Assessing causal relationships between COVID-19 and non-alcoholic fatty liver disease

To the Editor:
We read with great interest two manuscripts by Ji et al. and Valenti et al. Recently, Ji et al. found that patients with coronavirus disease 2019 (COVID-19) had a higher risk of liver injury or severe outcomes, and meanwhile, patients with non-alcoholic fatty liver disease (NAFLD) had an increased risk of progression to severe COVID-19 infection.1 However, more recently, Valenti et al. found that metabolic-associated fatty liver disease (MAFLD) may not associate with the susceptibility and severity of COVID-19.2 Thus, causal relationships between COVID-19 and NAFLD still remain unclear. In this work, we conducted a bidirectional two-sample Mendelian randomization (MR) analysis to examine causal relationships between COVID-19 susceptibility/severity (including COVID-19 infection, COVID-19 hospitalization, and severe COVID-19 symptoms) and NAFLD.

MR is a causal inference approach that uses genetic variants as instrumental variables (IVs) to draw causal inferences between risk factors and health outcomes in observational study settings. It could overcome problems of confounding and reverse causality and is widespread for assessing causal relationships.3,4

For COVID-19, we used data from the latest and largest genome-wide association study (GWAS), COVID-19 Host Genetics Initiative (COVID-19 HGI) (N = 2,586,691), which was released on June 15, 2021. Three subgroups of European ancestry COVID-19 case-control studies included Group 1: COVID-19-infected cases vs. population (case = 87,870 and control = 1,810,493), and Group 2: hospitalized COVID-19 cases vs. population (case = 17,992 and control = 7,02,801), and Group 3: very severe respiratory confirmed COVID-19 cases vs. population (case = 4,606 and control = 702,801). Group 1 and Group 2 used the latest Round 6 data, and Group 3 used Round 5 data due to lacking of the latest wave of European ancestry data. We used the latest European ancestry UK Biobank (UKB) GWAS results of NAFLD (1,687 cases and 398,277 controls), which were based on TOPMed imputed data and analyzed using SAIGE (https://pheweb.org/UKB-TOPMed/). The NAFLD outcome in UKB GWAS was defined according to the ICD-9 571.8 “Other chronic non-alcoholic liver disease”. Although this definition of NAFLD might not include all clinical or histological cases, we found that the top hits of this GWAS included previously identified loci for NAFLD. The COVID-19 GWAS was adjusted for sex, ancestry and date of sample collection, and the UKB GWAS was adjusted for sex, birth year and first 4 principal components.

All the genetic IVs selected in our study were genome-wide significant (p < 5×10^-8) with minor allele frequency greater than 1%. After linkage disequilibrium (LD) clumping (clumping with window = 10,000 kb, R^2 = 0.001 in PLINK 1.9), there were 7 single nucleotide polymorphisms (SNPs) remaining for COVID-19 infection, 10 SNPs for hospitalized COVID-19, 8 SNPs for severe COVID-19, and 4 SNPs for NAFLD, respectively (Fig. 1A).

In the main MR analysis, we performed separate two-sample MR analyses with three COVID-19 phenotypes (i.e., COVID-19 infection, COVID-19 hospitalization, and COVID-19 severity) as exposures and NAFLD as the outcome. To further understand the causal effect of NAFLD on each COVID-19 phenotype, 4 genome-wide significant SNPs for NAFLD were selected as IVs. For the main analysis, we implemented the multiplicative random effect inverse-variance weighted (IVW) method to generate a causal estimate using the TwoSampleMR package.5 Moreover, we conducted sensitivity analyses, including MR Egger regression, weighted median, and MR-PRESSO, to account for potential heterogeneity and horizontal pleiotropy (supplementary information). p values less than 0.05 were considered statistically significant.

As shown in Fig. 1, the IVW analysis generated little evidence of a causal effect of COVID-19 infection (odds ratio (OR) 1.054; 95% CI 0.696–1.596), COVID-19 hospitalization (OR 0.973; 95% CI 0.750–1.263), and severe COVID-19 (OR 0.919; 95% CI 0.839–1.007) on NAFLD. Meanwhile, NAFLD did not appear to have a causal effect on COVID-19 infection (OR 0.991; 95% CI 0.970–1.013), hospitalization (OR 0.991; 95% CI 0.896–1.096) or severe COVID-19 (OR 0.904; 95% CI 0.731–1.118). All the sensitivity analyses showed broadly consistent results. No significant horizontal pleiotropy was detected by MR Egger regression or MR-PRESSO. Detailed MR Egger regression intercepts and MR-PRESSO global test results are presented in Table S1–S2.

There was no evidence derived from this work to support causal relationships between COVID-19 susceptibility/severity and NAFLD, which are consistent with results from previous genetic studies.6,7 On the contrary, Innes et al. found that the rs738409 C>G variant in patatin-like phospholipase domain-containing protein 3 (PNPLA3), which is well studied in the genetic regulation of NAFLD and liver injury, played a protective role in COVID-19 severity.7 Previous observational studies revealed that COVID-19-infected patients had an increased risk of liver injury,8,9 however, unmeasured/unmeasurable confounding cannot be ruled out in conventional multivariable regression analysis. In fact, our findings derived from the MR analysis could minimize confounding bias or reverse causation. It is worthy of note that collider bias, which could distort the relationship between exposures and outcomes, existed in many observational studies and cannot be completely overcome by MR.10 To investigate the potential impact of collider bias on the MR estimation, we replicated the bidirectional two-sample MR analysis using data from a previously published GWAS by Namjou et al., where genetic associations might be biased due to adjusting for a collider variable (e.g., BMI). Details on methods, results, and directed acyclic graph are included in the

---

2. Valenti et al. (2021).
8. Innes et al. (2021).
10. Innes et al. (2021).
Fig. 1. The genome-wide significant variants and bidirectional MR results of COVID-19 infection, hospitalized COVID-19, severe COVID-19 cases, and NAFLD. (A) The genome-wide significant variants associated with COVID-19 infection, hospitalized COVID-19, severe COVID-19 cases, and NAFLD. (B) MR analysis results of COVID-19 on NAFLD. (C) MR analysis results of NAFLD on COVID-19. (D) Forest plot of MR results of COVID-19 on NAFLD. (E) Forest plot of MR results of NAFLD on COVID-19. *COVID-19, coronavirus disease 2019; EA, effect allele; EAF, effect allele frequency; MR, Mendelian randomization; NAFLD, non-alcoholic fatty liver disease; OA, other allele.
supplementary information. Future larger randomly sampled cohort studies on COVID-19 and NAFLD and rigorous statistical analysis might be helpful to minimize the impact of collider bias.

In conclusion, we found little evidence to support a causal relationship between COVID-19 susceptibility/severity and NAFLD.

Financial support
The authors received no financial support to produce this work.

Conflict of interest
The authors declare no conflicts of interest that relate to this work.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors’ contributions
All authors drafted and reviewed this manuscript.

Supplementary data
Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2021.11.014.

References
Author names in bold designate shared co-first authorship