

Assessing the genetic relationship between Gastro-esophageal Reflux Disease and risk of COVID-19 infection

Jue-Sheng Ong*¹, Puya Gharakhani¹, Thomas L. Vaughan², David Whiteman³, Bradley J. Kendall^{4,5} and Stuart MacGregor¹.

Affiliations

¹Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, 300 Herston Road, Herston, QLD 4006, Brisbane, Australia.

²Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA.

³Department of Population Health, QIMR Berghofer Medical Research Institute, 300 Herston Road, Herston, QLD 4006, Brisbane, Australia.

⁴Department of Medicine, The University of Queensland, Herston, QLD 4006, Brisbane, Australia.

⁵Department of Gastroenterology and Hepatology, Princess Alexandra Hospital, Woolloongabba, QLD 4102, Brisbane, Australia.

***Correspondence author: Dr. Jue-Sheng Ong**

E-mail: juesheng.ong@qimrberghofer.edu.au

Address: Statistical Genetics laboratory, Department of Genetics and Computational Biology, QIMR
Berghofer Medical Research Institute.

Tel/Fax: +61 7 3362 3214

Abstract

Background

Symptoms related with Gastro-esophageal reflux disease (GERD) were previously shown to be linked with increased risk for the 2019 coronavirus disease (COVID-19). We aim to interrogate the possibility of a shared genetic basis between GERD and COVID-19 outcomes.

Methods

Using published GWAS data for GERD (78,707 cases; 288,734 controls) and COVID-19 susceptibility (up to 32,494 cases; 1.5 million controls), we examined the genetic relationship between GERD and three COVID-19 outcomes: risk of developing severe COVID-19, COVID-19 hospitalization and overall COVID-19 risk. We estimated the genetic correlation between GERD and COVID-19 outcomes followed by Mendelian randomization (MR) analyses to assess genetic causality. Conditional analyses were conducted to examine whether known COVID-19 risk factors (obesity, smoking, type-II diabetes, coronary artery disease) can explain the relationship between GERD and COVID-19.

Results

We found small to moderate genetic correlations between GERD and COVID-19 outcomes (r_g between 0.06-0.24). MR analyses revealed a OR of 1.15 (95% CI: 0.96-1.39) for severe COVID-19; 1.16 (1.01-1.34) for risk of COVID-19 hospitalization; 1.05 (0.97-1.13) for overall risk of COVID-19 per doubling of odds in developing GERD. The genetic correlation/associations between GERD and COVID-19 showed mild attenuation towards the null when obesity and smoking was adjusted for.

Conclusions

Susceptibility for GERD and risk of COVID-19 hospitalization were genetically correlated, with MR findings supporting a potential causal role between the two. The genetic association between GERD and

COVID-19 was partially attenuated when obesity is accounted for, consistent with obesity being a major risk factor for both diseases.

UNCORRECTED MANUSCRIPT

Introduction

The Coronavirus Disease 2019 (COVID-19) outbreak has affected tens of millions globally, with the COVID-19 related death toll exceeding 1 million in October 2020. The COVID-19 is caused by the newly identified severe acute respiratory syndrome coronavirus (SARS-CoV-2), where the virus enters host cells via the transmembrane serine protease 2 (*TMPRSS2*) and the ACE-II cell receptor (*ACE2*) (1). Human hosts infected by the virus risk developing excessive immune reaction and pneumonia-like symptoms, which can be fatal if left untreated. Whilst promising progress on the COVID-19 vaccines had been made (2), risk factors associated with susceptibility for COVID-19 remain severely understudied.

More recently, symptoms related with gastro-esophageal reflux disease have been reported to be associated with COVID-19 in several epidemiological studies (3, 4). Gastro-esophageal reflux is a chronic disease characterised by the frequent regurgitation of acid arising from the stomach to the esophagus. Prior in-vivo studies have shown evidence for digestive systems to play a role in the pathogenesis of SARS-CoV-2 as the key receptors *ACE2* and *TMPRSS2* were shown to be co-expressed in both the upper epithelial and gland cells along the esophagus and the colon (5). On the other hand, the observational link between GERD and COVID-19 is unsurprising since both diseases share some major risk factors and present common symptoms (**Figure 1**) (6–10). For instance, obesity and smoking are established risk factors for both GERD and COVID-19 (9, 11). However, most of these studies were observational in nature and causality cannot be assumed. Genetic data offers an interesting avenue to validate these associations given the availability of large scale genome-wide association study (GWAS) data on both COVID-19 susceptibility and gastro-esophageal reflux disease (GERD). If there is a direct causal effect between GERD and COVID-19 we would expect that genetic variants which increase GERD risk will also increase COVID-19 risk. Genetic-derived findings can hence provide a complementary perspective into the biological mechanisms linking both diseases (12).

In this study, we interrogate the possibility of a shared genetic architecture between COVID-19 and GERD in two ways. We first estimated the genetic correlation between GERD diagnosis and several COVID-19 related outcomes, followed by a genetic instrumental variable analysis to evaluate whether genetically predicted GERD diagnosis is linked with risk of COVID-19 infection. We further implemented a conditional analysis on common risk factors between GERD and COVID-19 to evaluate potential mediation mechanisms.

Results

Genetic correlation between GERD and COVID-19 outcomes

The sample size for each COVID-19 outcome was shown in **Table 1**. Each of the examined COVID-19 outcomes showed moderate evidence of being heritable. Based on an estimated 8% population-wide COVID-19 infection rate in Europe with ~10% of cases developing severe symptoms, the estimated SNP-heritability under the liability scale was $h^2=0.206$ (se 0.038) for risk of developing severe COVID-19, $h^2=0.074$ (0.015) for risk of COVID-19 hospitalization and $h^2=0.01$ (0.004) for overall risk of developing COVID-19. The estimated genetic correlations between GERD and risk of severe COVID-19 was $r_g=0.145$ (se 0.050, $p=3.9e-3$). Similarly the correlation with risk of COVID-19 hospitalization was $r_g=0.239$ (se 0.053, $p=7.88e-06$). There was no evidence for a genetic correlation between GERD and overall risk of COVID-19 infection [$r_g=0.056$ (se 0.078, $p=0.47$)].

Estimate of genetic causal effect for GERD on COVID-19 outcomes

88 independent GERD-associated SNPs (GERD association p -value $< 5e-8$) were used as genetic proxies to evaluate whether higher genetic predisposition towards GERD influenced risk of COVID-19 infection (**Supplementary Table 1**). Power calculations indicated adequate statistical power for MR to detect

moderate effect sizes (**Supplementary Table 3**). Based on the IVW estimate, genetically doubling of odds of being diagnosed with GERD is associated with risk of severe COVID-19 (OR 1.15 [0.96 - 1.39]) and risk of COVID-19 hospitalisation (OR= 1.16 [1.01 - 1.34]), but not for the overall risk of COVID-19 infection (OR 1.05 [0.97 - 1.13]). Findings from alternative MR sensitivity models were not meaningfully different from the IVW estimates with largely overlapping 95% confidence intervals (**Figure 2**; complete summary in **Supplementary Table 4**). Moreover, the MR-Egger intercept estimates were very small (<0.01) with 95% CIs overlapping the null, indicating very weak evidence of directional pleiotropy (**Supplementary Table 5**). Similarly, there is limited evidence for heterogeneity among the SNP effect sizes as indicated by the Cochran Q statistics and the MR-PRESSO global heterogeneity tests (**Supplementary Table 6 and 7**). The MR scatter plots illustrating the fitted slope (causal effect) under various MR models are presented in **Supplementary Figure 1-3**. Finally, we found limited evidence for a bi-directional association from COVID-19 outcomes on GERD susceptibility (**Supplementary Table 8-9**).

The genetic correlation between GERD and COVID-19 outcomes after conditioning on known COVID-19 risk factors

The correlation matrix in **Figure 3(a)** reveals the estimated genetic correlation between GERD, COVID-19 risk factors and COVID-19. Each of these risk factors showed strong evidence of genetic overlap with GERD (r_g between 0.2 and 0.4). We also found moderate genetic correlations for BMI, CAD, cigarette smoked per day, risk of Type-2 diabetes with risk of COVID-19 hospitalization (COVID-A2) and risk of severe COVID-19 infection (COVID-B2) (**Figure 3(a)**). However, the estimated correlation of these risk factors with overall COVID-19 susceptibility (COVID-C2) is much weaker. We hence re-estimated the genetic correlation between GERD and COVID-19 outcomes after adjusting for these risk factors. Apart from obesity-adjusted COVID-19 and smoking-adjusted COVID-19 where our revised r_g estimate between GERD and COVID-19 outcomes showed evidence of attenuation towards the null, findings for other covariate-adjusted models remain widely consistent with the original unadjusted r_g estimates (see

Figure 3(b)).

Estimate of genetic causal effect for GERD on COVID-19 outcomes after conditioning on known COVID-19 risk factors

To assess our genetic instruments for potential horizontal pleiotropic association via known COVID-19 risk factors, we evaluated the association between each of the 88 variants and the genetic effect sizes on BMI, smoking phenotypes, risk of T2D and risk of CAD using publicly available summary statistics (13–16). In our analysis, at least 50 SNPs showed evidence of pleiotropy (i.e. absolute Z-scores > 5) on one or more of the aforementioned COVID-19 risk factors. The heatmap in **Figure 4** illustrates the distribution of Z-scores across these risk factors, indicating pervasive pleiotropic association between GERD instruments and established COVID-19 risk factors.

To control for potential genetic confounding between GERD and COVID-19 outcomes, we performed a conditional analysis adjusting for established COVID-19 risk factors in the COVID-19 GWAS summary statistics using GCTA-mtCOJO. Results for the conditional MR analysis are shown in **Figure 5**, showing minimal evidence of mediation through known COVID-19 risk factors, except for BMI - where the resultant MR estimate showed slight attenuation of the point estimate on COVID-19 hospitalization and risk of severe COVID-19 infection. However, the genetic association between GERD and overall risk of COVID-19 remains broadly unaffected in the conditional analyses. The pattern of associations under each adjusted mtCOJO model were broadly similar using alternative MR estimators, with concordant effect size estimates and direction of association (**Supplementary Table 10-13**). Finally, excluding the variant rs9940128 in the FTO gene in our leave-one-SNP-out analyses did not alter our original MR findings (See **Supplementary Figure 4-6**).

Discussion

Our genetic analyses reveal strong genetic evidence supportive of a potentially causal relationship between GERD and COVID-19 susceptibility. Genetically higher odds of having GERD was associated with ~15% increase in risk of severe COVID-19 and COVID-19 hospitalization, consistent with retrospective observational findings(3, 17). Adjustment for known COVID-19 risk factors, most notably obesity, partially weakens the association between GERD and COVID-19 although with 95% confidence intervals that overlap the unadjusted estimates.

Drawing a direct causal inference between GERD and COVID-19 can be difficult, as both diseases share common risk factors such as smoking, diabetes and obesity. For instance, obesity was previously shown to be causally associated with both GERD and COVID-19 outcomes in earlier MR findings.(18, 19) Genetic correlation analyses further revealed strong genetic overlap between these risk factors with GERD and COVID-19 risks. We controlled for these risk factors via mediation analyses, which typically reduced (but did not completely eliminate) the genetic correlations between GERD and COVID-19 outcomes (**Figure 3**). Similarly, our MR estimate on COVID-19 outcomes after adjusting for these risk factors showed partial attenuation towards the null, but remains widely consistent with a moderate adverse effect (**Figure 5**).

Our analysis primarily focused on GERD susceptibility as opposed to previous findings that evaluated the use of PPIs specifically.(3) We showed in our previous study that GERD diagnosis attained through self-report and medication use (i.e. GERD cases inferred through individuals using GERD-related medications such as PPIs) were genetically very similar (20) and hence our genetics based approach cannot reliably separate the cause and effect of PPI use. That is, our data cannot clarify whether increased risk from COVID-19 relates to GERD and its complications per se, or to associated treatments such as PPIs. Most

of our GERD instruments showed strong evidence of being replicated in an independent cohort (21), reducing chances of there being winner's curse bias (22). Our analysis also tried to control for genetic confounding arising from a comorbidity with obesity, smoking behaviour, diabetes and risk of cardiovascular complications - upon which the association with COVID-19 was unchanged apart from the model adjusting for obesity. While these findings are consistent with the role of PPIs promoting greater risk for COVID-19 through a mechanism independent of obesity and/or detection bias (5), further validations are required.

Several limitations ought to be considered. Firstly, our study focused chiefly on evaluating COVID-19 susceptibility and severity/hospitalization relative to the non-infected population; while GWAS data was available for severe vs non-severe COVID-19 within COVID-19 cases, the sample sizes were more limited and our power was limited. To obtain heritability estimates in the genetic liability scale one is required to provide an approximate prevalence (heritability does not change dramatically if this is misspecified), we tentatively assume a population-wide COVID-19 prevalence of 10% based on availability of data (23). We note however that the mis-specification of the prevalence has only negligible impact on the genetic correlation estimates (24). Instruments used for GERD susceptibility were derived from a symptom-based GERD definition (20). Even though we previously showed that the genetic architecture of our broad GERD definition is similar to those obtained via more robust clinical diagnosis, these instruments are highly heterogeneity in nature. We attempt to minimize biases arising from instrument misspecification by adopting MR models such as median and mode-based estimators, though these models are typically less powered. Whilst sample size for the analyses on overall COVID-19 infection outcome was the largest, we did not observe strong evidence for an association with GERD. One possible explanation is that the overall COVID-19 phenotype might be heterogeneous, which is reflected in the relatively lower estimated heritability. We performed post-hoc analyses using an earlier dataset (based on the COVID-19 HGI Release 3 in September 2020) before the availability of vaccines across US and Europe, in an attempt to capture a better phenotype. We did observe evidence for an association with

GERD, though the estimated effect sizes for overall risk of COVID-19 were still lower than those derived for COVID-19 severity and risk of COVID-19 hospitalization (**Supplementary Table 14**). Lastly, our analysis found very little evidence for reverse causality from COVID-19 susceptibility to GERD, though our power was limited as there are only a handful of COVID-19 associated variants.

Due to the nature of our study, we were also unable to directly compare our findings to those evaluating duration of PPIs used on COVID-19 severity (3). Genetically-derived findings were conceivably less biased by confounding and reverse causality, however issues of residual pleiotropy cannot be completely ruled out. Moreover, precise estimation of MR effect sizes between a binary-binary relationship such as those between GERD and COVID-19 susceptibility is difficult as the homogeneity and monotonicity assumptions (25) might not be satisfied. Whilst we present genetic evidence supporting a causal relationship between GERD and COVID-19, the precise magnitude of association cannot be reliably estimated and compared with observational findings. The derived instrumental variable estimates will hence reflect the average causal effect of GERD on COVID-19 susceptibility among genetically-predicted compliers (25) (i.e., GERD patients with genes that predict GERD diagnosis). Finally, although individuals included in the COVID-19 GWAS were predominantly of European ancestry, some of the smaller studies included participants from other/mixed ancestries.

In summary, genetic findings support a common genetic basis for both GERD and COVID-19 susceptibility, showing strong genetic correlation between GERD susceptibility and COVID-19 outcomes. Genetically predicted higher odds of being diagnosed with GERD increased risk of developing severe COVID-19 and COVID-19 hospitalization, as well as risk of being hospitalized due to COVID-19. Our conditional analyses show that the relationship between GERD and COVID-19 might be partly explained by shared comorbidities such as obesity, though the attenuation of effect sizes was only partial. Future studies are warranted to explore the biological mechanisms linking these risk factors, GERD and COVID-19.

Materials and Methods

Data source for GERD diagnosis

We obtained the largest GWAS summary data for GERD susceptibility from a study of 71,522 GERD cases and 261,039 controls of European ancestry from the UK Biobank and the Australian QSKIN cohorts (20). Both UK Biobank and QSKIN are population-based cohorts with predominantly middle aged participants.(26, 27) Details on the genotyping, genetic-QC and imputation for these studies had been previously described (26, 27). GERD cases in both studies were derived through a combination of self-report, ICD-10 codes (K21), GERD-related medication or heartburn status (QSKIN only), though our previous findings have found strong genetic correlations between broad and clinical GERD definitions (20). For the instrumental variable analyses, 88 GERD-associated variants were selected based on SNPs which reached genome-wide significance in a multi-trait GERD GWAS meta-analysis (21).

Data source for COVID-19 diagnosis

GWAS summary statistics for COVID-19 infection were obtained from the COVID-19 Host Genetic Initiative website (<https://www.covid19hg.org/results/>) (28, 29). For the genetic analyses, we used the genetic summary statistics for each COVID-19 trait excluding both the participants from 23andMe and UK Biobank (January 2021 data release) to prevent bias for the genetic causal inference analyses. Information on the demographics of the individual studies involved in the COVID-19 host genetic initiative are available in the COVID-19 HGI partners interactive website (<https://www.covid19hg.org/partners/>). In total, GWAS data from 26 individual studies were meta-analysed to generate the combined GWAS summary statistics including up to 32,494 cases and more than 1 million controls. Information on the analysis protocol and how the genetic data from each study was

quality controlled is provided in **Supplementary Notes**. We manually mapped the chromosome:basepair variant notation in the GWAS summary statistics back to RSID (under build 37) using the 1000G European reference panel for all our analyses. The three primary COVID-19 related outcomes evaluated in this study are provided below.

Exposure

Genetically predicted susceptibility towards GERD, derived through combination of self-report status, ICD-10 codes, hospital records and use of GERD-related medications such as omeprazole. Note that we have previously shown that the GERD defined through these definitions were highly consistent, i.e. the genetic correlations between self-report, medication-inferred and clinical GERD definitions were very close to one (20).

Primary outcome

Genetic estimates on risk of severe COVID-19 infection, risk of COVID-19 hospitalization and overall risk of developing COVID-19. See **Supplementary Notes** for the precise case definition for each COVID-19 outcome. Sample sizes for each COVID-19 outcome analysis was provided in **Table 1**.

Statistical Approach

Genetic correlation analysis: The LD-score regression (30) method implemented through the ldsc python software (available at <https://github.com/bulik/ldsc>) was used to estimate the pairwise genetic correlation estimate between GERD and COVID-19 susceptibility outcomes, which only requires summary level genetic data. We first estimated the SNP-heritability (i.e. proportion of phenotypic variance captured by common genetic variation) for each COVID-19 outcome via LD-score regression and converted the observed heritability estimate into the genetic liability scale (--h2 option; see Supplementary Notes for technical details) (31). The pairwise genetic correlation between GERD and COVID-19 outcomes were

estimated via bi-variate LD-score regression (--rg option). For all analysis, the European 1000 HGP reference panel was used to obtain the LD-score derived for each SNP among European populations. The genetic correlation estimate, r_g describes the genome-wide correlation of genetic effect sizes between two traits (30).

Genetic instrumental variable analysis (Mendelian randomization, MR): We obtained 88 genome wide significant SNP-markers associated with GERD susceptibility from our previous work (21). The individual SNP effect size and the allele frequencies for each of the 88 GERD-associated SNPs are provided in **Supplementary Table 1**. Power calculations using the mRnd online interface (accessible here: <https://cnsgenomics.com/shiny/mRnd/>) were performed to assess power for MR to detect clinically relevant endpoints (32). To avoid over-inflation of effect sizes from the multi-trait model, we adopted genetic effect size estimates for the 88 SNPs from the published (univariate) GERD GWAS (20).

Harmonisation of effect alleles were performed to remove variants with palindromic alleles that cannot be inferred via minor allele frequency ($MAF > 0.4$). We then fitted an inverse variance weighted (IVW) regression model for the genetic effect of GERD against the genetic effect of risk on COVID-19 outcomes, weighted by the inverse variance of the genetic effect on GERD. We repeated our MR analysis using other MR estimators (MR-Egger (33), MR-Weighted Mode (34), MR-Weighted median (34), MR-PRESSO (35), robust and penalized IVW (36) methods) to triangulate evidence for genetic causality under weak violation of key MR assumptions. Technical details of these alternative models have been previously described (35, 37, 38). To aid the interpretation of our association estimates, we scaled our results to reflect the $\log(OR)$ on COVID-19 outcomes per doubling of odds on developing GERD. This was done by multiplying the resultant genetically derived $\log(OR)$ estimate from the IVW model by $\log(2) = 0.693$ (25). We also checked for reverse causality by performing a MR analysis for COVID-19 outcomes on GERD susceptibility (see **Supplementary Notes**). All statistical analyses were performed using the statistical software R v4.0.3. Genetic instrumental variable analyses were conducted using the TwoSampleMR and MendelianRandomization R packages (39, 40).

Sensitivity analyses: Conditional-analyses on common risk factors between GERD and COVID-19 outcomes

To evaluate potential horizontal pleiotropy between GERD and COVID-19 susceptibility, we first estimated the genetic effect size for the 88 GERD SNPs on a series of established risk factors(9) on COVID-19 susceptibility. These risk factors include: body mass index (BMI), type-2 diabetes (T2D), smoking traits (cigarettes per day, smoking initiation) and coronary artery disease (CAD). The largest and most recent GWAS summary statistics for these risk factors were obtained from publicly available repositories (13–16). Details of the data source and curation of GWAS datasets for these phenotypes available in **Supplementary Note** (sources in **Supplementary Table 2**). Genetic effect size estimates for each risk factor are aligned based on the GERD-increasing allele.

We then explored whether the pattern of genetic correlation/causality changed when risk factors such as body mass index and smoking were adjusted for in the model. We implemented the GCTA-mtCOJO (mtCOJO) framework (41) to remove the genetic effect sizes of established COVID-19 risk factors from the COVID-19 GWAS. In brief, the mtCOJO model is a summary-level based approach which estimates the genetic estimate for a trait of interest by conditioning out the causal effect of potential mediating factors on the outcome of interest (estimated using the GCTA-GSMR model). The resultant conditional estimates had also been previously shown to be robust against collider bias (see for instance (42) on collider bias generated via adjusting for heritable covariates). Further technical details on the mtCOJO model have been previously described (the mtCOJO framework is available within the GCTA software available here <https://cnsgenomics.com/software/gcta>) (41).

We repeated our MR analyses between GERD and COVID-19 outcomes after adjusting for each COVID-19 risk factor in turn. This was done by regressing the GERD instrumental variables against the genetic estimate on COVID-19 susceptibility after conditioning on the aforementioned risk factors via mtCOJO.

Similarly, MR estimates of the covariate-adjusted model derived from alternative MR techniques were used to validate robustness of the IVW findings. Given that obesity is potentially a major risk factor for both GERD and COVID-19 susceptibility, we performed a leave-one-SNP-out MR analysis to assess whether the genetic association between these traits are potentially driven by specific variants, such as those in BMI-associated genes (e.g. the *FTO* gene).

UNCORRECTED MANUSCRIPT

Acknowledgement

QSkin was funded by grants from the National Health and Medical Research Council of Australia (NHMRC) APP1073898, APP1063061. SM acknowledges support from an NHMRC fellowship. S.M. acknowledges funding from the NHMRC grant 1123248. We thank all the genetics consortia and the COVID-19 Host Genetics Initiative for making COVID-19 related summary statistics publicly accessible for our analysis. We thank research staff and participants from the Australian QSKIN study (<https://www.qimrberghofer.edu.au/study/qskin/>) and the UK Biobank (<https://www.ukbiobank.ac.uk/>) for contributing to the GERD GWAS summary data. We acknowledge the use of UK Biobank data under application number 25331. We finally acknowledge the contribution of the participating studies in the COVID-19 Host Genetics Initiative listed at <https://www.covid19hg.org/acknowledgements/>. This study primarily uses GWAS summary data and does not involve the use of human test subjects. This research was approved by the QIMR Berghofer's Human Research Ethics Committee under project ID 3501.

Authorship contribution

JSO has full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: JSO and SM. Statistical analysis and interpretation of data: JSO, SM, PG, DW, BJK, TLV. Acquisition of data and obtained funding: JSO, SM, PG, DW. Data preparation: JSO, PG. Drafting of the manuscript: JSO and SM. Critical review of the manuscript for important intellectual content: JSO, SM, PG, DW, BJK, TLV. Study supervision: SM. All authors read and approved the final version of the manuscript for submission.

Role of sponsors

The funding bodies for our study had no role in the design or conduct of the study; collection, management, analysis and/or interpretation of the data; preparation, review or approval of the manuscript or the decision to submit the manuscript for publication.

Data availability

The summary statistics for the GERD GWAS meta-analysis can be downloaded from here (<https://doi.org/10.6084/m9.figshare.8986589>). GWAS summary statistics for the COVID-19 outcomes provided by the COVID-19 HGI (Release 5) can be accessed here (<https://www.covid19hg.org/results/r5/>). The individual GWAS summary statistics for traits used in the mediation analyses are publicly available with data sources provided in Supplementary Notes.

Conflict of interests

All authors declare no conflict of interests.

Reference

1. Yuki,K., Fujiogi,M. and Koutsogiannaki,S. (2020) COVID-19 pathophysiology: A review. *Clin. Immunol.*, **215**, 108427.
2. Tregoning,J.S., Brown,E.S., Cheeseman,H.M., Flight,K.E., Higham,S.L., Lemm,N.-M., Pierce,B.F., Stirling,D.C., Wang,Z. and Pollock,K.M. (2020) Vaccines for COVID-19. *Clin. Exp. Immunol.*, **202**, 162–192.
3. Almario,C.V., Chey,W.D. and Spiegel,B.M.R. (2020) Increased Risk of COVID-19 Among Users of Proton Pump Inhibitors. *Am. J. Gastroenterol.*, **115**, 1707–1715.
4. Charpiat,B., Bleyzac,N. and Tod,M. (2020) Proton Pump Inhibitors are Risk Factors for Viral Infections: Even for COVID-19? *Clin. Drug Investig.*, **40**, 897–899.
5. Ng,S.C. and Tilg,H. (2020) COVID-19 and the gastrointestinal tract: more than meets the eye. *Gut*, **69**, 973–974.
6. Grant,M.C., Geoghegan,L., Arbyn,M., Mohammed,Z., McGuinness,L., Clarke,E.L. and Wade,R.G. (2020) The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries. *PLoS One*, **15**, e0234765.
7. Kellerman,R. and Kintanar,T. (2017) Gastroesophageal Reflux Disease. *Prim. Care*, **44**, 561–573.
8. Du,J., Liu,J., Zhang,H., Yu,C.-H. and Li,Y.-M. (2007) Risk factors for gastroesophageal reflux disease, reflux esophagitis and non-erosive reflux disease among Chinese patients undergoing upper gastrointestinal endoscopic examination. *World J. Gastroenterol.*, **13**, 6009–6015.
9. Wolff,D., Nee,S., Hickey,N.S. and Marscholke,M. (2021) Risk factors for Covid-19 severity and fatality: a structured literature review. *Infection*, **49**, 15–28.
10. Clarrett,D.M. and Hachem,C. (2018) Gastroesophageal Reflux Disease (GERD). *Mo. Med.*, **115**, 214–218.
11. Argyrou,A., Legaki,E., Koutserimpas,C., Gazouli,M., Papaconstantinou,I., Gkiokas,G. and Karamanolis,G. (2018) Risk factors for gastroesophageal reflux disease and analysis of genetic contributors. *World J Clin Cases*, **6**, 176–182.
12. Evans,D.M. and Davey Smith,G. (2015) Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality. *Annu. Rev. Genomics Hum. Genet.*, **16**, 327–350.
13. Yengo,L., Sidorenko,J., Kemper,K.E., Zheng,Z., Wood,A.R., Weedon,M.N., Frayling,T.M., Hirschhorn,J., Yang,J., Visscher,P.M., *et al.* (2018) Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.*, **27**, 3641–3649.
14. Xue,A., Wu,Y., Zhu,Z., Zhang,F., Kemper,K.E., Zheng,Z., Yengo,L., Lloyd-Jones,L.R., Sidorenko,J., Wu,Y., *et al.* (2018) Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.*, **9**, 2941.

15. Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D.M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., *et al.* (2019) Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.*, **51**, 237–244.
16. Nelson, C.P., Goel, A., Butterworth, A.S., Kanoni, S., Webb, T.R., Marouli, E., Zeng, L., Ntalla, I., Lai, F.Y., Hopewell, J.C., *et al.* (2017) Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.*, **49**, 1385–1391.
17. Lee, S.W., Ha, E.K., Yeniova, A.Ö., Moon, S.Y., Kim, S.Y., Koh, H.Y., Yang, J.M., Jeong, S.J., Moon, S.J., Cho, J.Y., *et al.* (2021) Severe clinical outcomes of COVID-19 associated with proton pump inhibitors: a nationwide cohort study with propensity score matching. *Gut*, **70**, 76–84.
18. Aung, N., Khanji, M.Y., Munroe, P.B. and Petersen, S.E. (2020) Causal Inference for Genetic Obesity, Cardiometabolic Profile and COVID-19 Susceptibility: A Mendelian Randomization Study. *Front. Genet.*, **11**, 586308.
19. Green, H.D., Beaumont, R.N., Wood, A.R., Hamilton, B., Jones, S.E., Goodhand, J.R., Kennedy, N.A., Ahmad, T., Yaghootkar, H., Weedon, M.N., *et al.* (2020) Genetic evidence that higher central adiposity causes gastro-oesophageal reflux disease: a Mendelian randomization study. *Int. J. Epidemiol.*, **49**, 1270–1281.
20. An, J., Gharahkhani, P., Law, M.H., Ong, J.-S., Han, X., Olsen, C.M., Neale, R.E., Lai, J., Vaughan, T.L., Gockel, I., *et al.* (2019) Gastroesophageal reflux GWAS identifies risk loci that also associate with subsequent severe esophageal diseases. *Nat. Commun.*, **10**, 4219.
21. Ong, J.-S., An, J., Han, X., Law, M.H., Nandakumar, P., 23andMe Research team, Esophageal cancer consortium, Schumacher, J., Gockel, I., Bohmer, A., *et al.* (2021) Multitrait genetic association analysis identifies 50 new risk loci for gastro-oesophageal reflux, seven new loci for Barrett's oesophagus and provides insights into clinical heterogeneity in reflux diagnosis. *Gut*, 10.1136/gutjnl-2020-323906.
22. Palmer, C. and Pe'er, I. (2017) Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. *PLoS Genet.*, **13**, e1006916.
23. Meyerowitz-Katz, G. and Merone, L. (2020) A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. *Int. J. Infect. Dis.*, **101**, 138–148.
24. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M. and Wray, N.R. (2012) Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, **28**, 2540–2542.
25. Burgess, S. and Labrecque, J.A. (2018) Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur. J. Epidemiol.*, **33**, 947–952.
26. Olsen, C.M., Green, A.C., Neale, R.E., Webb, P.M., Cicero, R.A., Jackman, L.M., O'Brien, S.M., Perry, S.L., Ranieri, B.A., Whiteman, D.C., *et al.* (2012) Cohort profile: the QSkin Sun and Health Study. *Int. J. Epidemiol.*, **41**, 929–929i.
27. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., *et al.* (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature*, **562**, 203–209.

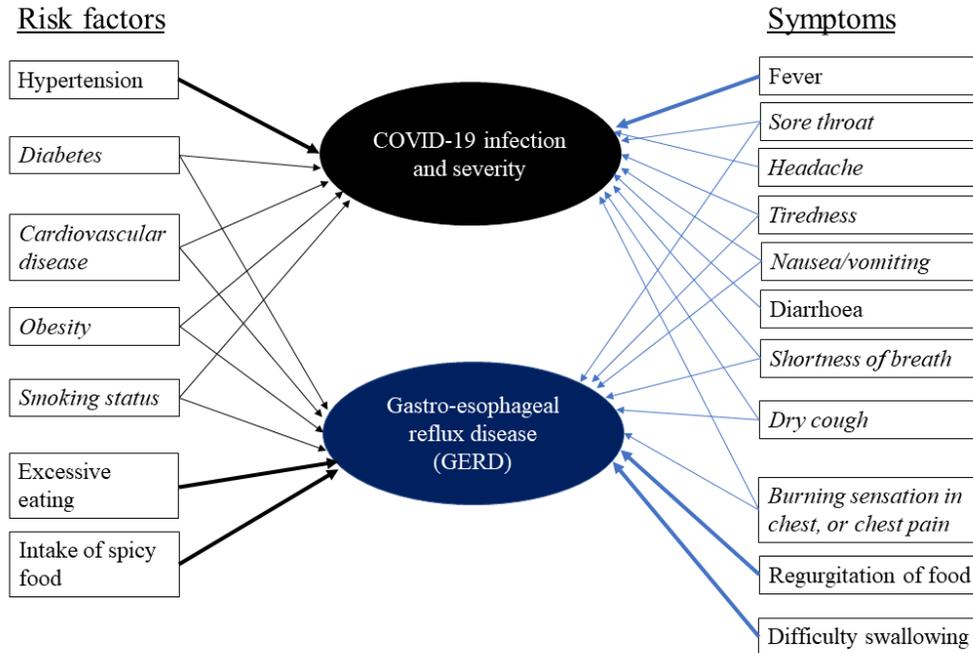
28. Severe Covid-19 GWAS Group, Ellinghaus,D., Degenhardt,F., Bujanda,L., Buti,M., Albillos,A., Invernizzi,P., Fernández,J., Prati,D., Baselli,G., *et al.* (2020) Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N. Engl. J. Med.*, **383**, 1522–1534.
29. COVID-19 Host Genetics Initiative (2020) The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet.*, **28**, 715–718.
30. Bulik-Sullivan,B.K., Loh,P.-R., Finucane,H.K., Ripke,S., Yang,J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson,N., Daly,M.J., Price,A.L. and Neale,B.M. (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.*, **47**, 291–295.
31. Lee,S.H., Wray,N.R., Goddard,M.E. and Visscher,P.M. (2011) Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.*, **88**, 294–305.
32. Brion,M.-J.A., Shakhbazov,K. and Visscher,P.M. (2013) Calculating statistical power in Mendelian randomization studies. *Int. J. Epidemiol.*, **42**, 1497–1501.
33. Bowden,J., Davey Smith,G. and Burgess,S. (2015) Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.*, **44**, 512–525.
34. Hartwig,F.P., Davey Smith,G. and Bowden,J. (2017) Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.*, **46**, 1985–1998.
35. Verbanck,M., Chen,C.-Y., Neale,B. and Do,R. (2018) Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.*, **50**, 693–698.
36. Rees,J.M.B., Wood,A.M., Dudbridge,F. and Burgess,S. (2019) Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates. *PLoS One*, **14**, e0222362.
37. Dimou,N.L. and Tsilidis,K.K. (2018) A Primer in Mendelian Randomization Methodology with a Focus on Utilizing Published Summary Association Data. *Methods Mol. Biol.*, **1793**, 211–230.
38. Hemani,G., Bowden,J. and Davey Smith,G. (2018) Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.*, **27**, R195–R208.
39. Hemani,G., Zheng,J., Elsworth,B., Wade,K.H., Haberland,V., Baird,D., Laurin,C., Burgess,S., Bowden,J., Langdon,R., *et al.* (2018) The MR-Base platform supports systematic causal inference across the human phenome. *Elife*, **7**.
40. Yavorska,O.O. and Burgess,S. (2017) MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.*, **46**, 1734–1739.
41. Zhu,Z., Zheng,Z., Zhang,F., Wu,Y., Trzaskowski,M., Maier,R., Robinson,M.R., McGrath,J.J., Visscher,P.M., Wray,N.R., *et al.* (2018) Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.*, **9**, 224.
42. Aschard,H., Vilhjálmsson,B.J., Joshi,A.D., Price,A.L. and Kraft,P. (2015) Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am. J. Hum. Genet.*, **96**, 329–339.

Table 1. Description of GWAS studies for GERD and COVID-19 outcome phenotypes involved in the present analysis

Trait	Phenotype definition	PMID	Total sample size	Ncases; ncontrols	Additional information	Studies involved
<i>Main exposure</i>						
GERD	cases inferred through self-reported GERD symptoms, use of GERD medication and ICD-10 codes	31527586	367,441	78,707; 288,734	Meta-analysis of GERD GWAS from UK Biobank and the Australian QSKIN stud. See An et al. for more information.	UK Biobank, QSKIN
<i>COVID-19 outcomes (based on COVID-19 Host Genetic Initiative phenotype definitions)</i>						
COVID-A2	Very severe respiratory confirmed COVID-19 cases vs. controls	32404885	1,059,456	4,792; 1,054,664	Please visit https://www.covid19hg.org/data-sharing/ for precise analysis protocol. Summary statistics based on January 2021 data release, including up to 30 studies.	Please visit https://www.covid19hg.org/results/ for the description of individual studies.
COVID-B2	Hospitalized COVID-19 cases vs. controls	32404885	1,557,411	8,316; 1,549,095		
COVID-C2	All COVID-19 cases vs. controls	32404885	1,348,701	32,494; 1,316,207		

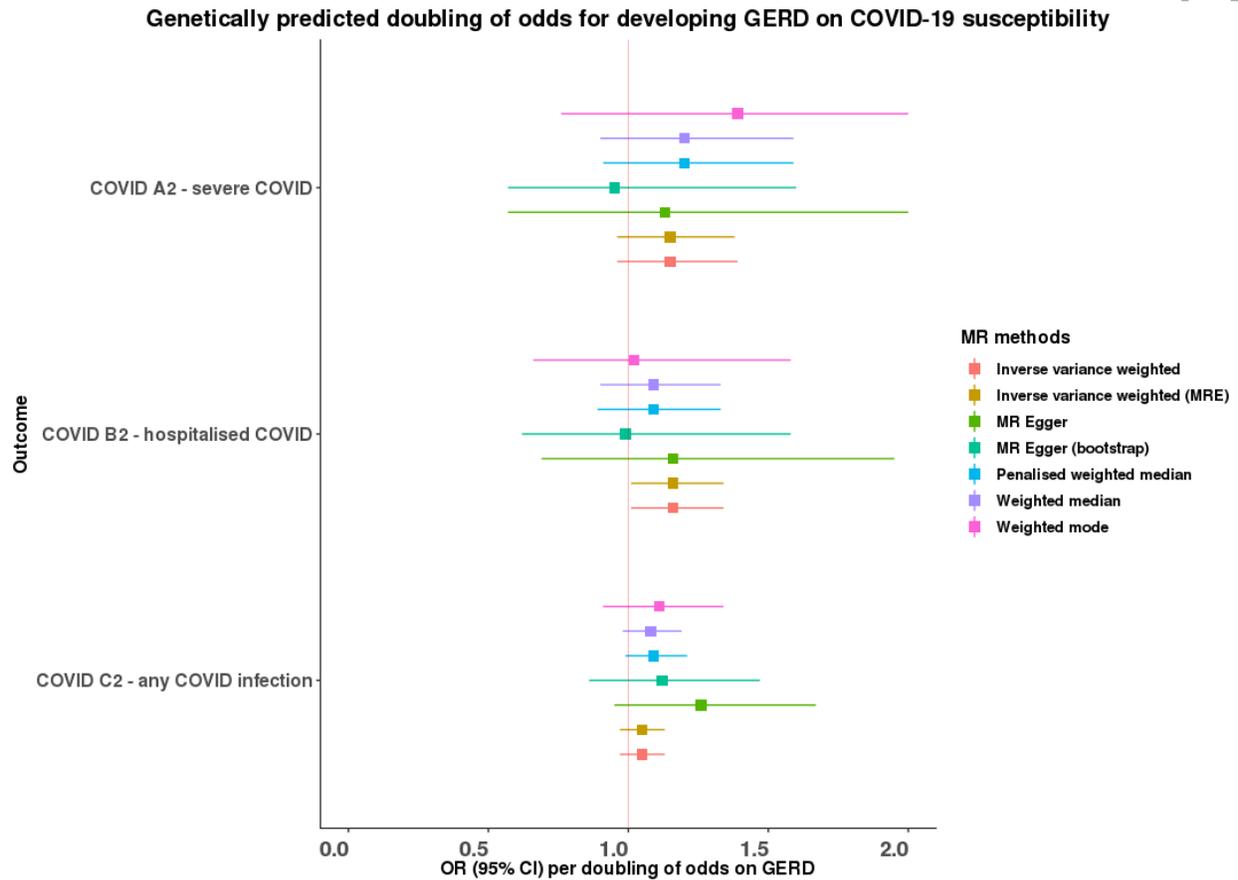
Figure captions

Figure 1. Diagram illustrating the observational risk factors and symptoms shared between COVID-19 outcomes and GERD susceptibility. Findings summarised from previous observational studies on GERD, risk of COVID-19 infection and severity (6–10). Arrows in bold are risk factors and symptoms unique to the disease.



UNCORRECTED

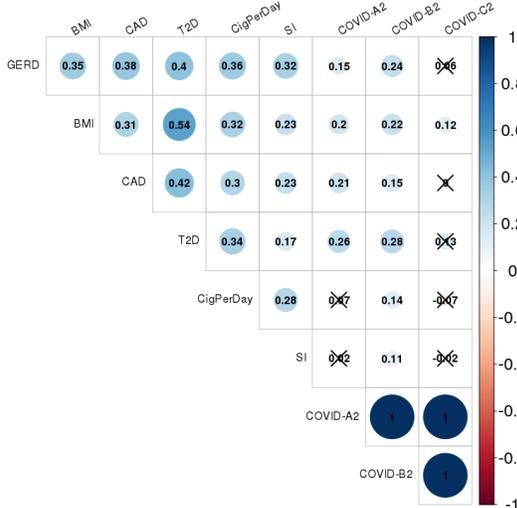
Figure 2. Mendelian randomization estimates on COVID-19 outcomes per doubling of odds in GERD susceptibility. COVID-A2: Severe COVID-19 infection; COVID-B2: COVID-19 hospitalization; COVID-C2: Overall COVID-19 susceptibility; MRE: Multiplication random effect. The model MR Egger (bootstrap) used 1000 iterations in the bootstrapping procedure. Error-bars represent the 95% CI around the derived OR estimate. For each COVID-19 outcome, every single row represents the OR estimate derived from different MR techniques. Whilst some estimators had lower precision, findings were largely consistent with those derived from the inverse-variance weighted model. Estimates derived using robust IVW and robust MR-Egger models are not presented here, but can be found in the Supplementary materials.



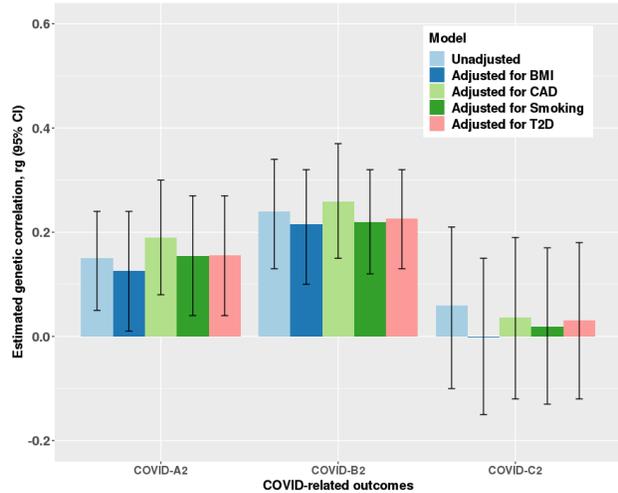
UNCORR

Figure 3. (Panel A) Pairwise genetic correlation between GERD, COVID-19 risk factors and COVID-19 outcomes. (Panel B) Genetic correlation estimates between GERD and risk factor-adjusted COVID-19 outcomes derived using bivariate LD-Score regression. COVID-A2: Severe COVID-19 infection; COVID-B2: COVID-19 hospitalization; COVID-C2: Overall COVID-19 susceptibility; SI: Smoking Initiation; CigPerDay: Cigarettes smoked per day; T2D: Type-II diabetes; CAD: Coronary artery disease; BMI: Body mass index. For panel A, estimates in cells that are crossed-off represent correlation estimates that were non-significant following correction for multiple testing. Note that the r_g estimates between COVID-19 outcomes were out of bound (i.e. $r_g > 1$) but are not significantly different from one. Panel B, error-bars represent the 95% CI around the point estimate. Adjustment for individual COVID-19 risk factors in the COVID-19 GWAS summary statistics performed via GCTA-mtCOJO. For the smoking-adjusted model, both cigarettes smoked per day and smoking initiation were conditioned on.

A) Estimated genetic correlation between GERD, COVID-19 outcomes and risk factors

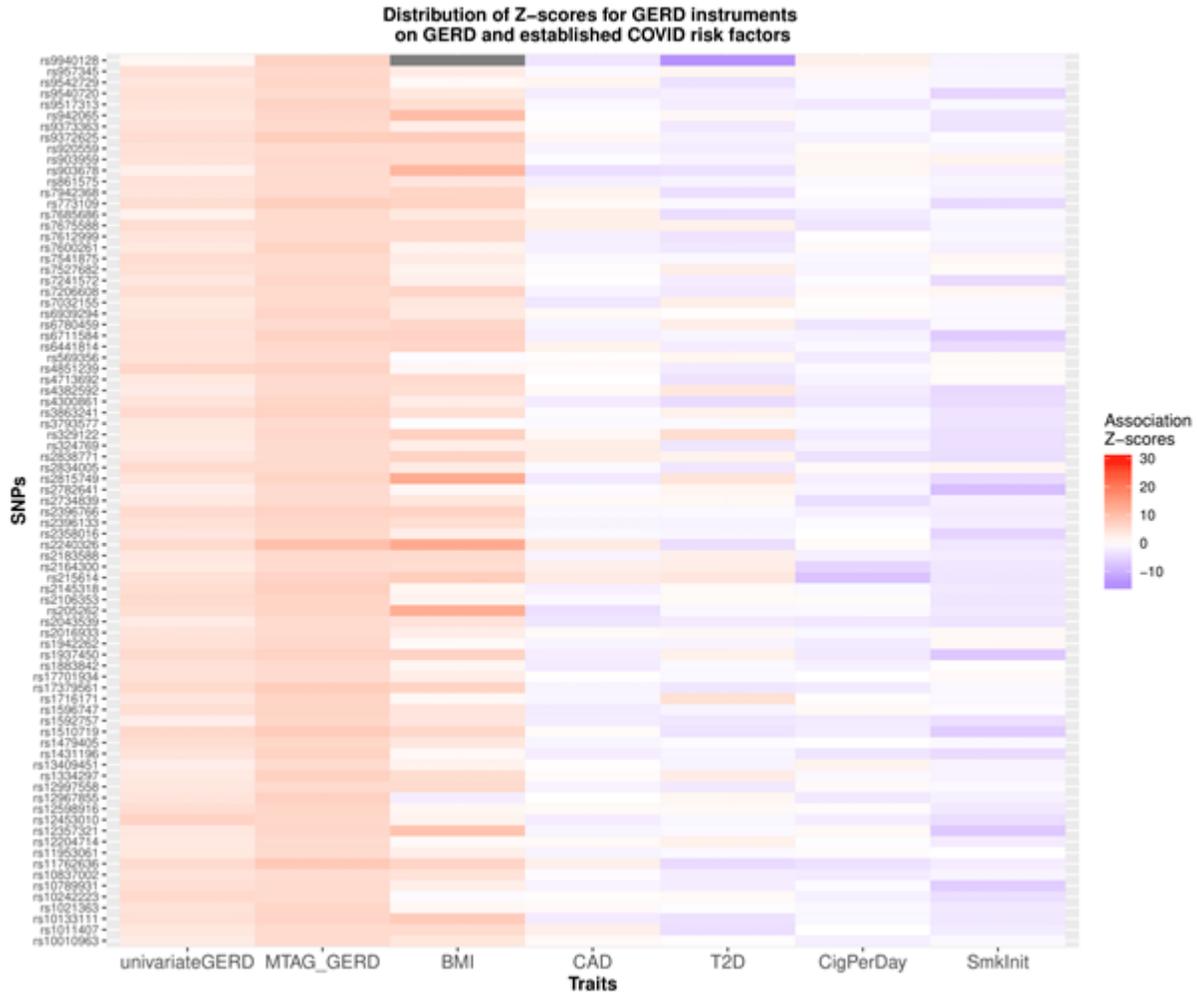


B) Estimated genetic correlation between GERD and COVID-19 outcomes adjusting for known COVID-19 risk factors via GCTA-mtCOJO



UNCORRECTED

Figure 4. Distribution of Z-Scores for the association between GERD SNP instruments and COVID-19 risk factors. COVID-A2: Severe COVID-19 infection; COVID-B2: COVID-19 hospitalization; COVID-C2: Overall COVID-19 susceptibility; BMI: Body mass index; CAD: Coronary artery disease; T2D: Type-II diabetes; CigPerDay: Cigarette smoked per day; SmkInit: Smoking initiation. The column on univariateGERD refers to the z-scores derived from the primary An et al. (2019) GERD GWAS. The column on MTAG_GERD refers to the z-scores derived from the multi-trait GERD GWAS model, which showed an overall enhancement in genetic signal on GERD by leveraging correlated traits. As shown above, there is evidence of pervasive pleiotropy between the GERD instruments and most of these risk factors.



UNCC

Figure 5. Mendelian randomization estimates on COVID-19 outcomes per doubling of odds in GERD susceptibility conditioned on known COVID-19 risk factors. COVID-A2: Severe COVID-19 infection; COVID-B2: COVID-19 hospitalization; COVID-C2: Overall COVID-19 susceptibility; IVW: Inverse variance weighted; BMI: Body mass index; CAD: Coronary artery disease; T2D: Type-II diabetes. For every COVID-19 outcome, each row represents the estimated IVW OR on risk of COVID-19 based on each unadjusted(original) model or model adjusted for known COVID-19 risk factors using GCTA-mtCOJO. Error-bars represent the 95% CI around the derived OR estimate. The smoking-adjusted model includes adjustment for both cigarette smoked per day and smoking initiation. Among all the aforementioned risk factors, the BMI-adjusted model showed the largest attenuation of effect towards the null. However, none of these covariate-adjusted estimates were meaningfully different from the unadjusted model.

