GLP-1R expression protects against 58 diseases but raises risk for 34 diseases and neonatal health

Robert H. Campbell [0009-0002-0212-224X] 1 & Melinda C. Mills [0000-0003-1704-0001]1,2,3

1 Leverhulme Centre for Demographic Science, Nuffield Department of Population Health, University of Oxford

2 Department of Genetics, University Medical Centre Groningen, The Netherlands

3 Department of Economics, Econometrics and Finance, University of Groningen, The Netherlands

Abstract

Cardiometabolic gene-target drugs such as Glucagon-Like Peptide 1 Receptor (GLP-1R) agonists are used extensively, yet risks and repurposing have not been systematically evaluated across a comprehensive set of diseases. We use Phenome-wide Mendelian Randomization (MR-PheWAS) to test GLP-1R and Cholesteryl Ester Transfer Protein (CETP), two promising cardiometabolic drug targets, on risk of 396 diseases based on sex-specific, ancestry-specific, genome-wide association studies (GWAS) in UK Biobank. We identify 92 causal effects (66 novel) of genetically proxied GLP-1R expression on disease, with 58 protective and 34 risk effects. Risks of GLP-1R on neonatal health raise concern for reproductive-aged women. We find GLP-1R substantially increases risk of Vitamin D deficiency. Although GLP-1R is hypothesized as anti-ageing, we find it increases risk of 22 age-related diseases. Conversely, we find CETP inhibition is narrowly cardioprotective. Results show benefits and risks of GLP-1R and CETP drugs, and ability of MR-PheWAS to identify drug repurposing opportunities.

Keywords: Phenome-wide Mendelian randomization, GLP-1R, Glucagon-Like Peptide 1 Receptor, CETP, Cholesteryl Ester Transfer Protein, Ageing, Maternal, Neonatal

Cardiovascular diseases are the leading cause of death worldwide[1], while prevalence of metabolic diseases such as type 2 diabetes has more than doubled since 1990, accounting for 10% of the annual UK National Health Service budget[2]. Multiple gene-target drugs have been developed to counter cardiometabolic disease, creating optimism and concern about rapid adoption. The most prominent is GLP-1R agonists, a class of drugs originally approved to treat diabetes. As of January 2026, GLP-1R agonist drugs are used by as many as one in eight Americans to treat overweight or obesity[3]. Due to upcoming patent expiry, the ongoing development of oral formulations, and plans to lower prices, GLP-1R agonist use is expected to further increase[4], meriting both optimism and caution. Recent observational evidence suggests that GLP-1R agonists may have additional risks and benefits outside their original target indications[5], including possibly as anti-ageing drugs[6].

Like the former morning-sickness drug thalidomide[7], GLP-1R agonists are untested for pregnancy safety but increasingly used on reproductive-aged women. However, GLP-1 has been shown to cause birth complications and fetal death in mouse models[8,9]. This presents a public health concern for women who may become pregnant given the multi-week half-life of GLP-1R drugs[10], and the fact that approximately 50% of pregnancies are unplanned[11].

Cholesteryl ester transfer protein (CETP) inhibitors are a second class of gene-target cardiometabolic drugs promising to disrupt the current best-in-class lipid lowering therapy (PCSK-9 inhibitors) with an oral formulation. CETP inhibitors are hypothesized to reduce cardiovascular disease by modifying multiple blood lipid parameters (LDL-C, Apo-B, HDL-C) including LP(a), a highly atherogenic particle that until recently was believed to be unmodifiable through lifestyle or pharmacology[12].

The CETP inhibitor drug Obicetrapib is currently under European Medicines Agency (EMA) review for treatment of high cholesterol[13]. However, previous CETP inhibitor drug programs, such as Torcetrapib, have been terminated due to the discovery of elevated blood pressure and increased deaths in the treatment arm, raising concerns about side effects[14]. It is also unknown whether CETP inhibition could modify risk of diseases beyond high cholesterol.

Due to the genetic mechanism of action of these drug classes, causal genomic analysis offers a unique opportunity to address the limitations in existing literature from clinical trial results, animal data, and observational research. Phenome-wide drug targeting Mendelian Randomization (MR-PheWAS) is a causal statistical technique that performs a “phenome scan” for the effect of perturbations of a genetic drug target on the phenome of human disease[15]. Unlike a traditional phenome scan, the drug target exposure is estimated using meta-analysis of quantitative trait loci (QTLs) of the drug target gene, and exposure-outcome relationships are estimated using the Mendelian Randomization method, which leverages the random-like assignment of alleles to estimate causal effects[16,17].

MR-PheWAS allows for natural-experiment tests of gene activity levels using high sample sizes in a human population over the course of an entire lifetime[17,18], and can test hypotheses in vulnerable populations such as pregnant women, where assignment to a treatment group would be unethical.

As of January 19th 2026, a PubMed search shows no phenome-wide causal research of either GLP-1R or CETP (Methods, Prior literature search). We found 44 and 75 individual-phenotype Mendelian randomization studies of GLP-1R and CETP respectively, which test for the effect of these drug targets on individual “candidate diseases”[19] rather than the full phenome of disease. These findings have mostly been in the protective direction. Genetic analysis of repurposing potential using MR-PheWAS can validate and expand on existing research[5], and test for unexamined risks and repurposing opportunities, such as for neonatal diseases.

We used MR-PheWAS to investigate the causal effects of genetically proxied CETP concentration and GLP-1R expression on a broad range of diseases, with a particular focus on neonatal diseases, age-related diseases, and sex differences.

Genetic instruments for CETP and GLP-1R were derived from published genome-wide, protein-quantitative, and expression-quantitative trait loci studies, using sex-specific instruments where available (CETP) and both-sex instruments where genetic effects were

assumed to be sex-conserved[47] (GLP-1R). Outcomes were drawn from hundreds of disease phenotypes defined using Global Burden of Disease (GBD)[20] hierarchies and mapped to ICD-10 codes with genome-wide association study (GWAS) summary statistics from two pan-phenome GWAS of the UK Biobank[21,22], analyzed separately in males, females, and both sexes. The combination of these powerful data sources allows for methodologically rigorous coverage of the human disease phenome, combined with the high statistical power of population-scale genomic data applied to granular, ICD-10-specific phenotyping.

Our MR-PheWAS used standard instrument quality control[16] and MR estimators (Wald ratio for single-instrument exposures, inverse-variance weighted methods for multi-instrument exposures)[23], followed by extensive robustness checks to mitigate bias from pleiotropy, heterogeneity, and reverse causation[16]. Significant exposure–outcome associations were identified after Benjamini-Hochberg false discovery rate correction ($q < 0.05$). We included positive and negative outcome controls for both GLP-1R (type 2 diabetes (E11) and hair color[24]) and CETP (hypercholesterolemia[25] and hair color[24]) to validate our instruments, as well as positive and negative exposure controls (smoking [26–28] and handedness[21]) to test our MR-PheWAS pipeline.

To identify the risk or protective effects of GLP-1R and CETP on disease categories, we grouped our MR-PheWAS results using the GBD disease ontology and performed permutation tests on in-group versus out-group disease signed p-values. We also applied permutation testing to compare MR effects across age-related diseases (ARDs) versus non-age-related diseases[29]. Sensitivity analyses examined sex-conserved GLP-1R effects, sex-divergent GLP-1R effects, and maternal and neonatal disease subsets. The study was conducted in accordance with STROBE-MR and established MR guidelines, adapting them for a hypothesis-free, phenome-wide context[16,30].

Results

GLP-1R – MR-PheWAS

Our MR-PheWAS analyses (Figure 1) of GLP-1R expression show phenome-wide significant effects across all categories of diseases, both protective and with risks, for women, men, and both sexes. While we observe differences in the proportions of risk and protective effects by disease category, permutation testing for category-wide effects was null after multiple testing correction.

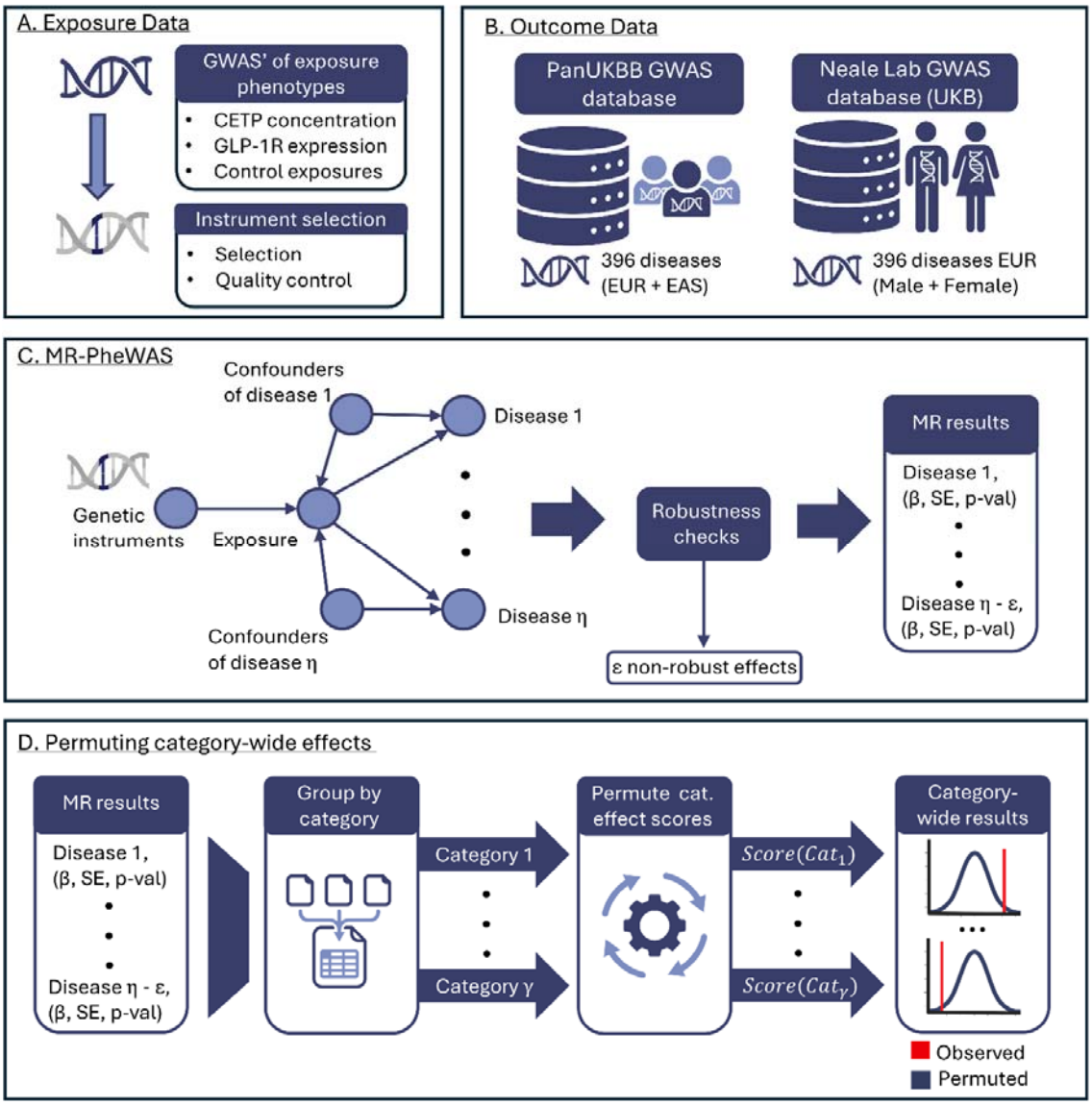


Figure 1, Overview of model design. Study design and analysis workflow for MR-PheWAS. **(A)** Exposure data comprised quantitative trait loci of CETP concentration and GLP-1R gene expression filtered for suitability in MR-PheWAS using standard thresholds. **(B)** Outcome data comprised the Genome-Wide Association Study summary statistics for 396 traits at the selected genomic loci from (A) made available through the PanUKBB and Neale Lab UKB GWAS Round 2 projects. **(C)** MR-PheWAS main model. The causal effect of the exposure from (A) is estimated on each outcome selected from (B) using two-sample MR. Robustness checks and multiple testing correction filter out non-robust results. **(D)** The estimated causal effects from (C) are grouped by disease category and permutation tested for category-wide effects of the exposure.

For the MR-PheWAS of GLP-1R, we used a shared set of single nucleotide polymorphisms (SNPs) available through eQTLGen as exposure instruments for GLP-1R gene expression[31,32] in females, males and both sexes (Table 1, Supplementary Table 1). 22 of 134 candidate GLP-1R SNPs passed exposure quality control and were used as exposure instruments for MR-PheWAS (Supplementary Table 2).

Table 1. MR-PheWAS summary results for CETP, GLP-1R and control exposures on sex-specific, ancestry-specific populations.

Exposure	Ancestry	Sex	Pleiotropy Screening Regex	Exposure Units	Instr. p-value threshold	Instr. count	Diseases passing robustness check (ARD/non-ARD)	Significant disease effects (risk/protection)	Category-wide effects
GLP1R	EUR	Both	None	SD GLP1R gene expression	5x10-8	22	148 (89/59)	92 (34/58)	FALSE
	EUR	Female	None	SD GLP1R gene expression	5x10-8	22	345 (170/175)	213 (101/112)	FALSE
	EUR	Male	None	SD GLP1R gene expression	5x10-8	22	298 (207/92)	202 (88/114)	FALSE
CETP	EUR	Both	None	SD CETP protein concentration	5x10-8	3	347 (197/150)	0	Cardiovascular disease, level two
	EUR	Female	None	SD CETP gene expression	5x10-2* **	1	345 (170/175)	0	NA
	EUR	Male	None	SD CETP gene expression	5x10-4*	1	304 (210/94)	0	NA
Left-handed	EUR	Both	None	log-OR left handed	5x10-8	2	0	NA	NA
Ambidextrous	EUR	Both	None	log-OR ambidextrous	5x10-8	1	118 (68/50)	0	NA
Smoking initiation	EUR	Both	none	log-OR ever smoker	5x10-8	92	300 (180/120)	50 (50/0)	FALSE
	EUR	Both	“educat”, “school”, "BMI", "Obes"	log-OR ever smoker	5x10-8	1	152 (91/61)	0	NA
	EAS	Both	None	log-OR ever smoker	5x10-8	4	20 (12/8)	0	FALSE
	EAS	Male	“educat”, “school”, "BMI", "Obes"	log-OR ever smoker	5x10-8	0	NA	NA	NA
	EAS	Female	“educat”, “school”, "BMI", "Obes"	log-OR ever smoker	5x10-8	0	NA	NA	NA

* Benjamini-Hocherg false discovery rate adjusted significant at $q < 0.05$
** F statistic is lower than 10.

Of the 321 diseases in the both-sex phenome with available GWAS data included in the both-sex MR-PheWAS of GLP-1R, 148 (89 ARD and 59 non-ARD) passed our robustness checks (Figure 2, Supplementary Table 3). Of the 148 remaining results shown in Figure 2, GLP-1R had a phenome-wide significant effect on 92 diseases defined by ICD-10 codes ($q < 0.05$), with a protective effect on 58 and a risk effect on 34.

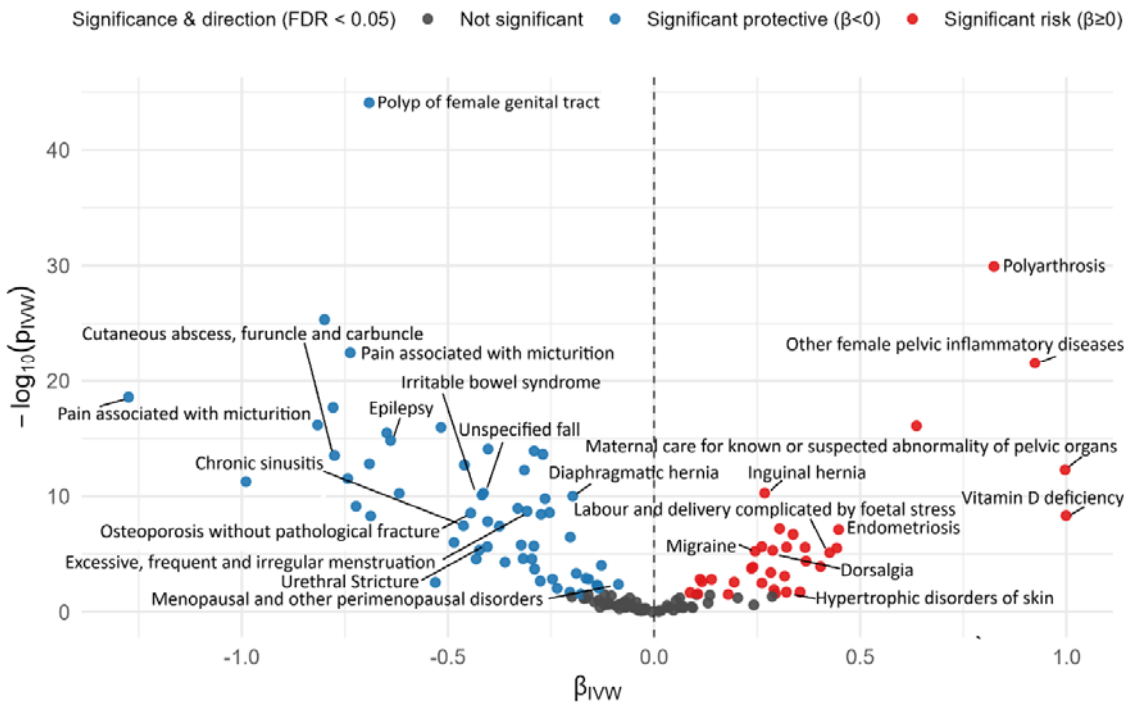


Figure 2, Volcano plot results of phenome-wide Mendelian randomization of GLP-1R expression. "MR" refers to Mendelian Randomization testing. "GLP-1R" refers to glucagon-like peptide 1 receptor expression. "FDR" refers to Benjamini-Hochberg false discovery rate correction. "βivw" refers to the causal effect of one standard deviation increase of GLP-1R expression on log-odds ratio of disease diagnosis as estimated using the inverse variance weighted method. The dotted line refers to a log-odds ratio of zero, implying no effect change in disease risk resulting from increased GLP-1R expression. Disease labels are applied to selected novel Mendelian randomization findings with selective labelling. Red dots imply increased GLP-1R expression raises risk for the named disease outcome, while blue dots imply increased GLP-1R expression lowers risk. Statistical significance is defined as q < 0.05 after Benjamini-Hochberg false discovery rate adjustment.

The disease with the largest protective effect size in the both-sex GLP-1R MR-PheWAS was "Pain associated with micturition" (OR 0.28), followed by Cardiac arrest (OR 0.37). The 10 largest protective effect size diseases included common healthspan-ending diseases, such as cardiovascular diseases (I46, I49), cancers (C79) and neurodegenerative conditions (G81).

The disease with the largest risk effect size in the both-sex GLP-1R MR-PheWAS was Vitamin D deficiency, with an odds ratio of 2.72 per standard deviation increase in GLP-1R expression, followed by "Maternal care for known or suspected abnormality of pelvic organs" (OR ≈ 2.72), "Other female pelvic inflammatory diseases" (OR 2.51), and "Polyarthrosis" (OR 2.28). Four of the 10 diseases with the highest risk were related to the uterus or female reproduction; others included "Endometriosis" (OR 1.55) and "Labor and delivery complicated by fetal stress" (OR 1.55). We note that the naïve odds ratios for female-specific diseases are likely to underestimate the true risk effect, as they are estimated over a sample population containing approximately 46% men, who are not in the risk set. We also note that these effect sizes reflect the effect of natural variation of GLP-1R expression levels among people, not the considerably larger effect of GLP-1R drug use (see Discussion).

We conducted literature searches for existing MR studies of GLP-1R to identify whether our 92 significant results for the both-sex population contained novel findings. We performed a primary search using PubMed for each of our significant results, checking against Perplexity[33] if the PubMed search returned no relevant results (Methods, Literature Search for Novel Findings). We found no prior MR studies for 66 of our 92 significant findings. Of these 66 novel MR findings, 41 were protective findings, where GLP-1R reduced risk of an ICD-10-coded disease, and 25 were risk findings, where GLP-1R increased risk of an ICD-10-coded disease (Supplementary Table 4).

Results from the MR-PheWAS of GLP-1R on the both-sex population grouped by Global Burden of Disease level-2 categories are shown in Extended Data Figure 1. The most protective-skewed category was the expected “Diabetes and kidney diseases,” and the most risk-skewed category was “Musculoskeletal disorders.” Our sensitivity check of “sex-conserved” findings that treated the female population as discovery and male population as validation returned a set of 98 phenome-wide significant effects of GLP-1R (Supplementary Table 5). Our sensitivity check of “sex-divergent” findings returned a set of 56 statistically significant sign disagreements between the effects of GLP-1R on males and females (Supplementary Table 6).

GLP-1R – MR-PheWAS – Sex-specific

Of the 345 ICD-10 codes in the female phenome with available GWAS surviving robustness checks (170 ARD, 175 non-ARD), GLP-1R expression had a statistically significant effect on 213 (101 risk; 112 protective) after false discovery rate correction ($q < 0.05$) (Supplementary Table 7, Supplementary Figure 1). Of the 298 ICD-10 codes with available GWAS in the male phenome surviving robustness checks (207 ARD, 92 non-ARD), GLP-1R expression had a statistically significant effect on 202 diseases (88 risk; 114 protective) (Supplementary Table 8, Supplementary Figure 2). These results show significant effects of GLP-1R on 64% and 67% of the phenotypes tested respectively.

To investigate the effects of GLP-1R on pregnancy, we examined the proportions of risky, protective and null effects of GLP-1R on the categories of “neonatal diseases,” which describe pregnancy risks originating from the neonate, and “maternal diseases,” which describe pregnancy risks originating from the mother (see Methods, Maternal and neonatal diseases, and Supplementary Table 9). Table 2 shows the estimated effects of GLP-1R on individual disease effects in both categories as measured in the female population. We found “neonatal diseases” was the most risk-affected category of the 14 level-2 disease categories examined (Extended Data Figure 2). For neonatal diseases, GLP-1R had a 25% protective and 75% risky effect among the four phenome-wide, significantly affected diseases within-category. The three significant risk effects on neonatal diseases were “O69 Labor and delivery complicated by umbilical cord complications,” “O00 Ectopic pregnancy,” and “O68 Labor and delivery complicated by fetal stress” (Table 2). O86 previously appeared as a risk effect in the both-sex MR-PheWAS of GLP-1R. The single significant protective effect on the neonatal phenome was “O02 Other abnormal products of conception”. Conversely, GLP-1R had a 63% protective and 37% risky effect on the 11/17 significantly affected maternal diseases.

Table 2. Estimated directional risk and protective effects of increased GLP-1R expression on maternal and neonatal diseases in female European population.

Maternal Diseases	Naïve p-value	BH-adjusted q-value	Robustness checks (passed / attempted)	Phenome-wide significant	Classification
O14 Gestational [pregnancy-induced] hypertension with significant proteinuria	3.67E-06	1.19E-05	(7/7)	TRUE	Significant risk
O16 Unspecified maternal hypertension	0.015	0.03	(7/7)	TRUE	Significant risk
O69 Labor and delivery complicated by umbilical cord complications	0.025	0.04	(7/7)	TRUE	Significant risk
O72 Postpartum hemorrhage	2.02E-05	5.78E-05	(7/7)	TRUE	Significant risk
O20 Hemorrhage in early pregnancy	2.98E-06	9.88E-06	(7/7)	TRUE	Significant protective
O21 Excessive vomiting in pregnancy	7.38E-04	1.50E-03	(7/7)	TRUE	Significant protective
O48 Prolonged pregnancy	2.87E-03	5.40E-03	(7/7)	TRUE	Significant protective
O75 Other complications of labor and delivery, not elsewhere classified	1.44E-20	4.03E-19	(7/7)	TRUE	Significant protective
O42 Premature rupture of membranes	8.05E-06	2.41E-05	(7/7)	TRUE	Significant protective
O46 Antepartum hemorrhage, not elsewhere classified	8.68E-05	2.15E-04	(7/7)	TRUE	Significant protective
O64 Obstructed labor due to malposition and malpresentation of foetus	0.031	0.05	(7/7)	TRUE	Significant protective
O03 Spontaneous abortion	0.046	0.07	(7/7)	FALSE	Not significant
O13 Gestational [pregnancy-induced] hypertension without significant proteinuria	0.634	0.68	(7/7)	FALSE	Not significant
O44 Placenta praevia	0.265	0.33	(7/7)	FALSE	Not significant
O62 Abnormalities of forces of labor	0.433	0.49	(7/7)	FALSE	Not significant
O63 Long labor	0.118	0.16	(7/7)	FALSE	Not significant
O70 Perineal laceration during delivery	0.145	0.19	(7/7)	FALSE	Not significant

Neonatal Diseases	Naïve p-value	BH-adjusted q-value	Robustness checks (passed / attemped)	Phenome-wide significant	Classification
O68 Labor and delivery complicated by fetal stress [distress]	0.022	0.04	(7/7)	TRUE	Significant risk
O69 Labor and delivery complicated by umbilical cord complications	0.025	0.04	(7/7)	TRUE	Significant risk
O00 Ectopic pregnancy	7.07E-09	3.29E-08	(7/7)	TRUE	Significant risk
O02 Other abnormal products of conception	2.35E-09	1.14E-08	(7/7)	TRUE	Significant protective
O03 Spontaneous abortion	0.046	0.07	(7/7)	FALSE	Not significant

GLP-1R Age-Related Disease

GLP-1R had at least one significant risk and protective effect in every level-2 GBD disease category containing age-related diseases in the both-sex, female, and male MR-PheWAS

(Supplementary Figures 3-5) and corresponding sex-specific ageing phenomes (Supplementary Tables 10-12). Figure 3 shows phenome-wide significant risk (22) and protective (32) effects for 54 of the 89 diseases (60%) in the both-sex age-related disease phenome. GLP-1R showed a concentrated protective effect on diabetic and kidney diseases within the ageing phenome, where it significantly protected against 5 of 6 diseases in this category (Figure 3). Conversely, GLP-1R showed a mixed-to-risky effect on musculoskeletal disorders in the ageing phenome, where it had a significant risk effect on 6 diseases and a protective effect on 4. Permutation tests for an ageing-phenome-wide effect of GLP-1R were null for women, men and both sexes (Supplementary Figures 6-8).

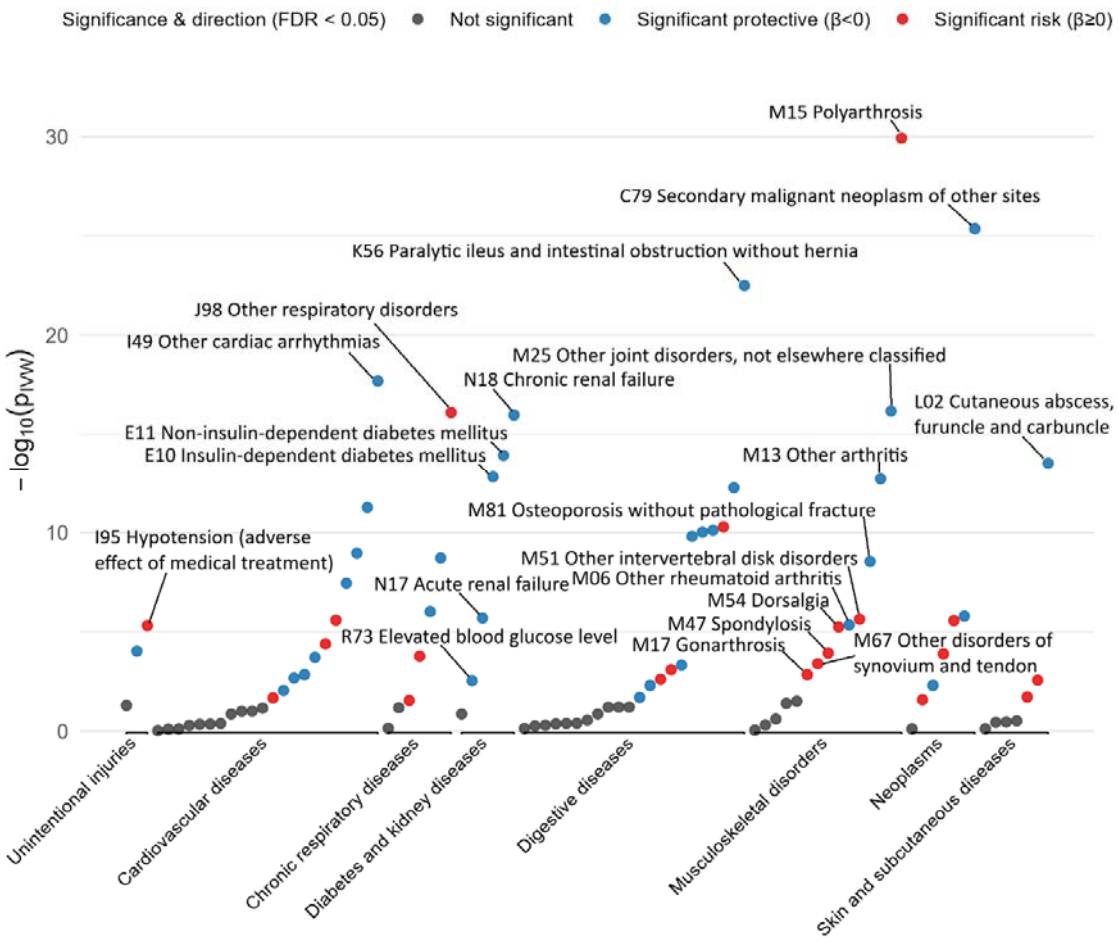


Figure 3, Manhattan plot of GLP-1R effects on the phenome of age-related diseases in the both-sex population. Diseases are grouped by Global Burden of Disease category (level 2). Disease labels are applied to the lowest p-value effect in each category, and to significantly affected diseases in the most risk-affected and protective-affected categories. Red dots imply increased GLP-1R expression raises risk for the named disease outcome, while blue dots imply increased GLP-1R expression lowers risk. Statistical significance is defined as q < 0.05 after Benjamini Hochberg false discovery rate adjustment. -log10(Pivw) represents the log-scaled p-value measuring the statistical significance of the measured effect.

GLP-1R – Disease Category Effect

To test for the statistical significance of the observed disease-group differences, we conducted permutation testing for a category-wide risk or protective effect of GLP-1R on all disease categories from the full GBD disease ontology using the top three disease hierarchy levels[20], including the three additional categories “neonatal diseases,” “maternal diseases,” and “age-related diseases.” After accounting for multiple testing burden, we found no category-wide significant results at any level for both-sex (Supplementary Figures 9-11), female (Supplementary Figures 12-14) and male (Supplementary Figures 15-17) analyses ($q < 0.05$). These permutation tests were adjusted for multiple testing burden ($q < 0.05$) at the disease hierarchy level (1,2,3). Relaxing multiple testing correction for the both-sex analyses results in significant group-wide protection of GLP-1R on “neoplasms” (level two, $p = 0.031$), and “other neoplasms” (level three, $p = 0.026$).

CETP – MR-PheWAS

The results for the MR-PheWAS of CETP protein concentration show a significant protective effect of CETP inhibition on cardiovascular disease at both the individual disease and disease-group level, with no other significant effects (Table 1, Supplementary Tables 1,15).

We used exposure SNPs from a GWAS of CETP concentration in our both-sex analysis, and a single sex-specific CETP concentration SNP for the sex-specific analyses[34,35] (Supplementary Tables 13-14). For the both-sex analysis shown in Figure 4, the MR-PheWAS pipeline fit IVW results for the 396 diseases with available GWAS data (224 ARD; 172 non-ARD) in UK Biobank, of which 347 passed robustness checks (197 ARD; 150 non-ARD) (Supplementary Table 15). Of these, we identified a phenome-wide significant risk effect of increased CETP concentration on Angina Pectoris (I20), an acute ischemia of the heart resulting in chest pain, typically caused by coronary artery disease.

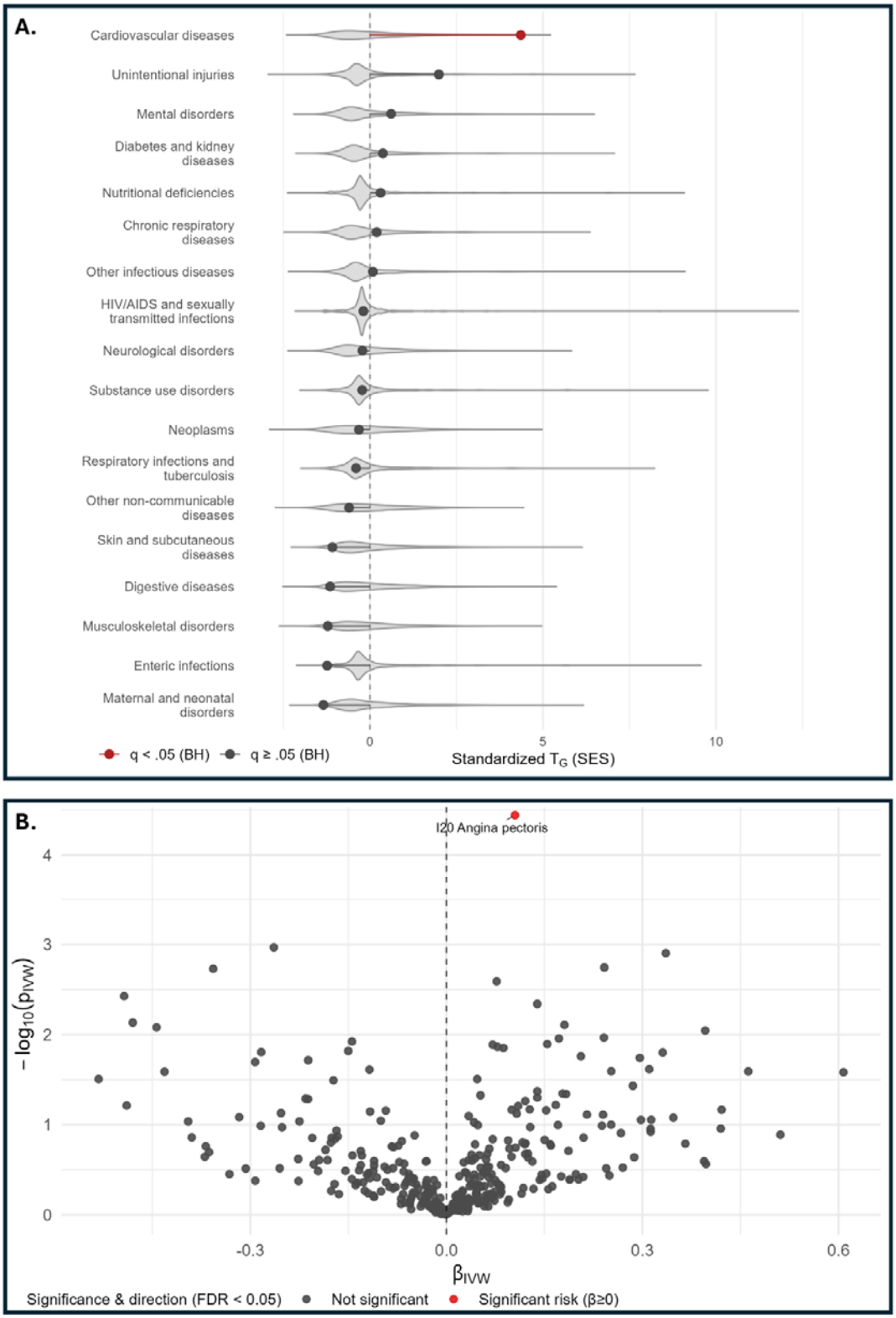


Figure 4. Summary effects of increased CETP concentration. (A) Permutation testing for category-wide effects of CETP concentration on the European both-sex population. Categories are drawn from level 2 of the Global Burden of Disease clinical ontology. Red dots imply increased CETP concentration raises risk for the diseases in the named disease category. Statistical significance is defined as q < 0.05 after Benjamini-Hochberg false discovery rate adjustment over the 18 categories at level 2. **(B)** Volcano plot results of phenome-wide Mendelian randomization of CETP concentration on the both-sex European population. “MR” refers to Mendelian

Randomization testing. “CETP” refers to cholesteryl ester transfer protein. “FDR” refers to Benjamini-Hochberg false discovery rate correction. “βivw" refers to the causal effect of one standard deviation increase of CETP concentration in whole blood on the log-odds ratio of disease diagnosis as estimated using the inverse variance weighted method. The dotted line refers to a log-odds ratio of zero, implying no effect change in disease risk resulting from increased CETP concentration.

Because the sex-specific MR-PheWAS of CETP used only a single exposure instrument, we used the Wald ratio as our primary MR estimator and disabled the robustness analysis pipeline, which is undefined on single-instrument analyses. For the female analysis, the MR-PheWAS pipeline fit Wald ratio results for the 345 diseases (170 ARD, 175 non-ARD) with available GWAS data (Supplementary Figure 18, Supplementary Table 16), while the male analysis fit Wald ratio results for 304 diseases (210 ARD, 94 non-ARD) with available GWAS data (Supplementary Figure 19, Supplementary Table 17). No results were phenome-wide significant.

CETP – Tests for Group Effects

Permutation tests for disease-group effects in the both-sex analysis show a statistically significant (Benjamini-Hochberg $q < 0.05$) risk effect of increased CETP protein concentration on cardiovascular diseases (Figure 4). This implies CETP inhibition is protective of cardiovascular diseases in general which is also the intended indication for CETP inhibitor drugs. This finding for cardiovascular disease occurred at level two of the GBD disease hierarchy and is Benjamini-Hochberg significant ($q < 0.05$) against a comparison set of 18 disease groups. Permutation tests for level one and level three disease category effects as well as ageing phenome effects were all null (Supplementary Figures 20-22).

Control Phenotypes – MR-PheWAS

We tested positive and negative outcome controls for GLP-1R and CETP in the both-sex European population. A negative control should have no causal relationship with the exposure (GLP-1R or CETP), and a positive control should have a causal relationship in the expected direction. We found that increased GLP-1R expression had the expected significant protective effect on type 2 diabetes (E11) ($p = 6.37 \times 10^{-10}$) (Supplementary Table 3) and expected null effect on brown hair color ($p = 0.31$) (Supplementary Table 18). We found increased CETP concentration, the opposite effect of a CETP inhibitor drug, had the expected significant risk effect on high cholesterol ($p = 6.25 \times 10^{-3}$) (Supplementary Table 19) and null effect on brown hair color ($p = 1.00$) (Supplementary Table 20).

We also tested positive and negative exposure controls (smoking initiation and handedness) to confirm our MR-PheWAS pipeline produced expected phenome-wide results of known-risky and known-harmless exposures. Our negative control left-handedness returned expected null results for the both-sex European population after all exposure-outcome effects failed to pass robustness checks. We tested a second negative control, ambidextrousness, which returned expected null results in the European both-sex population, with no effect reaching phenome-wide significance (Supplementary Table 21, Supplementary Figure 23). Our positive control smoking initiation had the expected phenome-wide significant risk effect on lung cancer (C34) ($q = 9.89 \times 10^{-5}$) and chronic obstructive pulmonary disease (J44) ($q = 4.49 \times 10^{-19}$), with phenome-wide risk effects on a

total of 50 diseases and protective effects on zero diseases in the European both-sex population (Supplementary Table 22, Supplementary Figures 24-25). Permutation tests for disease category effects did not reach statistical significance after multiple testing correction (Supplementary Figures 26-29). Results from MR-PheWAS of smoking initiation of East Asian both-sex and sex-specific analyses were either partial fits or non-fits due to data limitations, and uniformly without significant findings (Supplementary Figure 30, Supplementary Table 23). Additional detail on exposure control results is available in Supplementary Note 1.

Discussion

We found significant causal relationships between GLP-1R and ICD-10 coded diseases ranging across the human disease phenome for women, men, and both sexes. We identified 92 total effects and 66 novel effects of GLP-1R on the both-sex phenome, underscoring the power of MR-PheWAS to identify opportunities and risks for drug repurposing.

Of particular interest was the largest risk effect of GLP-1R on vitamin D deficiency (OR 2.72 per SD GLP-1R expression), not found in the existing Mendelian randomization literature. Curiously, GLP-1R levels appear to be positively correlated with vitamin D in human observational studies[36], while intervention studies of GLP-1R show risks of vitamin deficiencies[37]. Our findings align with intervention studies and suggest that GLP-1R agonist drugs may raise risk of vitamin D deficiency. This finding is particularly relevant to populations vulnerable to vitamin D deficiency, such as women and people with a migration background living in the low-sun climates of Europe and North America[38].

Secondly, we identify a risk effect of GLP-1R on neonatal health. Our results show mixed effects of GLP-1R on maternal diseases (37% risk effects), while we find a strong risk effect on neonatal diseases (75% risk effects). Of the two maternal disorders (O72 and O70) that GLP-1R protects against with established clinically meaningful effect sizes (OR > 1.1), both are obstetric traumas where neonate birth weight is a risk factor. For postpartum hemorrhage (O72), a meta-analysis of 800 million women places high birth weight (>4kg) in the largest effect size category of risk factors[39], and a second meta-analysis finds a monotonic relationship between birth weight and PPH risk with an odds ratio of 2.05 for birth weights above 4kg[40]. For perineal laceration during delivery (O70), the same meta-analysis finds an odds ratio of 1.90 for birth weights above 4kg. We suggest our findings are compatible with GLP-1R paradoxically protecting against obstetric trauma by inhibiting fetal growth, resulting in lower birth weights and reduced trauma during delivery. A similar effect is observed for cigarette smoking, which has been found to reduce obstetric trauma via inhibiting fetal growth[41].

Mouse studies directly support the hypothesis that GLP-1 activity inhibits fetal growth, finding GLP-1-exposed pregnancies have reduced birth weights[8,9] and increased risk of fetal death[9]. Emerging observational evidence in humans is mixed, with some studies finding a risk effect of GLP-1R agonists on maternal and birth complications[42], while others find no effect[43].

If GLP-1R agonism does present a risk to neonates, the anticipated increase in wider population usage and long washout period of current GLP-1R agonist drugs may present a significant public health concern. In the United States, the median time to become aware of a pregnancy is 5.5 weeks, with 23% of pregnancies being recognized at or after 7 weeks[44].

The common GLP-1R drug semaglutide (also known as Ozempic and Wegovy) has an effective clearance time of 5-6 weeks[10]. This suggests that GLP-1R-exposed pregnancies may have completed most or all of the first trimester under GLP-1R-exposure, even if drug use is discontinued immediately after the pregnancy is discovered. Our findings support the United States Food and Drug Administration current recommendation that pregnant women should not use GLP-1R agonists[45], and raise a further concern about GLP-1R agonist use among women who may unknowingly be pregnant.

Emerging evidence suggests that GLP-1R may function as an anti-ageing drug[6]. Despite the sizeable number of risk and protective effects of GLP-1R identified in this study, we do not find evidence of an ageing-phenome specific effect of GLP-1R for women, men and both sexes. This null finding for the effect of GLP-1R on age-related disease despite many significant risk and protective findings may suggest that GLP-1R agonism is a double-edged sword, protecting against some age-related diseases while increasing the risk for others. Figure 3 shows that GLP-1R has an expected within-ARD protective effect on diabetes and kidney diseases (5 protective, 0 risk), and a mixed-to-risky effect on musculoskeletal disorders (4 protective, 5 risk), with the risk concentrated on age-related bone degeneration such as Polyarthrosis (M15), Gonarthrosis (M17), and Spondylosis (M47).

Our findings show a stark contrast between the two molecular phenotypes of GLP-1R expression and CETP concentration. While GLP-1R expression affects most diseases we tested in the human phenome, CETP has only the narrow and anticipated effect on cardiovascular disease. This suggests that unlike GLP-1R agonists, CETP inhibitors do not show potential for drug repurposing, nor high risk for side effects assuming perfectly targeted CETP inhibition. This is relevant to the ongoing research and development of CETP inhibitor drugs[12,13], and supports the hypothesis that the hypertension risk observed in the first generation CETP inhibitor torcetrapib was due to an off-target effect rather than an unavoidable side-effect of CETP inhibition. We suggest this disparity between CETP and GLP-1R underscores the ability of phenome-wide Mendelian randomization to distinguish between drug targets with high and low potential for repurposing.

Our study was limited by both data and methodological constraints, which may affect the interpretation of the outcomes. Our exposure of GLP-1R gene expression levels is a related but distinct measure from the effect of GLP-1R agonist drugs. To address this concern, we used diabetes mellitus diagnosis (E11) as a positive control to check that GLP-1R gene-expression matched the known causal effect of GLP-1R agonism (Supplementary Note 1). This difference of exposure also complicates interpretation of log OR effect sizes. While our effect sizes are measured in standard deviations of circulating GLP-1R transcripts among the untreated population, the dosage of GLP-1R agonist drugs is typically measured in circulating (unbound) levels of the drug, a distinct measure[46]. We note that these levels are commonly 20 to 30 times greater in magnitude than endogenous levels, suggesting the change in GLP-1 activity driven by drug exposure may be significantly larger than natural within-population variation[46].

Our instruments for the sex-specific GLP-1R sensitivity analyses come from a both-sex GWAS[32] which may affect results if these instruments are uncalibrated in single-sex populations[47] (Supplementary Note 2).

Due to the “absolute risk” scale used to define ICD-10 outcome phenotypes in the Neale Lab round 2 GWAS, it was not possible to compare risk effect sizes across disease in our sex-specific analyses. This limits the interpretation of sex-specific analyses.

Our outcome data comes from published summary statistics of the UK Biobank, without external cohort validation. To address this limitation, we completed sex-conserved validations where possible by treating the female UKB cohort as discovery and the male UKB cohort as a validation set (Supplementary Note 3). This leaves the sex-specific findings, such as for maternal health, unvalidated.

We relaxed the p-value and quality control thresholds for instrument selection in the sex-specific analysis of CETP. As the sex-specific analyses of CETP did not produce any significant results, this preprocessing choice may not affect interpretation of the significant findings (Supplementary Note 4).

We identified a need for larger-than-typical statistical power for exposure GWAS, seemingly driven by the demanding nature of our quality control and robustness checks (Supplementary Note 5)**.** Of 8 runs we completed with fewer than five exposure SNPs, only one contained significant results, while the other seven were either null (1), partial failures to fit (3) or total failures to fit (3).

We used permutation testing to detect disease-group-wide effects of exposures, but found the approach less sensitive than intended. This was compounded by multiple testing adjustment; of three nominally significant disease-group-wide effects of smoking initiation, none remained significant after multiple testing correction. We attribute this to limitations of permutation testing, which compares within-group effects to a null of outside-group effects, rather than no effect. This comparison tends to bias results toward the null, particularly when an exposure (such as GLP-1R) has effects broadly distributed across the phenome rather than concentrated within a single disease group (as with CETP). We chose permutation testing due to data constraints (Supplementary Note 6).

Data Availability

The data used in this study are openly available from the Global Burden of Disease survey[20], eQTLgen[31,32] for GLP-1R expression, a GWAS of Netherlands Epidemiology of Obesity cohort for both-sex CETP concentration[34] and sex-specific CETP concentration from the Global Lipids Genetics Consortium multi-ancestry meta-analysis[35,48]. Exposure data for smoking comes from the Tobacco and Genetics (TAG) consortium[28] for European ancestry, and from BioBank Japan [26,27] for East Asian ancestry, both accessed via OpenGWAS catalog[49]. The phenome-wide GWAS summary statistic data come from the Neale Lab UKBB GWAS round 2[21] and the PanUKBB project for 7,228 phenotypes across 6 continental ancestry groups[22]. Data for ambidextrousness and left-handedness additionally come from Neale Lab UKBB GWAS round 2[21]. Both PanUKBB and Neale Lab GWAS round 2 are based on data from the UK Biobank[50].

Code Availability

The scripts used to perform all coded analyses and generate all figures in this analysis are contained in the supplement and available on a GitHub repo[51]. Main analysis including all MR-PheWAS and permutation tests were conducted with our R package "ardmr," a tool for conducting easy testing of exposures on the full human disease phenome, as well as the sex-specific age-related disease phenomes using the pipeline described in Methods. Full examples and documentation are available in the readme file of ardmr hosted on the package GitHub repo[52].

Online Methods

1. Overview of study design

We fit sex-specific (female, male and both sexes) phenome-wide Mendelian randomization analyses on CETP concentration and GLP-1R expression levels, then tested for statistically significant effects of these exposures on disease groups and the phenome of age-related diseases (Figure 1). We used genetic instruments from sex-specific Genome-Wide Association Studies for all exposures excluding GLP-1R, where we used both-sex exposure instruments under the assumption that genetic mutations affecting GLP-1R expression levels are not meaningfully sex specific[47]. The male-specific and female-specific outcome GWAS summary statistics we used are reported as absolute risk differences rather than log-odds ratios. This prevents comparison of MR effect sizes across diseases without referencing the baseline rate of disease; we therefore report p-values of male-specific and female-specific findings but not effect sizes. For both-sex analyses we report both p-values and effect sizes.

We used two types of positive and negative control phenotypes to separately check our exposure instruments and our MR-PheWAS pipeline. We used positive and negative outcome controls for GLP-1R and CETP to confirm our instruments for these exposures capture the known effects and non-effects of GLP-1R expression and CETP concentration. In addition, we used positive and negative exposure controls to confirm that our MR-PheWAS pipeline captured the expected effects of known risky exposures (smoking) and non-risky exposures (handedness) on the disease phenome (See Methods, Controls).

As a sensitivity check on the analysis of GLP-1R on the both-sex population, we generated a "sex conserved" list of significant GLP-1R disease effects by treating the female MR-PheWAS of GLP-1R as a discovery set, and validating significant findings using the results of the male MR-PheWAS with a Benjamini-Hochberg FDR q-value threshold of $q < 0.05$.

We applied permutation testing to test for significant differences of the beta coefficients generated by Mendelian randomization of age-related diseases and non-age-related diseases. Additionally, we applied permutation testing to Mendelian randomization beta coefficients grouped by GBD disease categories against outside-category beta coefficients to test for differential effects of the exposures on specific disease categories.

This study was designed with reference to the STROBE-MR guidelines for reporting Mendelian randomization analyses[30] and Burgess et al.'s "Guidelines for performing

Mendelian Randomization investigations,"[16] with deviations adapted to accommodate phenome-wide Mendelian randomization. This study was not preregistered.

2. Data sources

Phenotypic hierarchy data

The primary disease ontology we used in this paper comes from the Global Burden of Disease survey[20] (GBD). GBD is a cross-country observational study that aggregates population-level statistics from 204 countries on diseases causing death or disability. Outcomes resulting in death or disability measured by GBD are described as "causes" and are contained in a cause hierarchy of increasing specificity[20]. The sex-specific ageing phenomes, containing lists of globally conserved age-related diseases, come from a forthcoming paper developing sex-specific ageing phenomes from GBD disease incidence data[29].

Exposure data sources

The genomic data we used to proxy exposure to GLP-1R expression and CETP concentration came from published GWASs, pQTL studies, and expression quantitative trait loci (eQTL) studies. Each study design contains associations between individual genetic mutations and observed clinical and molecular phenotypes.

The exposure data we used for GLP-1R expression level come from eQTLgen, a consortium of 37 blood-based expression quantitative trait loci cohorts not overlapping with the UK Biobank, and have previously been used in Mendelian randomization studies of GLP-1R expression [31,32]. The exposure data we used for both-sex CETP come from a GWAS of CETP concentration in whole blood[34]. The exposure data we used for sex-specific CETP concentration come from a sex-specific analysis of CETP using data from the Global Lipids Genetics Consortium multi-ancestry meta-analysis[35,48].

The exposure data we used for smoking among Europeans came from the Tobacco and Genetics (TAG) consortium[28] and were accessed via the OpenGWAS catalog[49]. TAG contains mostly European participants. Smoking data in the TAG consortium is self-reported[28].

The exposure data we used for East Asian smoking comes from BioBank Japan, a single-cohort registry of participants with case status for one or more of 51 predetermined diseases [26,27]. The cohort is overwhelmingly of Japanese ancestry. These data were accessed via OpenGWAS catalog[49].

The exposure data we used for left-handedness and right-handedness comes from the UK Biobank, and was published by Neale Lab as part of their UKB Round 2 GWAS project[21]. These data were calculated using only European ancestry UKB participants.

Outcome data sources

We used previously published genetic data for our phenome-wide outcomes from two sources of genome-wide association study summary statistics, both of which are based on GWAS in the UK Biobank (UKB)[50]. The UKB is a prospective cohort study containing around 500,000 participants from the UK population, including genomic data, multi-omic data,

clinical data drawn from National Health Service (NHS) records and a participant questionnaire.

We used previously published summary statistics for sex-specific GWAS in UKB from the Neale Lab UKBB project round 2, an open-access project that calculated and hosts sex-specific GWAS summary statistics for 4,236 phenotypes in the UK biobank[21]. We accessed the manifest file describing the scope and structure of the available GWAS summary statistics using the Neale Lab website, and accessed the genomic variant manifest, sex-specific manifests, and sex-specific GWAS summary statistics from the Neale host server as part of an automated Mendelian randomization analysis pipeline[21]. All summary statistics downloaded from Neale Lab UKBB round 2 were calculated using the genomic reference build GRCh37 (hg19) and European-ancestry biobank participants only.

We used previously published both-sex summary statistics for ancestry-specific GWAS in UKB from the PanUKBB project, an open-access project associated with the Neale Lab that hosts GWAS summary statistics for 7,228 phenotypes across 6 continental ancestry groups[22]. We downloaded a spreadsheet file containing a summary of available genomic data from the PanUKBB website[22], and downloaded genomic files including variant manifests, GWAS summary statistic Tabix files, and selections of GWAS summary statistic files from the PanUKBB host server multiple times as part of an automated Mendelian randomization pipeline. All summary statistics downloaded from PanUKBB were calculated using the genomic reference build GRCh37 (hg19) and used combined male and female biobank participants.

We accessed summary statistics for our outcome control phenotypes from the IEUGWAS catalog[49] except where these controls were already tested inside the MR-PheWAS pipeline. The outcome data we used for the CETP positive outcome control high cholesterol diagnosis comes from a GWAS of age-related diseases in the predominantly European-ancestry UKB biobank[25] using a multi-ancestry framework. The outcome data we used for the GLP-1R positive control type 2 diabetes (E11) was measured in PanUKBB and was tested automatically through the MR-PheWAS pipeline[22]. The outcome data for our GLP-1R and CETP negative outcome control, naturally dark brown hair, comes from a GWAS in the UK Biobank[24]. The phenotype for brown hair color was defined as self-reported “dark brown hair color (natural, before greying).”

3. Mendelian Randomization PheWAS

Overview of Mendelian randomization assumptions

Mendelian randomization is a statistical method for estimating the causal effect of an exposure on an outcome using genomic instrumental variables[17]. MR leverages the recombination of the two parental genomes during conception as a source of random variation, and uses instrumental variable methods to estimate the effect of these randomly assigned alleles[17]. Mendelian randomization critically depends on three main assumptions to be valid[17].

1. Relevance: all genetic instruments used must impact the exposure through some causal pathway. This assumption is typically satisfied by selecting instruments using the results of GWAS on the exposure of interest. This method identifies instruments which are significantly associated with the MR exposure phenotype.

2. Independence: all instruments must be independent of confounders. This criterion is credibly met for within-family MR designs, where genetic confounding is automatically adjusted for. However, most implementations of MR, including ours, use “two sample MR,” which creates a risk of non-random assignment of genetic instruments under assortative mating or population stratification[17]. However, non-family-based MR studies typically reduce the risk of confounding from the standard set by traditional epidemiological studies.
3. Exclusion restriction: all instruments must impact the outcome only through their effect on the exposure. Horizontal pleiotropy, where an instrument impacts the outcome through a causal pathway that does not include the exposure, violates this assumption. This risk is typically adjusted for by removing instruments strongly associated with confounders of the intended outcome, also called pleiotropy screening.

3.1 Selection of genetic instruments

Exposure instruments

We used standard MR preprocessing steps for exposure instruments[16], with exceptions made for the sex-specific sensitivity analyses of CETP. We selected exposure instruments from exposure GWAS using the genome-wide significance p-value threshold ($p = 5 \times 10^{-8}$). For the sex-specific CETP sensitivity analysis we used the Benjamini-Hochberg adjusted significance threshold of $q < 0.05$ from the original publication of these instruments[35].

Once instruments were selected, we performed linkage disequilibrium (LD) clumping with threshold $r^2 = 0.001$ across a 10,000kb window, and dropped SNPs with $F < 10$, excluding sex-specific CETP instruments. We then applied pleiotropy screening on an as-needed basis for highly confounded (behavioral) phenotypes using the PheWAS function in the TwoSampleMR package (Table 1). We used regex to match phenotype strings to multiple PheWAS hits with uncertain phrasing, for example using “educat” to capture both “educational attainment”, and “years of education” (Table 1). SNPs with genome-wide significant PheWAS that matched to a specified regex string were dropped.

Existing MR-PheWAS literature is mixed on whether to use pleiotropy screening[53] or not[15]. Due to the phenome-wide nature of this MR study, we determined it would bias results to apply pleiotropy screening on outcomes that appear as phenome-wide endpoints, or on intermediate phenotypes such as lipid levels that are causally connected to diseases that appear as outcomes. We intended to take a hypothesis-free approach to the effects of each exposure and therefore did not want to rule out valid causal pathways leading to at least one disease present as an outcome phenotype. However, we assumed pleiotropy screening was useful on an as-needed basis to separate the effects of highly correlated behavioral phenotypes not likely to appear as ICD-10-coded endpoints. For this reason, we performed MR-PheWAS of our positive exposure control of smoking initiation both with and without pleiotropy screening for educational and adiposity phenotypes.

Outcome Instruments

We obtained instruments for the outcomes in the MR-PheWAS of each ageing phenome using publicly available GWAS summary statistics of the UKB[21,22].

We started the outcome-instrument gathering process by mapping diseases from the GBD to ICD-10 code GWASs with available summary statistics in the PanUKBB or Neale Lab round 2 UKB GWAS using a GBD-published mapping of diseases-to-codes. Then, we constructed an outcome plan that listed all outcome phenotypes to be analyzed in that MR-PheWAS run, along with their status as age-related diseases or not (ARD selection status) and GBD category values to allow for permutation tests of ARD and category enrichment. The outcome plan contained all GWAS-linked diseases in the Global Burden of Disease survey, but diseases differed in ARD selection status depending on the sex of the population analyzed. ARD selection status is defined by sex-specific ageing phenomes[29].

In cases of one-to-many mapping, where a single GBD disease was mapped to multiple ICD-10 codes, we preserved all codes as their own outcomes, and they inherited the same ARD status as their parent. We resolved cases of many-to-one mappings to the single ICD-10 code outcome. We decided ARD status for this surviving ICD-10 code outcome by majority vote of the ARD status on the multiple diseases mapped to it. We decided ties in favor of positive ARD status. We defined GBD hierarchy group values uninformatively from the first duplicate mapped disease with the majority vote ARD status.

Next, we gathered outcome SNPs for every GWAS into the outcome plan. We implemented this differently depending on whether the population studied included both male and female participants, or only a single sex. In the case of a both-sex population of any ancestry, we gathered outcome SNPs for each outcome phenotype using a Tabix interaction with the GWAS summary statistic files hosted on a PanUKBB-owned AWS cloud. Tabix is an indexing tool for genomic files that allows users to query summary statistics files by genomic coordinates, allowing for rapid retrieval[54]. We used Tabix to query only the outcome SNPs needed from the PanUKBB GWASs, allowing for streaming of the outcome SNPs in real time. Once a run had been completed, we saved a lightweight R dataset file containing only the streamed outcome SNPs in a local cache to save bandwidth for repeated runs.

In the case of a single-sex MR-PheWAS run, we extracted SNPs for a given outcome from a locally cached copy of the entire GWAS summary statistics for that outcome. If these files did not exist for all outcomes, we downloaded the missing files from a Neale Lab-owned AWS cloud containing their published GWAS summary statistics[21]. Once all outcome GWAS were available, we pulled the outcome SNPs required for the run with a non-Tabix search of each GWAS summary statistics file.

We then performed a pre-MR harmonization on the exposure and outcome SNPs. We aligned the effect allele and non-effect allele between exposure and outcome to prevent sign flips. We then resolved palindromic alleles using effect allele frequencies. Finally, we filtered and dropped allele mismatches.

Controls

We used the standard MR practice of testing our instruments for GLP-1R and CETP against positive and negative outcome control phenotypes[16]. This step was intended to confirm that MR using these instruments identify presumed-true causal relationships of our exposures (positive control) and fail to identify presumed-false causal relationships of our exposures (negative control). For GLP-1R exposure, we used type 2 diabetes[22] (E11) diagnosis as measured in the European both-sex population as our positive outcome control phenotype, as GLP-1R agonists were originally approved to treat type 2 diabetes. For our CETP

exposure, we used high cholesterol[25] as measured in the European both-sex population as our positive outcome control phenotype, as the CETP inhibitor drug Obicetrapib is under review to treat high cholesterol[13]. For both GLP-1R and CETP, we used brown hair color[24] in the European both-sex population as our negative outcome control phenotype.

In addition to using positive and negative outcome controls to validate our exposure instruments, we also tested our MR-PheWAS pipeline with positive and negative exposure controls. This was intended to confirm expected behavior of our MR-PheWAS pipeline under known-risky and known harmless exposures, including across multiple sexes and ancestry groups. For positive exposure controls, we used smoking initiation in the European both-sex population[28], smoking initiation in the East Asian both-sex population[26,27], and smoking initiation in the East Asian men and women[26,27] separately. These GWAS had no sample overlap with our UKB-based outcome GWASs, which we deemed necessary for positive exposure controls. We predicted smoking initiation would be a significant risk factor for "C34 Malignant neoplasm of bronchus and lung" and "J44 Other chronic obstructive pulmonary disease", and not protective of any disease from the level-2 GBD categories "Chronic respiratory diseases" or "Cardiovascular diseases". For negative exposure controls, we used left-handedness and ambidextrousness, both measured in the European both-sex population[21,26,28]. We predicted left-handedness and ambidextrousness would not have phenome-wide significant effects on any disease. The instruments for these two phenotypes come from GWASs in the UK Biobank, indicating nearly complete sample overlap with our phenome-wide outcome data. We deemed this acceptable for a negative but not positive control phenotype due to overlap driving results away from the null[55]. We predicted handedness would not be significantly associated with any disease.

3.2 MR-PheWAS Implementation

Primary estimators

After the exposure and outcome SNPs were quality controlled and harmonized, we conducted an MR-PheWAS by fitting a primary MR estimator of the exposure to each mapped outcome in the GBD-defined phenome. When conducting the MR-PheWAS, we used the Wald ratio as the primary estimator for exposure phenotypes with only one instrument surviving quality control[17]. The Wald ratio for exposure SNP J is shown in equation (1) and describes the estimated regression coefficient of SNP J on the outcome ($\widehat{\beta_{GY,J}}$), scaled by the estimated regression coefficient of the exposure SNP on the exposure ($\widehat{\beta_{GX,J}}$). The variance of the Wald ratio used to calculate confidence intervals shown in equation (2) is estimated using the delta method, which returns a linear approximation of the true variance[56].

(1) $$\widehat{\theta_J} = \frac{\widehat{\beta_{GY,J}}}{\widehat{\beta_{GX,J}}}$$

(2) $$\mathrm{Var}\left(\widehat{\theta_J}\right) \approx \frac{\mathrm{SE}\left(\widehat{\beta_{GY,J}}\right)^2}{\widehat{\beta^2_{GX,J}}} + \frac{\widehat{\beta^2_{GY,J}}\,\mathrm{SE}\left(\widehat{\beta_{GX,J}}\right)^2}{\widehat{\beta^4_{GX,J}}}$$

We used the inverse variance weighted (IVW) method as the primary estimator for exposure phenotypes with more than one instrument surviving quality control. IVW estimates the true effect of the exposure on the outcome using equation (3). Equations (3) and (4) show a precision-weighted estimate of the Wald ratios of K exposure SNPs, where the weight of

SNP j is determined using the estimated variance of the Wald ratio (4). The confidence intervals for $\widehat{\theta_{\mathrm{IVW}}}$ in equation (3) are estimated using the variance formula shown in equation (5).

(3) $$\widehat{\theta_{\mathrm{IVW}}} = \frac{\sum_{j=1}^{K} w_j \widehat{\theta_j}}{\sum_{j=1}^{K} w_j}$$

(4) $$w_j = \frac{1}{\mathrm{Var}(\widehat{\theta_j})}$$

(5) $$\mathrm{Var}\left(\widehat{\theta_{\mathrm{IVW}}}\right) = \frac{1}{\sum_{j=1}^{K} w_j}$$

Robustness Checks

Due to the hundreds of MR estimators fit in a single MR-PheWAS of a single population, it was infeasible to manually apply and interpret robustness checks as is typical for single-outcome MR[16]. Instead, we employed an automated robustness check procedure based on prior phenome-wide MR publications[15,53,57], aiming to meet or exceed the existing standard for robustness checks. We applied eight robustness checks to each exposure-outcome MR where the number of exposure SNPs (i.e., instrumental variables) was greater than one, then dropped outcomes from the MR-PheWAS that passed less than a pre-determined threshold number of tests (Supplementary Table 1). We chose our robustness checks using guidance from MR methods papers that recommend multiple robustness checks with varied assumptions to reduce the risk of hidden assumption violations, including at least one weighted median and one weighted mode robustness check[16,23]. Our robustness checks were as follows:

1. We calculated MR-Egger, which relaxes the exclusion-restriction assumption, allowing for horizontal pleiotropy but requiring that the pleiotropic effects are independent of the instrument-exposure association strength. This is also referred to as the InSIDE assumption[16]. We coded sign disagreement of the MR-Egger slope with the primary IVW slope as a failure.
2. We subjected the intercept of the MR-Egger regression to a hypothesis test against a null of H_0 : intercept = 0. If we found an intercept not equal to zero with a p-value < 0.05, we coded the test as a failure.
3. We calculated Cochran's Q, which tests for heterogeneity of instrument effects[57]. If we found heterogeneity with a p-value < 0.05, the test was recorded as a failure.
4. We calculated I^2 as a function of Cochran's Q[58], which estimated the proportion of observed heterogeneity that is attributed to true heterogeneity of instrument effect, rather than sampling error. If we found I^2 > 50%, the test was recorded as a failure.
5. We calculated weighted median of the Wald ratios and checked for sign disagreement of the slope with the IVW estimate[16]. Sign disagreements were recorded as a failure.
6. We calculated the weighted mode of the Wald ratios and checked for sign disagreement of the slope with the IVW estimate[16]. Sign disagreements were recorded as a failure.
7. We calculated Steiger directionality tests for each instrument to check whether the observed strength of the associations indicated that the instrument had a large impact on the outcome rather than the exposure, which would suggest pleiotropy or

reverse-causation[15]. If more than 30% of the instruments failed the Steiger directionality test, we recorded a robustness check failure.

8. We calculated a leave one out (LOO) analysis of the IVW estimate[59]. We then checked for sign flips of the slope under all the LOO folds, which would suggest heterogeneity of the effect between instruments. If we discovered a sign flip in any fold, the test was recorded as a failure.

We did not apply this robustness check procedure to exposures with only a single instrumental variable surviving initial Quality Control (QC), as all checks except for Steiger directionality could not be implemented. Instead, outcomes in single-instrument MR-PheWAS were dropped only if their Wald ratios were undefined. Due to the reduced confidence in the phenome-wide validity of these MR-PheWAS, we reported the individual MR results but did not conduct tests for a phenome-wide effect on age-related disease, or on any disease category.

After filtering exposure-outcome pairs that failed the MR sensitivity analysis check, we applied Benjamini-Hochberg false discovery rate adjustment to the p-values of surviving outcomes. We reported outcomes as significant if their β coefficient value differed from zero with q < 0.05 after multiple testing correction.

4. Phenome-wide effects on age-related disease

We used permutation testing to identify an aggregated effect of our exposures on the sex-specific ageing phenomes, which are a subset of the full phenome explored in the MR-PheWAS. For the permutation tests, we constructed signed disease scores s_i for each surviving MR outcome result by multiplying the negative log_{10} p-value of the IVW beta coefficient $p_{\mathrm{IVW},i}$ by the sign of the beta coefficient $\mathrm{sign}(\beta_{\mathrm{IVW},i})$, as shown in equation (6). This signed disease score is intended to capture both confidence and direction of an effect for that disease. Then, we constructed the observed test statistic T_{obs} that captures the precision-weighted mean of the signed disease scores across diseases in the ageing phenome shown in equation (7). As shown in equations (8) and (9), y_i is ARD selection status and w_i is weight given by the inverse of the squared standard error of the IVW estimate for that disease.

(6) $$s_i = \mathrm{sign}(\beta_{\mathrm{IVW},i})(-\log_{10}(p_{\mathrm{IVW},i}))$$

(7) $$T_{\mathrm{obs}} = \sum_{i=1}^{n} y_i \; w_i \, s_i$$

(8) $$y_i = 1 \text{ if outcome } i \text{ is ARD, else } 0$$

(9) $$w_i = 1/SE_i^2$$

We then conducted permutation testing on T_{obs} against a null hypothesis that ARD selection status (y_i) is irrelevant to the distribution of T. We did this by shuffling the ARD selection status labels over b (=10,000) permutations or using the exact permutations if this number was less than 10,000. From this we constructed an estimate of the mean and standard deviation of T under the null hypothesis, shown in equation (10) and (11). Using equations (10) and (11), we conducted our hypothesis test on T_{obs}, yielding a signed enrichment score (SES) shown in equation (12) and a p-value for SES shown in equation (13). These

represent the standardized distance of the observed ARD disease scores T_{obs} from the null hypothesis, and the degree of confidence that $T_{\text{obs}} \neq T_{\text{null}}$.

(10) $$\mu = \frac{1}{B}\sum_{b=1}^{B} T^{(b)}$$

(11) $$\sigma = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(T^{(b)} - \mu)^2}$$

(12) $$\text{SES} = \frac{T_{\text{obs}} - \mu}{\sigma}$$

(13) $$p = \frac{1 + \sum_{b=1}^{B} 1\left(\left|T^{(b)} - \mu\right| \geq \left|T_{\text{obs}} - \mu\right|\right)}{1 + B}$$

GBD Category Testing

We also applied the same permutation testing methods to examine whether exposures had a category-wide effect on diseases by GBD category. We conducted category-wide permutation tests for each category in GBD hierarchy levels one, two and three. In this case, T_{obs} was defined as the precision-weighted sum of the signed disease scores included in that category, or for the children of diseases included in that category. For example, the level two disease “neoplasms” has a level one parent “non-communicable disease” and level three child “leukemia.” Accordingly, the permutation test of “neoplasms” included the beta coefficient for the child “leukemia” but not for the parent “non-communicable disease.” Due to the risk of multiple testing, we applied Benjamini-Hochberg FDR adjustment within each hierarchy. Where all results were null, we tested relaxing the multiple testing correction for the both-sex analyses. We did not test relaxing the multiple testing correction on the female and male specific analyses.

Literature Search for Novel Findings

For each significant GLP-1R disease finding in the both-sex analysis, we searched PubMed for terms “Mendelian randomization GLP-1R <disease>” in all search fields. We specified <disease> first as the ICD-10 coded name without the code, then using variations, simplifications and synonyms if no results were found. If no results were found through PubMed, we fell back to the Perplexity AI search engine, asking, “can you find a Mendelian randomization study of GLP-1R on <disease>?[33]” If no relevant results were found in either the PubMed searches or Perplexity searches, we recorded the finding as a novel Mendelian randomization finding.

Maternal and Neonatal Diseases

We wished to analyze the effect of GLP-1R on maternal diseases, and separately on the effect of GLP-1R on developing fetuses. We identified the GBD category “Maternal and neonatal disorders” as a close match for this aim, but noted that many ICD-10 codes (15 of 33) mapped to this category referred to billable but non-diagnostic maternal medical events, such as screening (Z36 Antenatal screening), care (Z39 Postpartum care and examination), non-pathological medical outcomes (O80 Single spontaneous delivery), or procedures (O04 Medical abortion).

To capture disease diagnosis risk, we excluded the 15 ICD-10 codes not referring to diagnoses of maternal or neonatal pathologies, retaining 18 diseases. We then separated the remaining diseases into the categories “neonatal diseases” and “maternal diseases,”

which we drew from obstetrics literature[60], using ICD-10 code definitions from the World Health Organization 2019 issue of the International Statistical Classification of Diseases and Related Health Problems 10th Revision. The category “neonatal diseases” captures health risks to either the mother or fetus originating from the fetus or from pathological fetal development, such as “O68 Labor and delivery complicated by fetal stress.” The category “Maternal diseases” captures health risks to either the mother or fetus not originating from the fetus, such as “O21 Excessive vomiting in pregnancy.” We assigned diseases originating from the umbilical cord as belonging to both the mother and the neonate as the umbilical cord has no clear owner following birth.

Prior literature search

For both CETP and GLP-1R, we searched for prior MR and MR-PheWAS studies of the form shown in Equation 14 with results sorted by date. Searches were performed on January 19th 2026.

(14) **(TARGET[Title/Abstract]) AND (Mendelian Randomization[Title/Abstract])**
(TARGET ∈ {CETP, GLP1R, GLP-1R, glucagon-like peptide 1 receptor})

5. Code Availability

The scripts used to perform all coded analyses and generate all figures in this analysis are available on GitHub[51]. The MR-PheWAS and downstream analysis listed in Table 1, excluding the outcome control columns, were completed using the R package “ardmr”[52].

The software used to apply these methods is listed according to Methods section:

2, Data Sources. This analysis was conducted in R v4.4.3 using packages readr v2.1.5, dplyr v1.1.4, stringr v1.5.1, tidyr v1.3.1, ggplot2 v3.5.2, forcats v1.0.0, grid v4.4.3, ggrepel v0.9.6, scales v1.4.0, purrr v1.1.0, tibble v3.3.0

3, MR-PheWAS. This analysis was conducted in base R version 4.4.3 and ardmr v1.0.0, an MR-PheWAS tool built to study the ageing phenome[52].

Supplementary Notes

Supplementary Note 1: Control Phenotypes

We used exposure control phenotypes to test our MR-PheWAS pipeline, as well as outcome control phenotypes to test the behavior of our GLP-1R and CETP exposure instruments.

We tested the effect of smoking initiation in the European both-sex population both with and without pleiotropy filtering exposure SNPs on obesity and education-related outcomes. This was intended as a sensitivity analysis of our positive exposure control of smoking. When including pleiotropy filtering, only 154 of 396 outcomes had available data for the single SNP instrument (Table 1, Supplementary Table 1). No results reached significance at Benjamini-Hochberg $q < 0.05$. Permutation testing was not performed because robustness checks were not applied due to the singular SNP exposure instrument. To increase the power of our positive control instruments, we abandoned the pleiotropy filtering on educational and adiposity related outcomes and reran the MR-PheWAS pipeline on smoking initiation (Table 1). In this relaxed run, reported in Results, 92 exposure SNPs passed quality control, and

300 disease outcomes tested passed robustness checks. Smoking initiation appeared as a phenome-wide significant risk factor for 50 diseases, and protective of no diseases (Supplementary Table 22, Supplementary Figures 24,25). The two strongest hits, which differed by ~20 orders of magnitude, were “Mental and behavioral disorders due to use of tobacco” ($q = 6.18 \times 10^{-41}$, ICD-10 code = F17) and “Other chronic obstructive pulmonary disease” ($q = 4.49 \times 10^{-19}$, ICD-10 code = J44).

Naïve grouping by category shows a risk effect of smoking on nine diseases from the “chronic respiratory disease” category (GBD cause level 2), as well as four mental and substance use disorders including alcoholism (F10) and depressive episodes (F32). Permutation tests for disease-group effects on age-related disease and GBD disease categories were directionally expected but did not reach statistical significance after multiple testing correction (Supplementary Figures 26-29). The strongest GBD group findings prior to multiple testing correction were for “substance use disorders” ($p = 0.005$), “chronic obstructive pulmonary disease” ($p = 0.022$), and “atrial fibrillation and flutter” ($p = 0.037$).

We additionally ran MR-PheWAS of smoking initiation in the East Asian both-sex population. We did not include pleiotropy screening due to concerns about loss of statistical power. For the East Asian both-sex population, 20 phenotypes (12 ARD, 8 non-ARD) passed robustness checks. Of these, none were significant. Permutation analysis was null at all levels.

We attempted sex-specific MR-PheWAS of smoking initiation in the East Asian female and male populations. No instruments survived exposure quality control, terminating these runs before results.

For the negative control phenotype of left-handedness in the European both-sex population, two exposure SNPs passed quality control, and none of the 396 disease outcomes tested (143 ARD and 172 non-ARD) passed MR robustness checks. For the negative control of ambidextrousness, only one exposure SNP passed quality control, and 118 outcomes (68 ARD and 50 non-ARD) had available outcome SNP data. We did not apply the robustness check pipeline as it is undefined on single-instrument exposures. No surviving results were significant at Benjamini-Hochberg adjusted $q < 0.05$ (Supplementary Table 21, Supplementary Figure 23). We did not conduct permutation testing on the left-handed phenotype as it had no robust results, or on the ambidextrous phenotype as it could not be robustness checked.

Supplementary Note 2: GLP-1R Instruments

Our main analysis of GLP-1R used exposure instruments derived from eQTLgen, an international consortium of gene expression quantitative trait loci from both-sex human blood[32]. We used these both-sex instruments in all analyses of GLP-1R, including for the male and female specific runs. This data limitation presents a potential concern for the sex-specific analyses if the both-sex instruments are uncalibrated for predicting GLP-1R levels in a single-sex population. To mitigate this concern, we searched the literature for and did not find evidence that GLP-1R has meaningfully sex-specific QTLs. This finding concurs with the broader set of findings that blood QTLs generally show little evidence of sex-specificity[47].

Supplementary Note 3: GLP-1R Validation

To address our lack of a validation cohort for our GLP-1R findings, we performed a within-UKB biobank sensitivity analysis by treating the MR-PheWAS in women as discovery, and the MR-PheWAS in men as validation. This technique was made possible by the sex-specific GWAS published by the Neale Lab round 2 GWAS of the UK Biobank[21], which allows for partitioning the participants without access to the individual level data. While this validation method cannot validate sex-specific effects, we suggest it robustly confirms the sex-conserved nature of effects observed in the both-sex population. Accordingly, sex-specific findings from the MR-PheWAS of GLP-1R should be treated as unvalidated.

Supplementary Note 4: CETP Instruments

Data limitations forced a trade-off between using both-sex instruments for sex-specific analysis as we did with GLP-1R, or alternatively using sex-specific instrument(s) selected with lower-than-standard selection criteria. We opted for the latter option, marking deviations from standard best practice with "*" and "**" in Table 1 and Supplementary Table 1. Our sex-specific analyses of CETP did not produce any significant results, thus we do not believe this preprocessing choice undermines the significant findings of this paper.

Supplementary Note 5: Robustness Checks

Our robustness check procedure commonly left zero or only a few surviving outcomes, which reduced the "phenome-wide" nature of the findings. Due to the hundreds of MR estimators fit in a single MR-PheWAS, it was infeasible to manually interpret each set of robustness checks. To address this, we implemented a broader array of robustness checks than would be used for a single-outcome MR with the intention of capturing most sources of assumption violations across phenotypes[16], meeting the standard set by other phenome-wide Mendelian randomization studies[15,53,57]. However, we hypothesize that the automated nature of these tests makes them less discerning than manually interpreted tests, and forces a larger sensitivity-specificity trade-off. We chose to prioritize specificity, which resulted in null findings or complete or partial failures-to-fit tests for 9 of the 14 MR-PheWAS we ran. We suggest the importance of further development of automated robustness checks in phenome-wide Mendelian randomization.

Supplementary Note 6: Permutation Testing

Permutation testing was chosen as a method for group effect testing due to data constraints. The preferred approach would have been a t-test of the mean within-group MR effect against zero, but this requires estimating the covariance structure of MR beta coefficients. Assuming zero covariance was considered inappropriate, as it would systematically bias results away from the null. The permutation test instead makes weaker assumptions: under the null, disease scores and weights are assumed exchangeable between ARD and non-ARD labels. While the sign of MR effects should be randomly distributed around zero if MR assumptions hold, the associated p-values and standard errors may not be exchangeable. This assumption may be violated if ARD and non-ARD outcomes differ in GWAS power or in their likelihood of passing MR robustness checks, potentially biasing results toward or away from the null.

Future work could address limitations to this approach by permuting only effect signs, ignoring effect size and precision, or by using individual-level data to estimate covariance structures or to test category-wide effects directly through aggregated GWAS outcomes.

This first option would reduce power while reducing bias, while the second, if possible, may increase power while reducing bias.

Bibliography

1. 20 years of Nature Reviews Cardiology. *Nat Rev Cardiol* **21**, 735–736 (2024).

2. Goodall, R. *et al.* Trends in type 2 diabetes mellitus disease burden in European Union countries between 1990 and 2019. *Sci Rep* **11**, 15356 (2021).

3. Kaiser Family Foundation (KFF). KFF Health Tracking Poll: Prescription Drug Costs, Views on Trump Administration Actions, and GLP-1 Use. *KFF* https://www.kff.org/public-opinion/kff-health-tracking-poll-prescription-drug-costs-views-on-trump-administration-actions-and-glp-1-use/ (2025).

4. Mallapaty, S. Cheaper versions of blockbuster obesity drugs are being created in India and China. *Nature* **630**, 797–798 (2024).

5. Xie, Y., Choi, T. & Al-Aly, Z. Mapping the effectiveness and risks of GLP-1 receptor agonists. *Nat Med* **31**, 951–962 (2025).

6. Are GLP-1s the first longevity drugs? *Nat Biotechnol* **43**, 1741–1742 (2025).

7. Vargesson, N. Thalidomide-induced teratogenesis: History and mechanisms. *Birth Defects Research Part C: Embryo Today: Reviews* **105**, 140–156 (2015).

8. Qiao, L., Lu, C., Zang, T., Dzyuba, B. & Shao, J. Maternal GLP-1 receptor activation inhibits fetal growth. *Am J Physiol Endocrinol Metab* **326**, E268–E276 (2024).

9. Ramamoorthy, R. *et al.* GLP-1-Mediated Pregnancy and Neonatal Complications in Mice. *J Dev Biol* **13**, 29 (2025).

10. Granhall, C. *et al.* Safety and Pharmacokinetics of Single and Multiple Ascending Doses of the Novel Oral Human GLP-1 Analogue, Oral Semaglutide, in Healthy Subjects and Subjects with Type 2 Diabetes. *Clin Pharmacokinet* **58**, 781–791 (2019).

11. Singh, S., Sedgh, G. & Hussain, R. Unintended Pregnancy: Worldwide Levels, Trends, and Outcomes. *Studies in Family Planning* **41**, 241–250 (2010).

12. Hsieh, A. *et al.* Obicetrapib Demonstrates Significant Reductions Of Lp(a) On Top Of High-intensity Statins. *Journal of Clinical Lipidology* **18**, e563–e564 (2024).

13. Menarini Group's Obicetrapib and Obicetrapib/Ezetimibe Marketing Authorisation Applications Accepted for Review by the European Medicines Agency (EMA) for the Treatment of adults with primary hypercholesterolaemia (heterozygous familial and non-familial). https://www.menarini.com/en-us/news/news-detail/menarini-group-s-obicetrapib-and-obicetrapib-ezetimibe-marketing-authorisation-applications-accepted-for-review-by-the-european-medicines-agency.html.

14. News in brief. *Nature Reviews Drug Discovery* **6**, 14–15 (2007).

15. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet* **52**, 1122–1131 (2020).

16. Burgess, S. *et al.* Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Res* **4**, 186 (2019).

17. Sanderson, E. *et al.* Mendelian randomization. *Nat Rev Methods Primers* **2**, 6 (2022).

18. Schmidt, A. F. *et al.* Genetic drug target validation using Mendelian randomisation. *Nat Commun* **11**, 3255 (2020).

19. Reitz, J. *et al.* Genetically modeled GLP1R and GIPR agonism reduce binge drinking and alcohol-associated phenotypes: a multi-ancestry drug-target Mendelian randomization study. *Mol Psychiatry* **30**, 6119–6133 (2025).

20. Global Burden of Disease Collaborative Network & Institute for Health Metrics and Evaluation (IHME). Global burden of disease study 2021 (GBD 2021) results. (2022).

21. Neale Lab. UK Biobank GWAS Round 2 results (released 1 Aug 2018). Broad Institute (2017).

22. Karczewski, K. J. *et al.* Pan-UK Biobank genome-wide association analyses enhance discovery and resolution of ancestry-enriched effects. *Nat Genet* **57**, 2408–2417 (2025).

23. Gill, D. *et al.* Common pitfalls in drug target Mendelian randomization and how to avoid them. *BMC Medicine* **22**, 473 (2024).

24. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat Genet* **50**, 906–908 (2018).

25. Dönertaş, H. M., Fabian, D. K., Valenzuela, M. F., Partridge, L. & Thornton, J. M. Common genetic associations between age-related diseases. *Nat Aging* **1**, 400–412 (2021).

26. Matoba, N. *et al.* GWAS of smoking behaviour in 165,436 Japanese people reveals seven new loci and shared genetic architecture. *Nat Hum Behav* **3**, 471–477 (2019).

27. Ishigaki, K. *et al.* Large scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet* **52**, 669–679 (2020).

28. Furberg, H. *et al.* Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441–447 (2010).

29. Campbell, R. Translational Geroprotective Therapies Leveraging Age Related Disease Phenotype (Forthcoming). https://osf.io/5j3g7 (2026).

30. Skrivankova, V. W. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. *JAMA* **326**, 1614–1621 (2021).

31. Sun, Y. *et al.* Association of glucagon-like peptide-1 receptor agonists with risk of cancers-evidence from a drug target Mendelian randomization and clinical trials. *International Journal of Surgery* **110**, 4688 (2024).

32. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* **53**, 1300–1310 (2021).

33. *Perplexity*. https://www.perplexity.ai/.

34. Blauw, L. L. *et al.* CETP (Cholesteryl Ester Transfer Protein) Concentration. *Circulation: Genomic and Precision Medicine* **11**, e002034 (2018).

35. Kanoni, S. *et al.* Implicating genes, pleiotropy, and sexual dimorphism at blood lipid loci through multi-ancestry meta-analysis. *Genome Biol* **23**, 268 (2022).

36. Pazarci, Ö., Dogan, H. O., Kilinc, S. & Çamurcu, Y. Evaluation of Serum Glucagon-Like Peptide 1 and Vitamin D Levels in Elderly Patients with Bone Fractures. *Med Princ Pract* **29**, 219–224 (2020).

37. Johnson, B. *et al.* Investigating nutrient intake during use of glucagon-like peptide-1 receptor agonist: a cross-sectional study. *Front. Nutr.* **12**, (2025).

38. Spiro, A. & Buttriss, J. L. Vitamin D: An overview of vitamin D status and intake in Europe. *Nutr Bull* **39**, 322–350 (2014).

39. Yunas, I. *et al.* Causes of and risk factors for postpartum haemorrhage: a systematic review and meta-analysis. *The Lancet* **405**, 1468–1480 (2025).

40. Beta, J. *et al.* Maternal and neonatal complications of fetal macrosomia: systematic review and meta-analysis. *Ultrasound Obstet Gynecol* **54**, 308–318 (2019).

41. Räisänen, S., Vehviläinen-Julkunen, K., Gissler, M. & Heinonen, S. Smoking during Pregnancy Is Associated with a Decreased Incidence of Obstetric Anal Sphincter Injuries in Nulliparous Women. *PLoS One* **7**, e41014 (2012).

42. Westergaard, D. *et al.* O-046 Obesity, implantation and early pregnancy - does GLP1 play a role? *Hum Reprod* **40**, deaf097.046 (2025).

43. Dao, K. *et al.* Use of GLP1 receptor agonists in early pregnancy and reproductive safety: a multicentre, observational, prospective cohort study based on the databases of six Teratology Information Services. *BMJ Open* **14**, e083550 (2024).

44. Watson, K. & Angelotta, C. The frequency of pregnancy recognition across the gestational spectrum and its consequences in the United States. *Perspect Sex Reprod Health* **54**, 32–37 (2022).

45. U.S. Food and Drug Administration WEGOVY Label. https://www.accessdata.fda.gov/drugsatfda_docs/label/2025/215256s023lbl.pdf.

46. Müller, T. D. *et al.* Glucagon-like peptide 1 (GLP-1). *Mol Metab* **30**, 72–130 (2019).

47. Porcu, E. *et al.* Limited evidence for blood eQTLs in human sexual dimorphism. *Genome Med* **14**, 89 (2022).

48. Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).

49. Elsworth, B. *et al.* The MRC IEU OpenGWAS data infrastructure. 2020.08.10.244293 Preprint at https://doi.org/10.1101/2020.08.10.244293 (2020).

50. Allen, N. *et al.* UK Biobank: Current status and what it means for epidemiology. *Health Policy and Technology* **1**, 123–126 (2012).

51. Campbell, R. Robert-Henry-Campbell/GLP1R_CETP_scripts. https://github.com/Robert-Henry-Campbell/GLP1R_CETP_scripts (2025).

52. Campbell, R. Robert-Henry-Campbell/ARD-Mendelian-Randomization. https://github.com/Robert-Henry-Campbell/ARD-Mendelian-Randomization (2025).

53. Went, M. *et al.* Phenome-wide Mendelian randomisation analysis of 378,142 cases reveals risk factors for eight common cancers. *Nat Commun* **15**, 2637 (2024).

54. Li, H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* **27**, 718–719 (2011).

55. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* **40**, 597–608 (2016).

56. Thomas, D. C., Lawlor, D. A. & Thompson, J. R. Re: Estimation of bias in nongenetic observational studies using ‘Mendelian triangulation’ by Bautista et al. *Ann Epidemiol* **17**, 511–513 (2007).

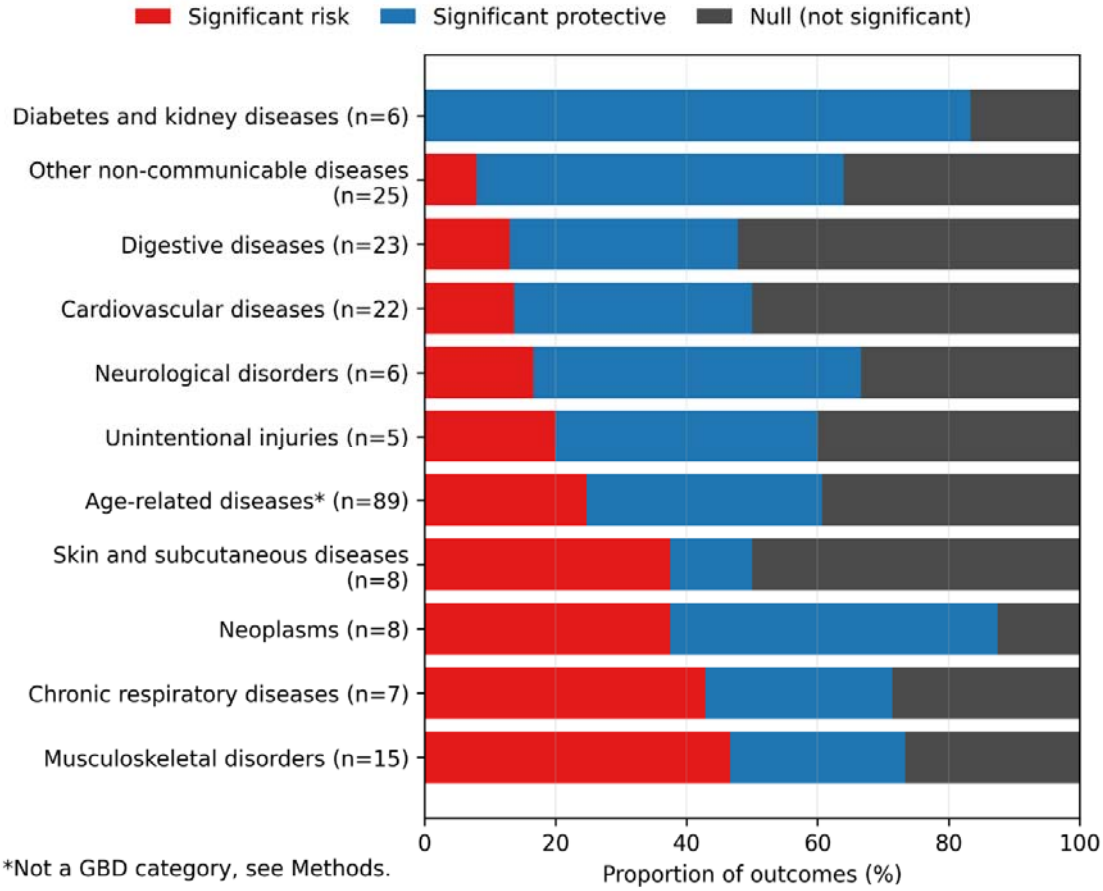
57. Zhang, Y., Wang, X. & Yang, J. Phenome-wide Mendelian randomization analysis reveals new risk factors for 4 inflammatory skin diseases. *Medicine (Baltimore)* **104**, e45569 (2025).

58. Julian, T. H. *et al.* Phenome-wide Mendelian randomisation analysis identifies causal factors for age-related macular degeneration. *eLife* **12**, e82546 (2023).

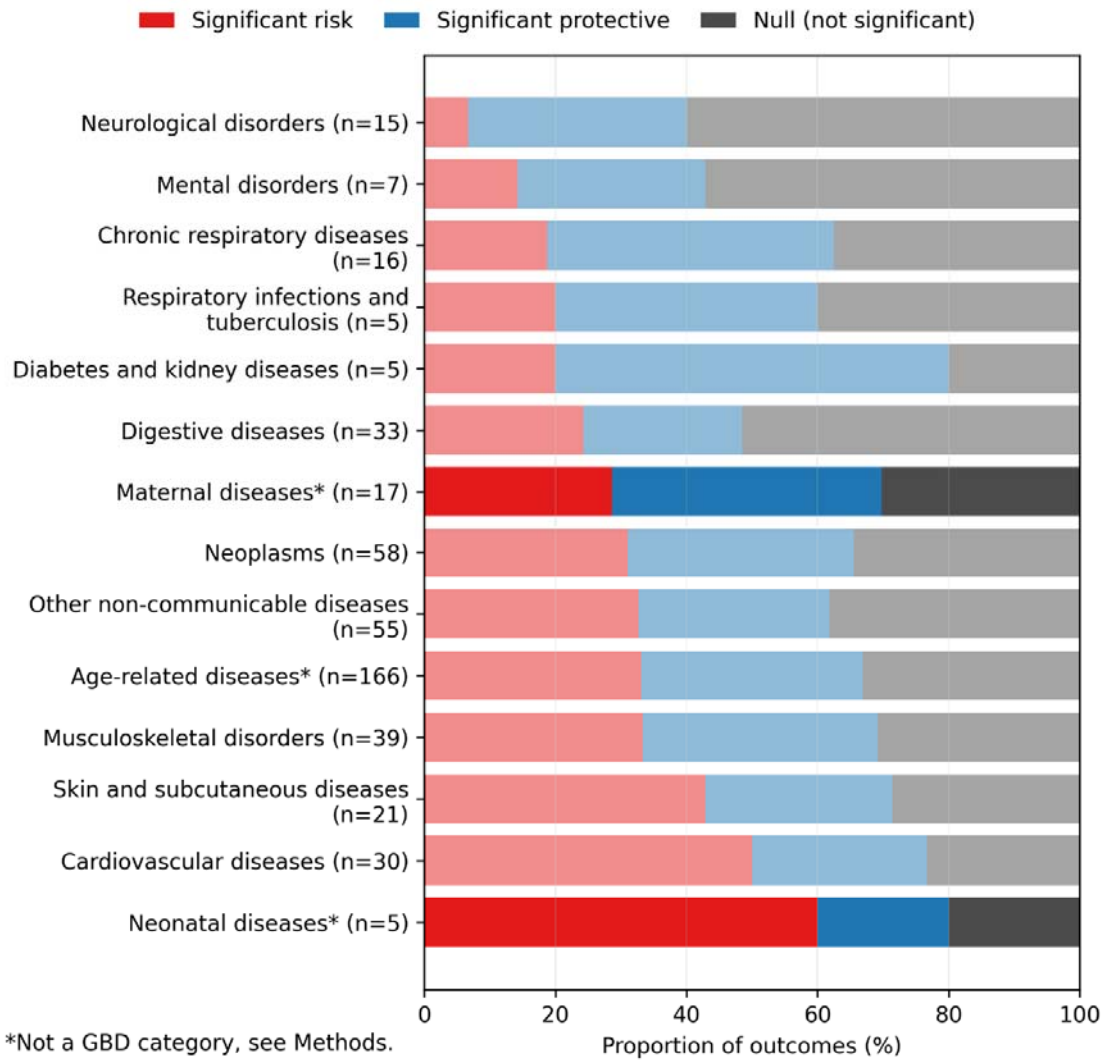
59. Saunders, C. N. *et al.* Searching for causal relationships of glioma: a phenome-wide Mendelian randomisation study. *Br J Cancer* **124**, 447–454 (2021).

60. Hong, J. *et al.* Comparison of Maternal Labor-Related Complications and Neonatal Outcomes Following Elective Induction of Labor at 39 Weeks of Gestation vs Expectant Management: A Systematic Review and Meta-analysis. *JAMA Netw Open* **6**, e2313162 (2023).

Extended Data



Extended Data Figure 1, GLP-1R disease effects grouped by Global Burden of Disease level-2 category (both-sex population). Colored bars represent categories of disease drawn from level two of the Global Burden of Disease ontology. The bracketed number (n = X) next to the category name represents the number of ICD-10 coded diseases tested in the MR-PheWAS of GLP-1R expression mapped to that category. The red colored portion of each bar represents the proportion of diseases sorted into that category on which increased GLP-1R expression has robust, phenome-wide significant risk effect. The blue colored portion of each bar represents the proportion of diseases sorted into that category on which increased GLP-1R expression has a robust, phenotype-wide significant protective effect. The gray colored portion of bar represents the proportion of diseases sorted into that category on which increased GLP-1R expression does not have a directionally significant effect. Categories marked with an asterisk were not derived from the Global Burden of Disease ontology. Bars are sorted from the lowest proportion of (red) risk effects to the highest. Categories with fewer than five mapped diseases are excluded.



Extended Data Figure 1, female sex-specific GLP-1R disease effects grouped by Global Burden of Disease level-2 category (maternal and neonatal diseases highlighted). Colored bars represent categories of disease, with maternal and neonatal diseases highlighted. The bracketed number (n = X) next to the category name represents the number of ICD-10 coded diseases tested in the MR-PheWAS of GLP-1R expression mapped to that category. The red colored portion of each bar represents the proportion of diseases sorted into that category on which increased GLP-1R expression has robust, phenome-wide significant risk effect. The blue colored portion of each bar represents the proportion of diseases sorted into that category on which increased GLP-1R expression has a robust, phenotype-wide significant protective effect. The gray colored portion of bar represents the proportion of diseases sorted into that category on which increased GLP-1R expression does not have a directionally significant effect. Categories marked with an asterisk were not derived from the Global Burden of Disease ontology. Bars are sorted from the lowest proportion of (red) risk effects to the highest. Categories with fewer than five mapped diseases are excluded.

Acknowledgements

R.H.C. is supported by the Leverhulme Trust Biopsychosocial Doctoral Training Programme (DS-2020-053), granted to M.C.M. M.C.M. is supported by ESSGN (HORIZON-MSCA-DN-

2021 [101073237]), ESRC/UKRI Connecting Generations (ES/W002116/1) and Einstein Foundation Berlin (EZ-2019-555-2). R.H.C. and M.C.M are supported by the Leverhulme Trust Large Centre Grant LCDS (RC-2018-003).

We thank Benjamin Woolf for his expert advice on the strengths and limitations of Mendelian Randomization.

Author information

Authors and affiliations

Leverhulme Centre for Demographic Science, Nuffield Department of Population Health, University of Oxford and Nuffield College, Oxford, UK

Robert H. Campbell and Melinda C. Mills

Department of Genetics, University Medical Centre Groningen, Groningen, The Netherlands

Melinda C. Mills

Department of Economics, Econometrics and Finance, University of Groningen, Groningen, The Netherlands

Melinda C. Mills

Author ORCID IDs

Robert H. Campbell, https://orcid.org/0009-0002-0212-224X

Melinda C. Mills, https://orcid.org/0000-0003-1704-0001

Contributions

R.H.C. conceptualized the project, led the study design and data analysis, created all display items, the R package and drafted the initial manuscript. M.C.M. contributed to drafting the initial manuscript, revision and secured funding and supervised the project. All authors reviewed the manuscript for accuracy, edited subsequent revisions of the draft, and approved the final version.

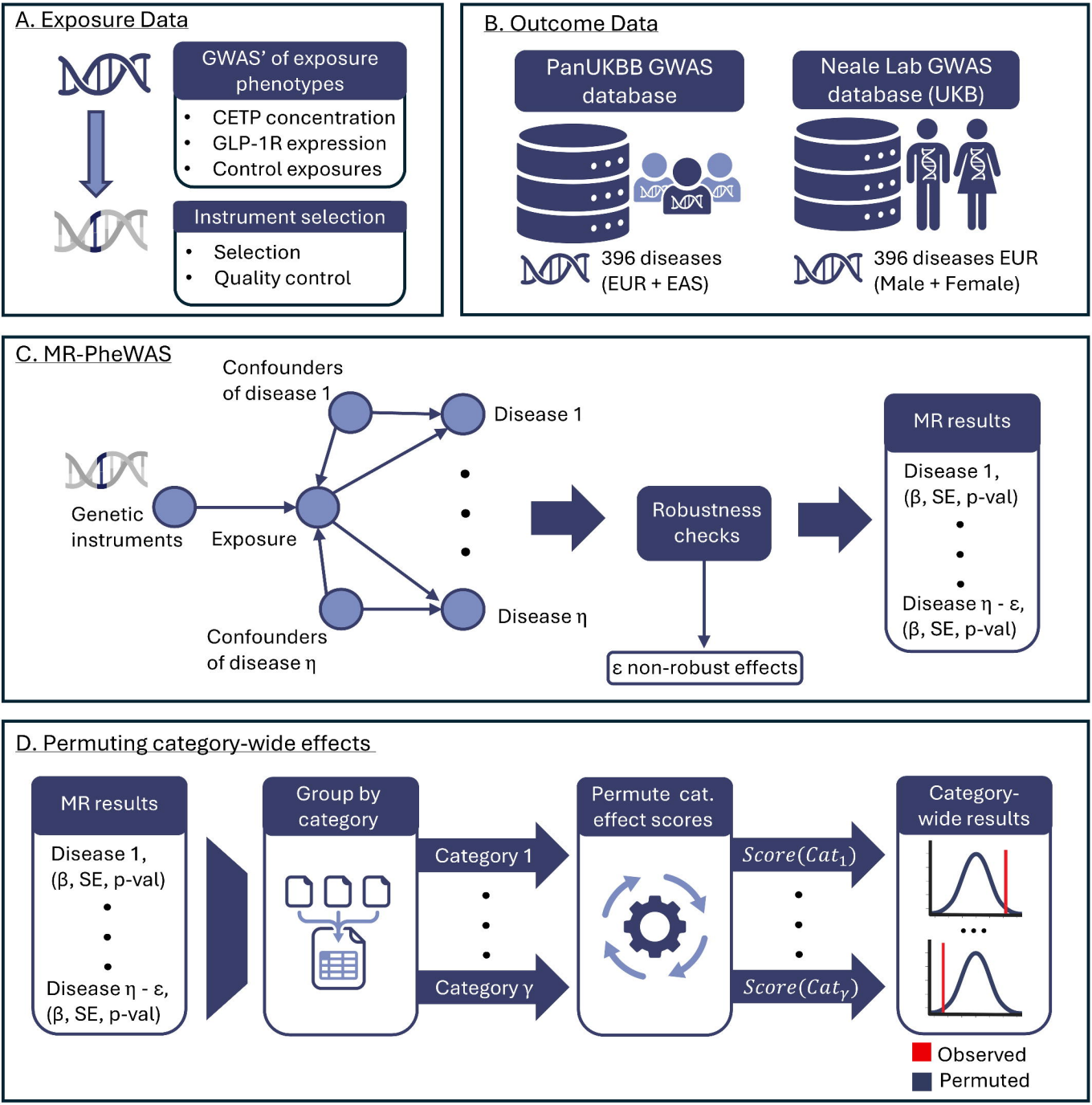
Corresponding authors

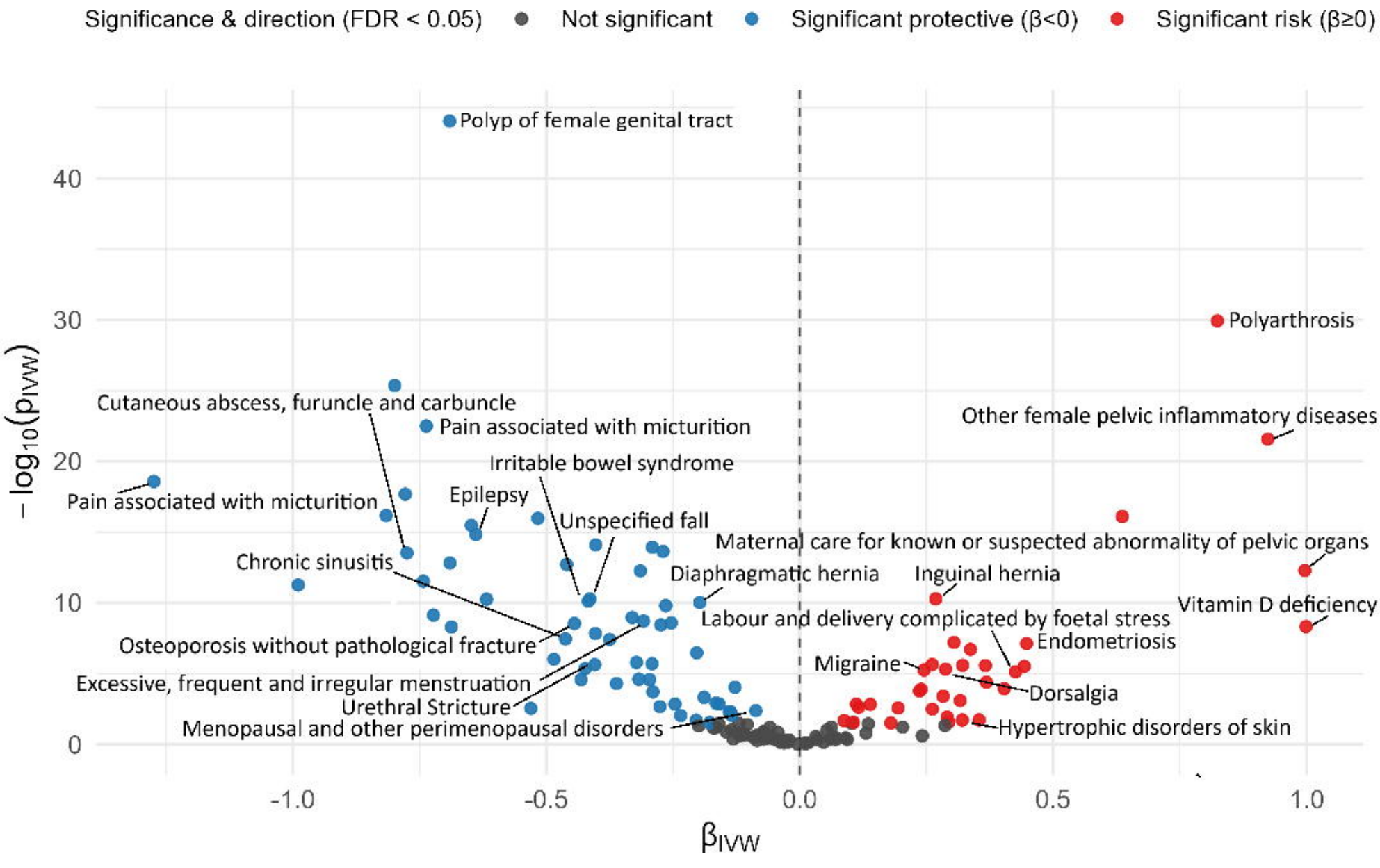
Robert H. Campbell

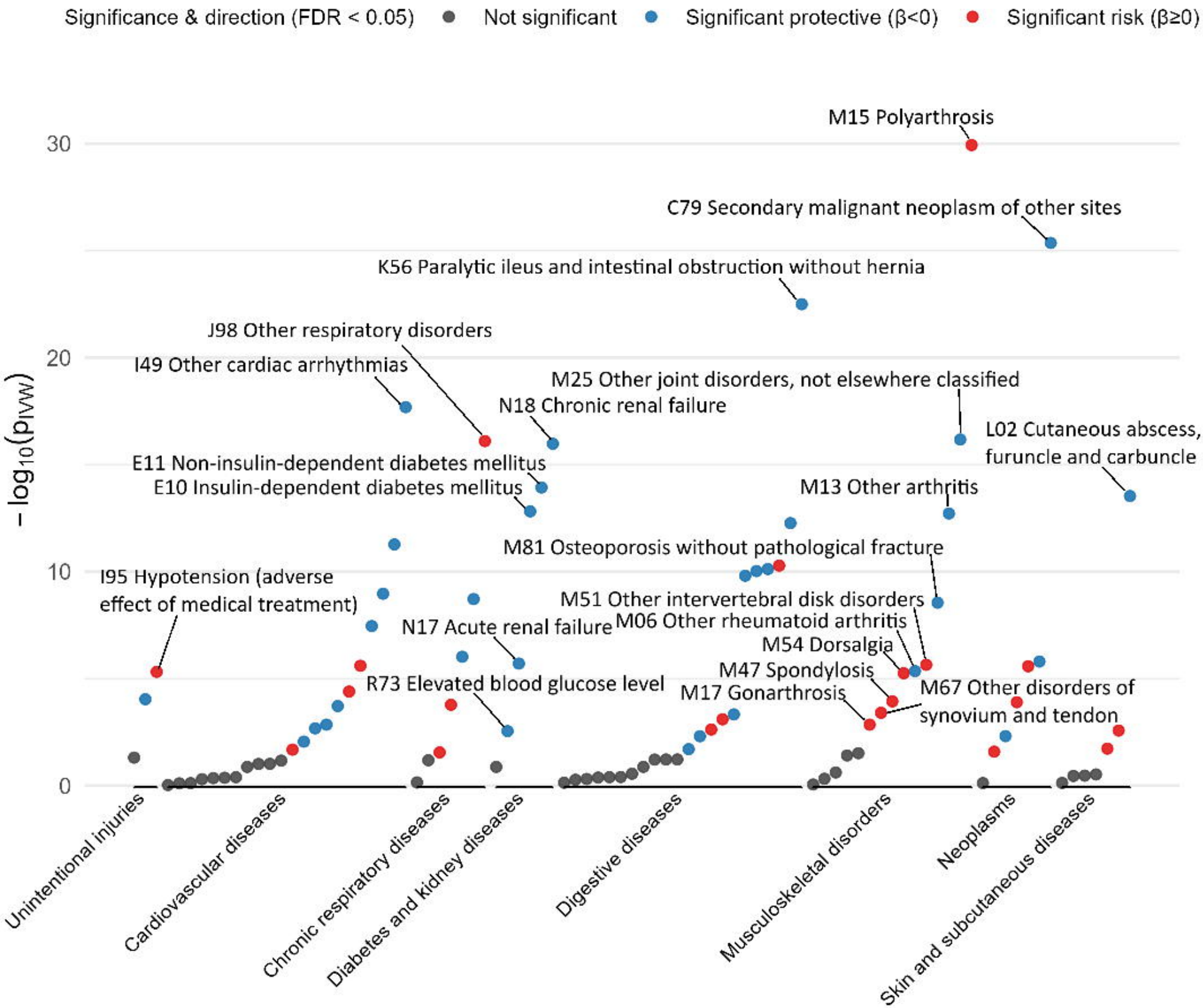
Ethics declarations

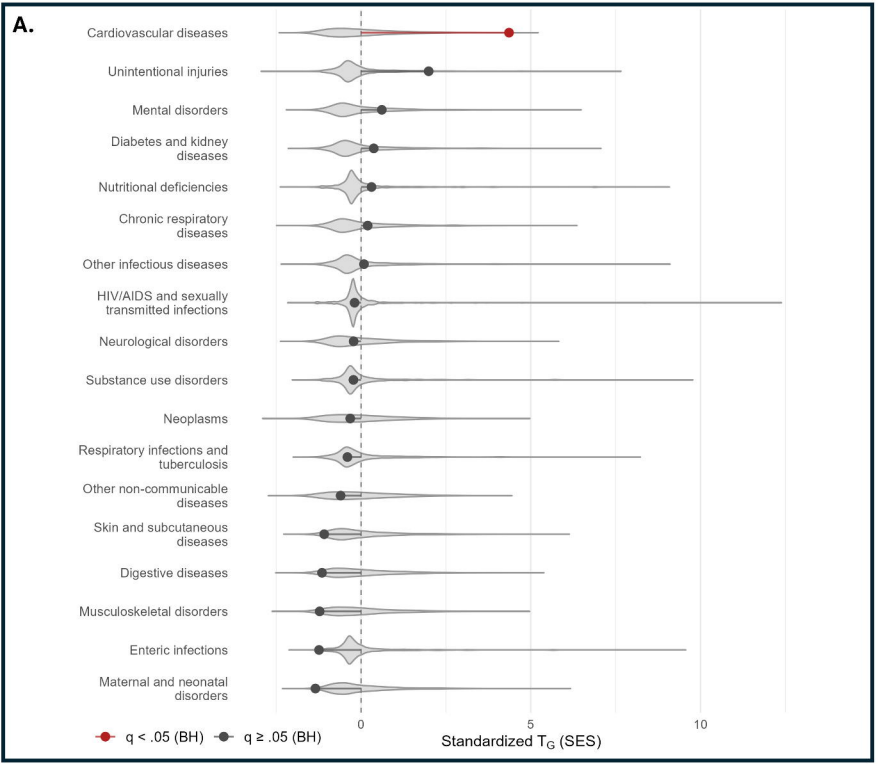
Competing interests

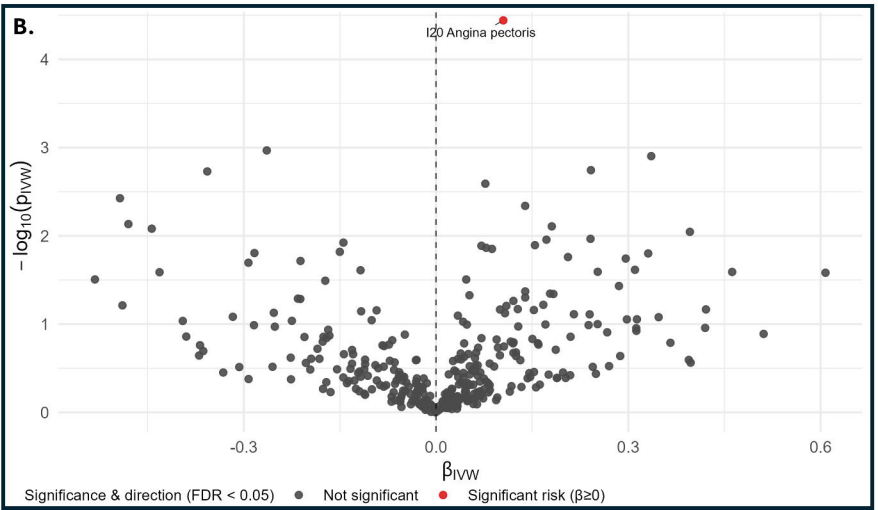
M.C.M. is a Trustee and on the ethics advisory board of the UK Biobank, on the scientific and ethics advisory boards of OFH and Netherlands Lifelines Biobank, and on the data management advisory board of the US Health and Retirement Survey and UK CLS Cohort Studies. R.H.C. is employed by and holds stock options in iuvantium plc, a precision immunology company.

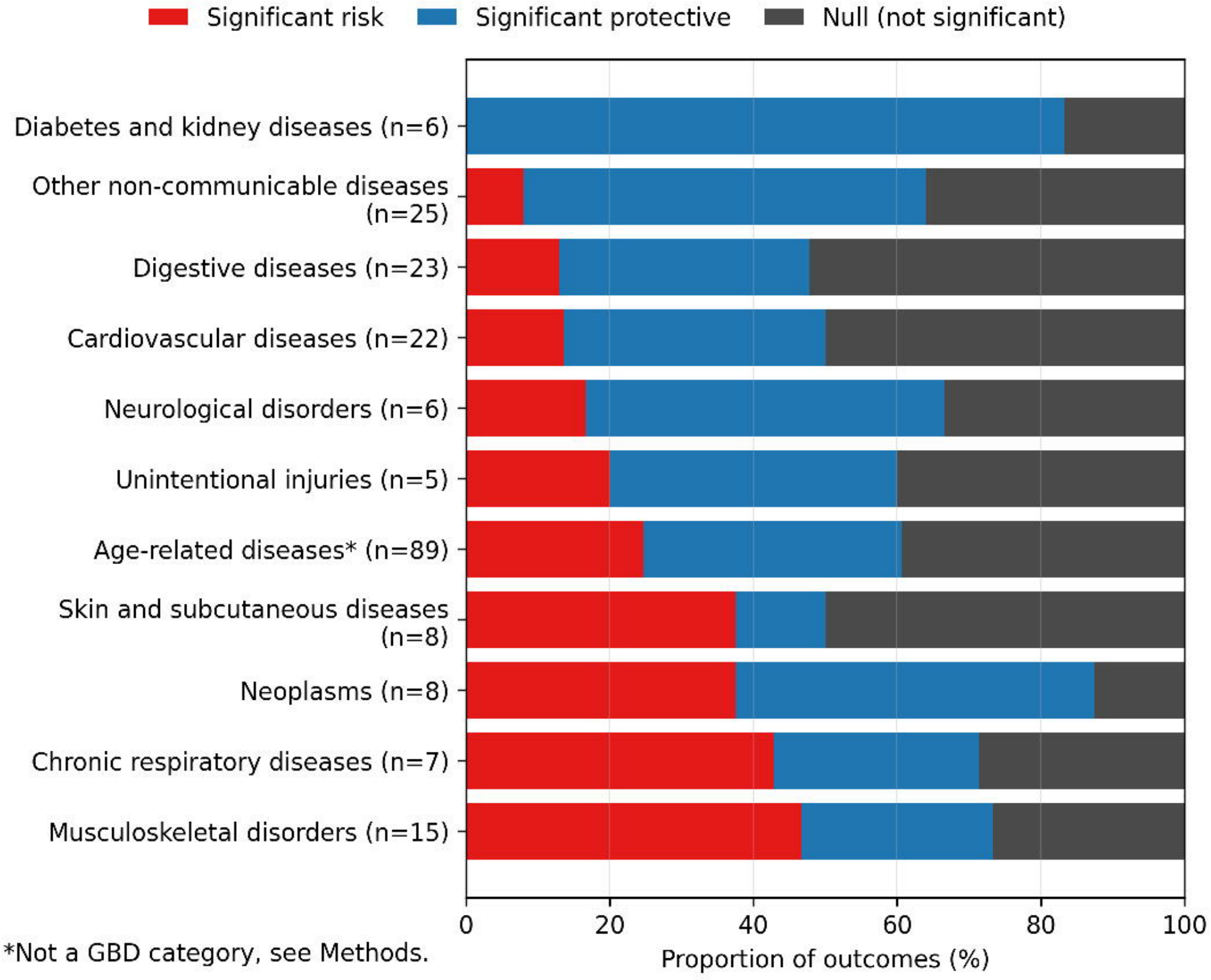

A. Exposure Data
GWAS' of exposure phenotypes
CETP concentration
GLP-1R expression
Control exposures
Instrument selection
Selection
Quality control
B. Outcome Data
PanUKBB GWAS database
396 diseases (EUR + EAS)
Neale Lab GWAS database (UKB)
396 diseases EUR (Male + Female)
C. MR-PheWAS
Confounders of disease 1
Disease 1
Genetic instruments
Exposure
Confounders of disease η
Disease η
Robustness checks
ε non-robust effects
MR results
Disease 1, (β, SE, p-val)
Disease η - ε, (β, SE, p-val)
D. Permuting category-wide effects
MR results
Disease 1, (β, SE, p-val)
Disease η - ε, (β, SE, p-val)
Group by category
Category 1
Category γ
Permute cat. effect scores
$Score(Cat_1)$
$Score(Cat_\gamma)$
Category-wide results
Observed
Permuted


Significance & direction (FDR < 0.05)
Not significant
Significant protective (β<0)
Significant risk (β≥0)
Polyp of female genital tract
Polyarthrosis
Cutaneous abscess, furuncle and carbuncle
Pain associated with micturition
Other female pelvic inflammatory diseases
Irritable bowel syndrome
Pain associated with micturition
Epilepsy
Unspecified fall
Maternal care for known or suspected abnormality of pelvic organs
Chronic sinusitis
Diaphragmatic hernia
Inguinal hernia
Vitamin D deficiency
Labour and delivery complicated by foetal stress
Endometriosis
Osteoporosis without pathological fracture
Migraine
Excessive, frequent and irregular menstruation
Dorsalgia
Urethral Stricture
Menopausal and other perimenopausal disorders
Hypertrophic disorders of skin
40
30
20
10
0
-1.0
-0.5
0.0
0.5
1.0
$-\log_{10}(p_{IVW})$
β_{IVW}

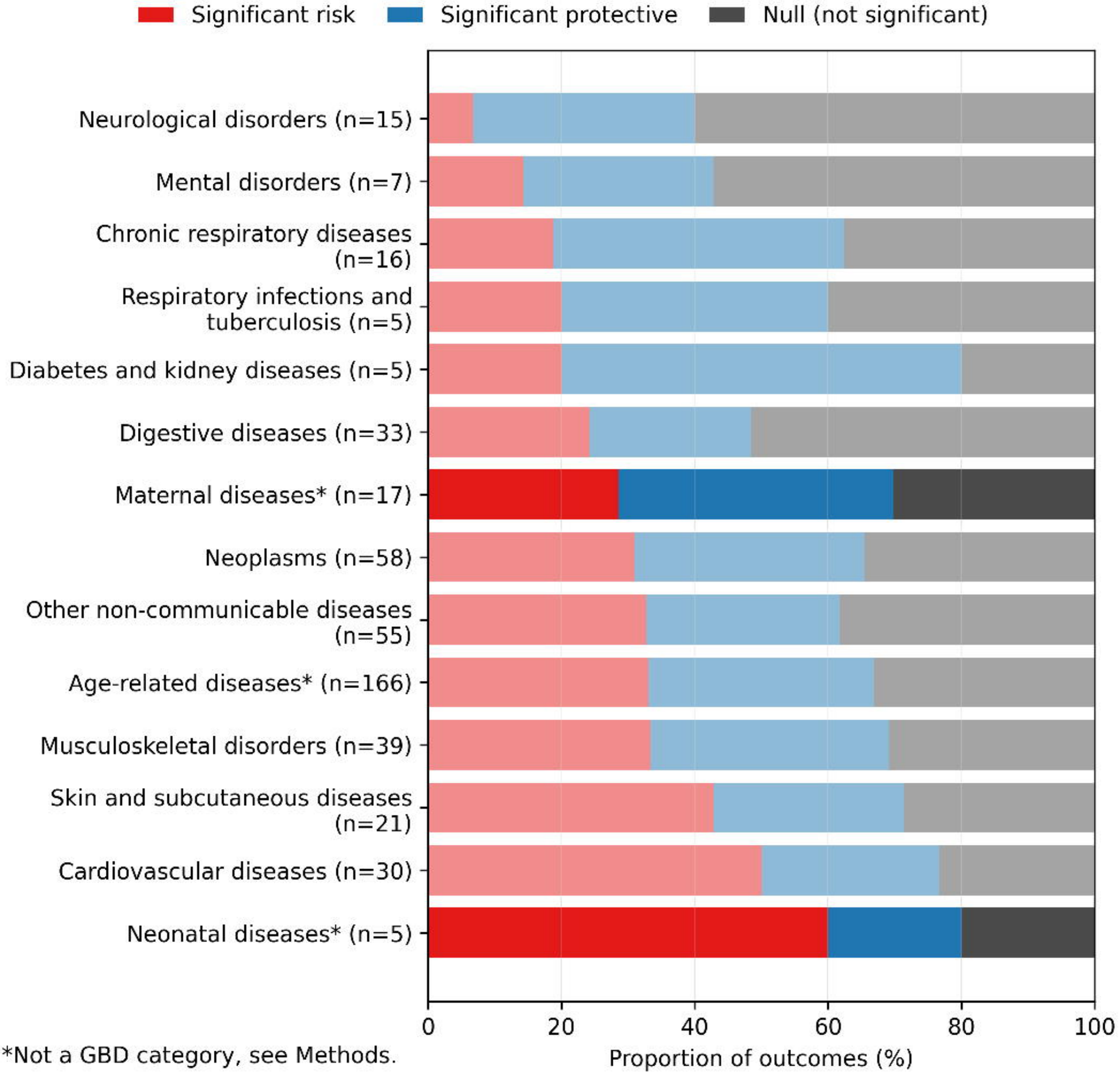

Significance & direction (FDR < 0.05)
Not significant
Significant protective (β<0)
Significant risk (β≥0)
M15 Polyarthrosis
C79 Secondary malignant neoplasm of other sites
K56 Paralytic ileus and intestinal obstruction without hernia
J98 Other respiratory disorders
I49 Other cardiac arrhythmias
M25 Other joint disorders, not elsewhere classified
N18 Chronic renal failure
L02 Cutaneous abscess, furuncle and carbuncle
E11 Non-insulin-dependent diabetes mellitus
E10 Insulin-dependent diabetes mellitus
M13 Other arthritis
M81 Osteoporosis without pathological fracture
I95 Hypotension (adverse effect of medical treatment)
M51 Other intervertebral disk disorders
M06 Other rheumatoid arthritis
N17 Acute renal failure
M54 Dorsalgia
M47 Spondylosis
R73 Elevated blood glucose level
M17 Gonarthrosis
M67 Other disorders of synovium and tendon
30
20
10
0
$-\log_{10}(p_{IVW})$
Unintentional injuries
Cardiovascular diseases
Chronic respiratory diseases
Diabetes and kidney diseases
Digestive diseases
Musculoskeletal disorders
Neoplasms
Skin and subcutaneous diseases


A.
Cardiovascular diseases
Unintentional injuries
Mental disorders
Diabetes and kidney diseases
Nutritional deficiencies
Chronic respiratory diseases
Other infectious diseases
HIV/AIDS and sexually transmitted infections
Neurological disorders
Substance use disorders
Neoplasms
Respiratory infections and tuberculosis
Other non-communicable diseases
Skin and subcutaneous diseases
Digestive diseases
Musculoskeletal disorders
Enteric infections
Maternal and neonatal disorders
0
5
10
q < .05 (BH)
q ≥ .05 (BH)
Standardized T_G (SES)


B.
I20 Angina pectoris
4
3
2
1
0
$-\log_{10}(p_{IVW})$
-0.3
0.0
0.3
0.6
β_{IVW}
Significance & direction (FDR < 0.05)
Not significant
Significant risk (β≥0)



*Not a GBD category, see Methods.


Significant risk
Significant protective
Null (not significant)
Neurological disorders (n=15)
Mental disorders (n=7)
Chronic respiratory diseases (n=16)
Respiratory infections and tuberculosis (n=5)
Diabetes and kidney diseases (n=5)
Digestive diseases (n=33)
Maternal diseases* (n=17)
Neoplasms (n=58)
Other non-communicable diseases (n=55)
Age-related diseases* (n=166)
Musculoskeletal disorders (n=39)
Skin and subcutaneous diseases (n=21)
Cardiovascular diseases (n=30)
Neonatal diseases* (n=5)
0
20
40
60
80
100
Proportion of outcomes (%)
*Not a GBD category, see Methods.