Imboden M, Harris SE, Marten J, Enroth S, Kerr SM, Surakka I, Vitart V, Lehtimäki T, Allen RJ, Bakke PS, Beaty TH, Bleecker ER, Bossé Y, Brandsma C-A, Chen Z, Crapo JD, Danesh J, DeMeo DL, Dudbridge F, Ewert R, Gieger C, Gulsvik A, Hansell AL, Hao K, Hoffman JD, Hokanson JE, Homuth G, Joshi PK, Joubert P, Langenberg C, Li X, Li L, Lin K, Lind L, Locantore N, Luan Ja, Mahajan A, Maranville JC, Murray A, Nickle DC, Packer R, Parker MM, Paynton ML, Porteous DJ, Prokopenko D, Qiao D, Rawal R, Runz H, Sayers I, Sin DD, Smith BH, Soler Artigas M, Sparrow D, Tal-Singer R, Timmers PRHJ, Van den Berge M, Whittaker JC, Woodruff PG, Yerges-Armstrong LM, Troyanskaya OG, Raitakari OT, Kähönen M, Polašek O, Gyllensten U, Rudan I, Deary IJ, Probst-Hensch NM, Schulz H, James AL, Wilson JF, Stubbe B, Zeggini E, Jarvelin M-R, Wareham N, Silverman EK, Hayward C, Morris AP, Butterworth AS, Scott RA, Walters RG, Meyers DA, Cho MH, Strachan DP, Hall IP, Tobin MD, Wain LV, Understanding Society Scientific G. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nature Genetics* 2019: 51(3): 481-493.
2. https://www.ukbiobank.ac.uk/. [cited; Available from:
3. Trehearne A. Genetics, lifestyle and environment. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 2016: 59(3): 361-367.
4. Ruth Mitchell GH, Tom Dudding, Laura Corbin, Sean Harrison, Lavinia Paternoster. UK BioBank Genetic Data: MRC-IEU Quality Control, version 2. 2019.
5. Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nat Genet* 2018: 50(7): 906-908.
6. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, Patterson N, Price AL. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015: 47(3): 284-290.
7. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 2018: 7.
8. IEU. https://gwas.mrcieu.ac.uk/. 2020.
9. Burgess S, Thompson SG, Collaboration CCG. Avoiding bias from weak instruments in Mendelian randomization studies. *International journal of epidemiology* 2011: 40(3): 755-764.
10. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS genetics* 2017: 13(11): e1007081.

11. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *Am J Clin Nutr* 2016: 103(4): 965-978.
12. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International journal of epidemiology* 2015: 44(2): 512-525.
13. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol* 2017: 32(5): 377-389.
14. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. 2016: 40(4): 304-314.
15. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Human Molecular Genetics* 2018: 27(R2): R195-R208.
16. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, Webb TR, Zeng L, Dehghan A, Alver M, Armasu SM, Auro K, Bjonnes A, Chasman DI, Chen S, Ford I, Franceschini N, Gieger C, Grace C, Gustafsson S, Huang J, Hwang S-J, Kim YK, Kleber ME, Lau KW, Lu X, Lu Y, Lyytikäinen L-P, Mihailov E, Morrison AC, Pervjakova N, Qu L, Rose LM, Salfati E, Saxena R, Scholz M, Smith AV, Tikkanen E, Uitterlinden A, Yang X, Zhang W, Zhao W, de Andrade M, de Vries PS, van Zuydam NR, Anand SS, Bertram L, Beutner F, Dedoussis G, Frossard P, Gauguier D, Goodall AH, Gottesman O, Haber M, Han B-G, Huang J, Jalilzadeh S, Kessler T, König IR, Lannfelt L, Lieb W, Lind L, Lindgren CM, Lokki M-L, Magnusson PK, Mallick NH, Mehra N, Meitinger T, Memon F-U-R, Morris AP, Nieminen MS, Pedersen NL, Peters A, Rallidis LS, Rasheed A, Samuel M, Shah SH, Sinisalo J, Stirrups KE, Trompet S, Wang L, Zaman KS, Ardissino D, Boerwinkle E, Borecki IB, Bottinger EP, Buring JE, Chambers JC, Collins R, Cupples LA, Danesh J, Demuth I, Elosua R, Epstein SE, Esko T, Feitosa MF, Franco OH, Franzosi MG, Granger CB, Gu D, Gudnason V, Hall AS, Hamsten A, Harris TB, Hazen SL, Hengstenberg C, Hofman A, Ingelsson E, Iribarren C, Jukema JW, Karhunen PJ, Kim B-J, Kooner JS, Kullo IJ, Lehtimäki T, Loos RJF, Melander O, Metspalu A, März W, Palmer CN, Perola M, Quertermous T, Rader DJ, Ridker PM, Ripatti S, Roberts R, Salomaa V, Sanghera DK, Schwartz SM, Seedorf U, Stewart AF, Stott DJ, Thiery J, Zalloua PA, O'Donnell CJ, Reilly MP, Assimes TL, Thompson JR, Erdmann J, Clarke R, Watkins H, Kathiresan S, McPherson R, Deloukas P, Schunkert H, Samani NJ, Farrall M. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature genetics* 2015: 47(10): 1121-1130.
17. Malik R. MEGASTROKE. 2018 [cited; Available from: megastroke.org/index.html
18. Hartwig FP, Tilling K, Smith GD, Lawlor DA, Borges MC. Bias in two-sample Mendelian randomization when using heritable covariable-adjusted summary associations. *bioRxiv* 2020: 816363.
19. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese A-K, van der Laan SW, Gretarsdottir S, Anderson CD, Chong M, Adams HHH, Ago T, Almgren P, Amouyel P, Ay H, Bartz TM, Benavente OR, Bevan S, Boncoraglio GB, Brown RD, Jr., Butterworth AS, Carrera C, Carty CL, Chasman DI, Chen W-M, Cole JW, Correa A, Cotlarciuc I, Cruchaga C, Danesh J, de Bakker PIW, DeStefano AL, den Hoed M, Duan Q, Engelter ST, Falcone GJ, Gottesman RF, Grewal RP, Gudnason V, Gustafsson S, Haessler J, Harris TB, Hassan A, Havulinna AS, Heckbert SR, Holliday EG, Howard G, Hsu F-C, Hyacinth HI, Ikram MA, Ingelsson E, Irvin MR, Jian X, Jiménez-Conde J, Johnson JA, Jukema JW, Kanai M, Keene KL, Kissela BM, Kleindorfer DO, Kooperberg C, Kubo M, Lange LA, Langefeld CD, Langenberg C, Launer LJ, Lee J-M, Lemmens R, Leys D, Lewis CM, Lin W-Y, Lindgren AG, Lorentzen E, Magnusson PK, Maguire J, Manichaikul A, McArdle PF, Meschia JF, Mitchell BD, Mosley TH, Nalls MA, Ninomiya T, O'Donnell MJ, Psaty BM, Pulit SL, Rannikmäe K, Reiner AP, Rexrode KM, Rice K, Rich SS, Ridker PM, Rost NS, Rothwell PM, Rotter JI, Rundek T, Sacco RL, Sakaue S, Sale MM, Salomaa V, Sapkota BR, Schmidt R, Schmidt CO, Schminke U, Sharma P, Slowik A, Sudlow CLM, Tanislav C, Tatlisumak T, Taylor KD, Thijs VNS, Thorleifsson G, Thorsteinsdottir U, Tiedt S, Trompet S, Tzourio C, van Duijn CM, Walters M, Wareham NJ, Wassertheil-Smoller S, Wilson JG, Wiggins KL, Yang Q, Yusuf S, Consortium AF, Cohorts for H, Aging Research in Genomic Epidemiology C, International

Genomics of Blood Pressure C, Consortium I, Starnet, Bis JC, Pastinen T, Ruusalepp A, Schadt EE, Koplev S, Björkegren JLM, Codoni V, Civelek M, Smith NL, Trégouët DA, Christophersen IE, Roselli C, Lubitz SA, Ellinor PT, Tai ES, Kooner JS, Kato N, He J, van der Harst P, Elliott P, Chambers JC, Takeuchi F, Johnson AD, BioBank Japan Cooperative Hospital G, Consortium C, Consortium E-C, Consortium EP-I, International Stroke Genetics C, Consortium M, Neurology Working Group of the CC, Network NSG, Study UKYLD, Consortium M, Sanghera DK, Melander O, Jern C, Strbian D, Fernandez-Cadenas I, Longstreth WT, Jr., Rolfs A, Hata J, Woo D, Rosand J, Pare G, Hopewell JC, Saleheen D, Stefansson K, Worrall BB, Kittner SJ, Seshadri S, Fornage M, Markus HS, Howson JMM, Kamatani Y, Debette S, Dichgans M. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature genetics* 2018: 50(4): 524-537.