Smoking is significantly associated with increased risk of COVID-19 and other respiratory infections

Daniel B. Rosoff1,2,3, Joyce Yoo1,3 & Falk W. Lohoff1✉

Observational studies suggest smoking, cannabis use, alcohol consumption, and substance use disorders (SUDs) may impact risk for respiratory infections, including coronavirus 2019 (COVID-2019). However, causal inference is challenging due to comorbid substance use. Using summary-level European ancestry data (>1.7 million participants), we performed single-variable and multivariable Mendelian randomization (MR) to evaluate relationships between substance use behaviors, COVID-19 and other respiratory infections. Genetic liability for smoking demonstrated the strongest associations with COVID-19 infection risk, including the risk for very severe respiratory confirmed COVID-19 (odds ratio (OR) = 2.69, 95% CI, 1.42, 5.10, $P$-value $= 0.002$), and COVID-19 infections requiring hospitalization (OR = 3.49, 95% CI, 2.23, 5.44, $P$-value $= 3.74 \times 10^{-8}$); these associations generally remained robust in models accounting for other substance use and cardiometabolic risk factors. Smoking was also strongly associated with increased risk of other respiratory infections, including asthma-related pneumonia/sepsis (OR $= 3.64$, 95% CI, 2.16, 6.11, $P$-value $= 1.07 \times 10^{-6}$), chronic lower respiratory diseases (OR $= 2.29$, 95% CI, 1.80, 2.91, $P$-value $= 1.69 \times 10^{-11}$), and bacterial pneumonia (OR $= 2.14$, 95% CI, 1.42, 3.24, $P$-value $= 2.84 \times 10^{-4}$). We provide strong genetic evidence showing smoking increases the risk for COVID-19 and other respiratory infections even after accounting for other substance use behaviors and cardiometabolic diseases, which suggests that prevention programs aimed at reducing smoking may be important for the COVID-19 pandemic and have substantial public health benefits.
Since the first reported cases in Wuhan, China in December 2019, coronavirus disease 2019 (COVID-19) has subsequently affected more than 200 countries and continues to be a global pandemic of substantial worldwide morbidity and mortality. More broadly, upper and lower respiratory infections (URIs and LRIs, respectively) and other respiratory diseases (i.e., asthma, chronic obstructive pulmonary disease (COPD), etc.) are leading causes of yearly worldwide morbidity and mortality. For example, the Global Burden of Disease Study estimated that LRIs caused more than two million deaths globally in 2016, while approximately 2.3 million people died from COPD in 2015. Respiratory infection and diseases are also a large economic burden: URIs result in more than 40 million missed days of school and work per year.

Substance use (tobacco smoking, cannabis use, and alcohol consumption) are risk factors linked with adverse lung and respiratory outcomes. For example, observational data has shown chronic heavy alcohol consumption to be associated with increased risk for pneumonia and acute respiratory distress syndrome (ARDS), while cannabis smoke has been shown to contain many of the same toxins and irritants as smoke derived from tobacco, but may differ from tobacco in its association with bronchitis and other respiratory infections. In addition, it has been suggested that chronic alcohol abuse may compromise the ability of immune cells to destroy bacteria in the lungs, which may result in an increased vulnerability to respiratory infections like pneumonia and tuberculosis.

Paralleling the COVID-19 pandemic have been increases in substance use, which combined with data showing approximately 10.8% of US adults suffering from a substance use disorder (SUD) and recent work using electronic health records (EHRs) to show that individuals with a SUD are at increased risk for reduced precision (hospitalized COVID-19 vs population: SVMR OR = 2.42, 95% CI, 1.46, 4.01, P-value = 6.09 × 10^−4; MVMR OR = 2.62, 95% CI, 1.46, 4.71, P-value = 0.001; and hospitalized vs not hospitalized COVID-19: SVMR OR = 3.44, 95% CI, 1.72, 6.87, P-value = 4.60 × 10^-4; MVMR OR = 3.61, 95% CI, 1.63, 8.01, P-value = 0.002) (Table 1; Supplementary Data 8, 10, and 11). This association remained robust in second sensitivity analyses excluding UK Biobank participants in the COVID-19 outcome GWAS, but with reduced precision (hospitalized COVID-19 vs population: SVMR OR = 2.42, 95% CI, 1.46, 4.01, P-value = 6.09 × 10^−4; MVMR OR = 2.62, 95% CI, 1.46, 4.71, P-value = 0.001; and hospitalized vs not hospitalized COVID-19: SVMR OR = 3.27, 95% CI, 1.15, 9.33, P-value = 0.03; MVMR OR = 4.84, 95% CI, 1.46, 15.39, P-value = 0.008) (Supplementary Data 8, 10, and 11). Importantantly, these associations were consistent across complementary SVMR and MVMR methods, including single variable GSRM (Supplementary Data 8, 10, and 11). Leave-one-out analyses highlight variants with heterogeneous causal effects that would be flagged as invalid by MR PRESSO and MVMR Lasso and removed for outlier corrected results (Supplementary Data 9).

Given the strong associations of lifetime tobacco smoking and COVID-19 risk, we further evaluated robustness by performing MVMR analyses accounting for cardiometabolic disorders (CAD, T2D, and obesity) previously reported as risk factors for COVID-19 risk. Genetic liability for lifetime tobacco smoking generally remained associated with increased risk for COVID-19 hospitalization (e.g., accounting for CAD, hospitalized COVID-19 vs. population: SVMR OR = 3.18, 95% CI, 2.06, 4.92, P-value = 1.80 × 10^−7; and accounting for Type 2 diabetes, MVMR OR = 4.16, 95% CI, 2.51, 6.92, P-value = 3.76 × 10^−8; accounting for obesity, MVMR OR = 3.75, 95% CI, 2.25, 6.25, P-value = 4.01 × 10^−7) (Supplementary Data 12).

### Associations of substance use and SUDs with COVID-19 infection risk

COVID-19 results comparing SVMR and MVMR results are presented in Table 1. Supplementary Data 8–12 present the full COVID-19 results. Broadly, among all substance use exposures, the genetic liability for lifetime tobacco smoking consistently demonstrated the strongest associations with COVID-19 infection risk, including the risk for very severe respiratory confirmed COVID-19 (SVMR odds ratio (OR) = 2.69, 95% CI, 1.42, 5.10, P-value = 0.002, and also the risk for COVID-19 infection requiring hospitalization (hospitalized COVID-19 vs population: SVMR odds ratio (OR) = 3.49, 95% CI, 2.23, 5.44, P-value = 3.74 × 10^−8; MVMR accounting for substance use disorders OR = 3.61, 95% CI, 2.19, 5.95, P-value = 4.92 × 10^−7; and hospitalized vs not hospitalized COVID-19: SVMR OR = 3.44, 95% CI, 1.72, 6.87, P-value = 4.60 × 10^−4; MVMR OR = 3.61, 95% CI, 1.63, 8.01, P-value = 0.002) (Table 1; Supplementary Data 8, 10, and 11). This association remained robust in secondary sensitivity analyses excluding UK Biobank participants in the COVID-19 outcome GWAS, but with reduced precision (hospitalized COVID-19 vs population: SVMR OR = 2.42, 95% CI, 1.46, 4.01, P-value = 6.09 × 10^−4; MVMR OR = 2.62, 95% CI, 1.46, 4.71, P-value = 0.001; and hospitalized vs not hospitalized COVID-19: SVMR OR = 3.27, 95% CI, 1.15, 9.33, P-value = 0.03; MVMR OR = 4.84, 95% CI, 1.46, 15.39, P-value = 0.008) (Supplementary Data 8, 10, and 11). Importantantly, these associations were consistent across complementary SVMR and MVMR methods, including single variable GSRM (Supplementary Data 8, 10, and 11). Leave-one-out analyses highlight variants with heterogeneous causal effects that would be flagged as invalid by MR PRESSO and MVMR Lasso and removed for outlier corrected results (Supplementary Data 9).
As with COVID-19 infection risk results, we found that the genetic liability of lifetime tobacco smoking was the substance use risk factor with the strongest associations, including results that were robust in MVMR models. Tobacco smoking, for example, was associated with increased risk of asthma-related infections and asthma-related pneumonia/asthma (SVMR OR = 2.52, 95% CI 1.59, 3.97, P-value = 7.29 × 10^{-3}; accounting for substance use disorders, MVMR OR = 3.64, 95% CI 2.16, 6.11, P-value = 1.07 × 10^{-6}), but for neither bronchitis nor the common cold (Table 3; Supplementary Data 13–15). Tobacco smoking was also associated with chronic lower respiratory diseases (SVMR OR = 2.23, 95% CI 1.73, 2.87, P-value = 5.69 × 10^{-10}; MVMR OR = 2.29, 95% CI 1.80, 2.91, P-value = 1.69 × 10^{-11}) and several pneumonia-related outcomes, including bacterial pneumonia (SVMR OR = 2.22, 95% CI 1.57, 3.15, P-value = 7.32 × 10^{-6}; MVMR OR = 2.14, 95% CI 1.42, 3.24, P-value = 2.84 × 10^{-4}) (Table 5, Supplementary Data 13–15).

As with the smoking-COVID-19 findings, we tested robustness of the smoking-respiratory infection risk results using additional MVMR models that accounted for cardiometabolic disorders (CAD, T2D, and obesity) with evidence for an impact on respiratory infection risk results using additional neous causal effects that would be fl

| Table 1 Single variable and multivariable MR results of the genetic liability of alcohol, cannabis and lifetime smoking exposures on COVID-19 outcomes. |
|---|---|---|---|---|---|---|---|---|---|
| | Single-variable MR | | Multivariable MR | |
| | N SNPs | OR | 95% CI Lower | 95% CI Upper | P-value | N SNPs | MV OR | 95% CI Lower | 95% CI Upper | P-value |
| **Very severe respiratory confirmed COVID-19 vs. population** | | | | | | | | | |
| Tobacco smoking | 91 | 2.69 | 1.42 | 5.10 | **0.002** | 111 | 2.72 | 1.27 | 5.82 | 0.010 |
| Cannabis use | 28 | 1.17 | 0.92 | 1.50 | 0.207 | 114 | 1.03 | 0.76 | 1.42 | 0.805 |
| CUD | 22 | 0.97 | 0.82 | 1.15 | 0.748 | 111 | 1.02 | 0.82 | 1.22 | 0.805 |
| Drinks per week | 22 | 1.39 | 0.52 | 3.74 | 0.511 | 114 | 0.83 | 0.33 | 2.30 | 0.698 |
| AUD | 9 | 0.91 | 0.75 | 1.11 | 0.344 | 111 | 0.95 | 0.82 | 1.09 | 0.442 |
| **Hospitalized COVID-19 vs. not hospitalized COVID-19** | | | | | | | | | |
| Tobacco smoking | 91 | 3.44 | 1.72 | 6.87 | **4.60E–04** | 111 | 3.61 | 1.63 | 8.01 | **0.002** |
| Cannabis use | 28 | 0.87 | 0.68 | 1.11 | 0.270 | 114 | 1.02 | 0.77 | 1.38 | 0.883 |
| CUD | 21 | 1.07 | 0.91 | 1.25 | 0.404 | 111 | 1.04 | 0.87 | 1.25 | 0.627 |
| Drinks per week | 21 | 0.69 | 0.27 | 1.76 | 0.432 | 114 | 0.37 | 0.15 | 1.45 | 0.034 |
| AUD | 9 | 0.83 | 0.69 | 0.99 | 0.035 | 111 | 0.94 | 0.81 | 1.10 | 0.451 |
| **Hospitalized COVID-19 vs. population** | | | | | | | | | |
| Tobacco smoking | 91 | 3.49 | 2.23 | 5.44 | **3.74E–08** | 111 | 3.61 | 2.19 | 5.95 | **4.92E–07** |
| Cannabis use | 28 | 0.99 | 0.85 | 1.16 | 0.887 | 114 | 1.00 | 0.83 | 1.27 | 0.964 |
| CUD | 21 | 0.99 | 0.90 | 1.00 | 0.871 | 111 | 1.01 | 0.90 | 1.13 | 0.915 |
| Drinks per week | 21 | 1.01 | 0.57 | 1.81 | 0.964 | 114 | 0.59 | 0.32 | 1.80 | 0.079 |
| AUD | 8 | 0.94 | 0.82 | 1.08 | 0.375 | 111 | 0.98 | 0.89 | 1.07 | 0.633 |
| **COVID-19 vs. population** | | | | | | | | | |
| Tobacco smoking | 91 | 1.21 | 0.97 | 1.52 | 0.095 | 111 | 1.25 | 0.95 | 1.64 | 0.104 |
| Cannabis use | 28 | 1.11 | 1.02 | 1.21 | 0.022 | 114 | 1.04 | 0.95 | 2.83 | 0.404 |
| CUD | 21 | 0.98 | 0.93 | 1.03 | 0.436 | 111 | 0.98 | 0.92 | 1.04 | 0.411 |
| Drinks per week | 21 | 1.09 | 0.81 | 1.47 | 0.565 | 114 | 0.93 | 0.68 | 2.52 | 0.622 |
| AUD | 8 | 1.08 | 1.00 | 1.16 | 0.039 | 111 | 1.03 | 0.98 | 1.08 | 0.310 |

Notes: Results from two sample SVMR inverse-variance weighted MR analysis; outliers identified by MR PRESSO global test and, for MVMR, MR MV R Lasso penalization were removed. Estimated associations reported as odds ratios with 95% confidence intervals. Boldface indicates statistical significance after correction for multiple comparisons (P < 0.0025). Genetic instruments selected from five GWASs; selection threshold P < 5 × 10^{-6} or P < 5 × 10^{-5} (CUD and AUD), clumped at linkage disequilibrium (LD) r^2 = 0.001 (10 000 kilobase pair window); N SNPs differs across outcomes depending on number of genetic instruments found in outcome GWASs. AUD: alcohol use disorder, CUD: cannabis use disorder, COVID-19: coronavirus 2019, MVMR: Mendelian randomization, GWAS: genome-wide association study, N SNPs: number of single-nucleotide polymorphism (genetic instruments), OR: odds ratio, CI: confidence interval.

**Discussion**

Using large summary-level GWAS data and complementary two-sample MR methods, we show that the genetic liability for tobacco smoking has potential causal relationships with several respiratory infection and disease outcomes, including COVID-19. These tobacco smoking-respiratory-finding results were supported by multivariable MR analyses accounting for alcohol and cannabis use and abuse, which in addition to the broadly consistent IVW results (within the IVW MR 95% confidence interval but typically less precise) with estimates from the weighted median, weighted mode, and MR Egger sensitivity analyses strengthen causal inference. Further, in single variable MR, we identify potential adverse impacts of CUD on lower respiratory infections, the common cold, and several asthma-related infections, suggesting evidence for a dose-dependent impact of cannabis use where heavy cannabis use may be harmful to the respiratory system. In parallel, we find little evidence for an alcohol-respiratory infection relationship suggesting that previous observational data may be due to confounding.

Our COVID-19 results extend recent MR studies showing adverse effects of smoking on COVID-19 risk by accounting for highly comorbid alcohol consumption, cannabis use, and SUDs, which when combined with reports suggesting smoking intensifies the severity of COVID-19 symptoms33,34, the risk for being PRESSO and MV R Lasso and removed for outlier corrected results (Supplementary Data 17).
admitted to an intensive care unit or requiring ventilation34, and recent transcriptomics-based work showing that smoking may increase the expression of angiotensin converting enzyme 2 (ACE2), the putative receptor for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (the virus that causes COVID-19)35, suggests smoking may be an important modifiable risk factor for COVID-19.

Our genetics-based findings support and extend the observational literature identifying tobacco smoking as a risk factor for respiratory infection and diseases9,23,26, and add to the recent MR literature identifying potential causal links of smoking with reduced lung function16, lung cancer37, and increased mortality due to respiratory disease38. Potential mechanisms by which smoking increases respiratory infection risk include structural changes to the respiratory tract and a dysregulated cellular and humoral immune response, including peribronchiolar inflammation, decreased levels of circulating immunoglobulins, and changes to pathogen adherence. For example, smoking has been shown to stimulate the release of catecholamine and corticosteroids, which may, in turn, increase circulating CD8+ lymphocytes and suppress the host defense against infections. Notably, many immunological effects related to smoking may resolve within six weeks of smoking cessation, which suggests that smoking cessation programs may have an important impact on reducing respiratory infections.

Regarding cannabis use, while we failed to find evidence of any relationships, smoking cannabis, like tobacco smoking, may prompt the onset of coughing, which could consequently increase viral transmission, or may possibly exacerbate respiratory symptoms.

As cannabis is the most used drug worldwide—an estimated 188 million recreational users worldwide—this aspect of cannabis use may have important implications for the spread of COVID-19. In contrast, the single-variable MR CUD results demonstrated adverse effects on several respiratory outcomes, but not COPD, which supports the existing literature39–41; however, accounting for lifetime tobacco smoking attenuated the CUD results, thus highlighting the complex nature of these relationships. Further, habitual cannabis smoking may have several effects on respiratory and immune systems that may impact respiratory infection susceptibility. For example, structural abnormalities in alveolar macrophages and coincident dysregulated cytokine production and antimicrobial activity have been reported. While our study provides preliminary genetic evidence suggesting potential causal relationships between heavy cannabis use and respiratory infection, additional triangulating lines of evidence (i.e., immune monitoring studies) are required to further elucidate the CUD-respiratory infection relationship. However, given that the toxic and irritant profiles of cannabis and tobacco smoke are similar11, the direct route of administration via inhalation for these

### Table 2 Single variable and multivariable MR results of the genetic liability of alcohol, cannabis and lifetime smoking exposures on asthma-related respiratory infections.

| Outcome | Exposure | Single-variable MR | Multivariable MR |
|---------|----------|--------------------|------------------|
|         |          | N SNPs | OR | 95% Cl Lower | 95% Cl Upper | P-value | N SNPs | OR | 95% Cl Lower | 95% Cl Upper | P-value |
| Asthma related acute respiratory infections | Tobacco smoking | 116 | 2.06 | 1.39 | 3.05 | 3.15E−04 | 137 | 2.07 | 1.31 | 3.26 | 0.002 |
|         | Cannabis Use | 35 | 1.07 | 0.98 | 1.16 | 0.128 | 142 | 1.03 | 0.88 | 1.22 | 0.704 |
|         | CUD | 27 | 1.10 | 1.03 | 1.19 | 0.007 | 137 | 1.01 | 0.92 | 1.11 | 0.835 |
|         | Drinks Per Week | 32 | 0.92 | 0.56 | 1.53 | 0.757 | 142 | 1.06 | 0.65 | 1.72 | 0.816 |
|         | AUD | 11 | 1.04 | 0.93 | 1.17 | 0.456 | 137 | 1.01 | 0.93 | 1.10 | 0.826 |
| Asthma related infections | Tobacco smoking | 115 | 2.31 | 1.61 | 3.31 | 5.69E−06 | 127 | 2.15 | 1.45 | 3.17 | 1.21E−04 |
|         | Cannabis Use | 35 | 1.02 | 0.94 | 1.09 | 0.672 | 142 | 1.06 | 0.91 | 1.23 | 0.484 |
|         | CUD | 27 | 1.08 | 1.00 | 1.16 | 0.040 | 127 | 0.98 | 0.91 | 1.07 | 0.662 |
|         | Drinks Per Week | 32 | 0.94 | 0.61 | 1.44 | 0.773 | 142 | 0.96 | 0.61 | 1.49 | 0.845 |
|         | AUD | 11 | 1.04 | 0.95 | 1.13 | 0.385 | 127 | 1.03 | 0.96 | 1.10 | 0.358 |
| Asthma-related pneumonia | Tobacco smoking | 116 | 2.52 | 1.59 | 3.97 | 7.29E−05 | 138 | 3.64 | 2.16 | 6.11 | 1.07E−06 |
|         | Cannabis Use | 36 | 0.95 | 0.85 | 1.06 | 0.382 | 136 | 1.07 | 0.90 | 1.27 | 0.500 |
|         | CUD | 27 | 1.07 | 0.98 | 1.18 | 0.120 | 138 | 0.95 | 0.85 | 1.06 | 0.380 |
|         | Drinks Per Week | 31 | 1.47 | 0.83 | 2.59 | 0.187 | 136 | 1.09 | 0.61 | 1.93 | 0.272 |
|         | AUD | 11 | 1.09 | 0.97 | 1.21 | 0.132 | 138 | 0.99 | 0.90 | 1.08 | 0.755 |
| Asthma-related pneumonia or sepsis | Tobacco smoking | 116 | 2.54 | 1.61 | 4.02 | 6.54E−05 | 138 | 3.66 | 2.17 | 6.16 | 1.04E−06 |
|         | Cannabis Use | 35 | 0.96 | 0.87 | 1.07 | 0.472 | 141 | 1.08 | 0.90 | 1.31 | 0.401 |
|         | CUD | 27 | 1.07 | 0.98 | 1.18 | 0.122 | 138 | 0.95 | 0.85 | 1.07 | 0.399 |
|         | Drinks Per Week | 32 | 1.85 | 1.04 | 3.26 | 0.035 | 141 | 1.15 | 0.66 | 2.00 | 0.616 |
|         | AUD | 11 | 1.09 | 0.97 | 1.21 | 0.138 | 138 | 0.98 | 0.89 | 1.08 | 0.734 |

Notes: Results from two sample SVMR inverse-variance weighted MR analysis; outliers identified by MR PRESSO global test and, for MVMR, MV MR Lasso penalization were removed; estimated associations reported as odds ratios with 95% confidence intervals. Boldface indicates statistical significance after correction for multiple comparisons (P < 0.000074). Genetic instruments selected from 5 GWASs, selection threshold P < 5 × 10−8 or P < 5 × 10−6 (CUD and AUD), clumped at linkage disequilibrium (LD) r2 = 0.001 (10 kilobase pair window); N SNPs differs across outcomes depending on number of genetic instruments found in outcome GWASs. CUD cannabis use disorder, AUD alcohol use disorder, MR Mendelian randomization, GWAS genome wide association study, N SNPs number of single-nucleotide polymorphism (genetic instruments), OR odds ratio, CI confidence interval.
substances could result in dysregulated pulmonary physiology which may, in turn, increase infection risk.

In contrast to our tobacco smoking findings, we failed to find genetic evidence of respiratory implications due to alcohol consumption not meeting the threshold of AUD, or binge drinking, suggesting that previous observational literature may be due to confounding from other comorbid behaviors—such as smoking—that may be the true causal risk factors for respiratory infections. For example, observational and genetic evidence have shown a strong association between alcohol consumption and smoking. It may be that previous observational literature may be due to confounding from other comorbid behaviors such as smoking.
has been estimated that 85% of smokers consume alcohol\textsuperscript{42-44} and alcohol drinkers are 75% more likely than abstainers to smoke\textsuperscript{45}. Therefore, it is possible that the observational study-based alcohol-respiratory infection links may be due, instead, to tobacco smoking; however, future work will be needed to confirm this hypothesis. In addition, it is important to note that our results should not be interpreted as suggesting that alcohol does not impact overall lung health and structure, which has been previously reported\textsuperscript{7}. Further, while we failed to find evidence that weekly alcohol consumption impacted COVID-19 risk, the Centers for Disease Control recently showed that dining at on-site locations, such as restaurants and bars, is associated with increased COVID-19 risk; since alcohol consumption may lower inhibition and increase impulsivity, individuals consuming alcohol may take social distancing less seriously, and thereby unintentionally spread the SARS-CoV-2 virus.

This study has several strengths including the use of multiple alcohol consumption and cannabis use variables, which enabled us to evaluate various dimensions of substance use and abuse and identify possible causal relationships of substance use disorders and respiratory outcomes. In addition, our main single variable analyses included multiple MR methods, each relying on orthogonal assumptions, which provide confidence in robustness of the results and strengthen causal inference\textsuperscript{46}. Our multivariable two-sample MR design, the most appropriate design given the strong correlation between tobacco smoking, alcohol consumption and cannabis use, yielded estimates that account for these correlated behaviors for each exposure on COVID-19 risk and other respiratory outcomes. Another strength is our extension of MVMR to test the robustness of the main tobacco smoking findings by incorporating other potential confounders that may impact infectious disease risk (obesity, cardiovascular disease, and T2D).

This study also has limitations. A main limitation is the possibility of collider bias—especially with regards to the COVID-19 datasets\textsuperscript{17,49}. Collider bias may occur when analyses are controls or selects the sample based upon a collider variable that is caused by both the exposure and outcome variables and distorts the true underlying association\textsuperscript{48,49}. The recent commentaries by Griffith et al. (2020) and Tattan-Birch et al. (2020) discuss in detail the potential for collider bias in COVID-19 datasets\textsuperscript{17,49}, and are important for context when interpreting COVID-19 findings based upon observational data. For example, an observational study from early in the COVID-19 pandemic reported an apparent protective effect of tobacco smoking on COVID-19 risk\textsuperscript{50}; however, as Tattan-Birch et al. discuss, both smoking and COVID-19 may cause coughing, which, during the COVID-19 pandemic, may increase the likelihood for smokers to be tested and their subsequent overrepresentation among clinical study participants testing negative for COVID-19\textsuperscript{49}. As a result, among samples with COVID-19 tests, smoking may appear to have a protective effect\textsuperscript{49}. While it is often not possible to ensure the absence of collider bias\textsuperscript{47}, we aimed to design our study incorporating measures that may mitigate its impact. For example, we used the most recently released version of publicly available COVID-19 data (from January 18, 2021)\textsuperscript{51} that may include participants more representative of the general population compared to samples collected earlier in the COVID-19 pandemic. Reassuringly, we also found similar smoking effect estimates in several respiratory-related infection outcomes, which suggests a broader impact of smoking on the respiratory system that extends to COVID-19.

In addition, as with all self-reported substance use literature, these exposures may be either under- or over-reported\textsuperscript{52}. Because many of the datasets included UK Biobank participants, who are more educated, lead healthier lifestyles, and have fewer health problems than the UK population\textsuperscript{53}, this discrepancy may limit the applicability of our findings to other populations. Regarding our mainly null alcohol-respiratory infection results, it is possible that alcohol may have indirect impacts on infection risk through a modified immune response\textsuperscript{48}, or other system dysregulations that may modulate infection risk that we were not able to directly assess using MR. However, like other recent psychiatric MR studies where the exposure instruments included a relaxed statistical threshold, our binge drinking and AUD instruments were comprised of independent SNPs associated with the respective drinking behavior (i.e., $P$-value $< 5 \times 10^{-8}$) for SNP inclusion due to the lack of conventionally GWS SNPs.

| Outcome Exposure | Outcome Exposure | Single-variable MR | Multivariable MR |
|------------------|------------------|-------------------|------------------|
|                  |                  | N SNPs OR 95% CI Lower 95% CI Upper P-value | N SNPs OR 95% CI Lower 95% CI Upper P-value |
| Bacterial pneumonia | Tobacco smoking | 117 2.22 1.57 3.15 | 732E-06 138 2.14 1.42 3.24 |
|                  | Cannabis Use | 36 1.05 0.96 1.15 | 0.268 142 0.97 0.84 1.12 |
|                  | CUD | 27 1.05 0.98 1.13 | 0.145 138 1.02 0.93 1.12 |
|                  | Drinks Per Week | 33 1.17 0.72 1.89 | 0.530 142 1.19 0.78 1.83 |
|                  | AUD | 11 1.04 0.95 1.14 | 0.426 138 1.01 0.93 1.08 |
| All Pneumonia | Tobacco smoking | 115 1.52 1.22 1.88 | 1.34E-04 134 1.46 1.15 1.85 |
|                  | Cannabis Use | 34 1.01 0.96 1.06 | 0.833 136 1.03 0.95 1.12 |
|                  | CUD | 27 1.05 1.01 1.09 | 0.014 134 0.98 0.94 1.04 |
|                  | Drinks Per Week | 31 0.97 0.76 1.24 | 0.814 136 0.96 0.75 1.23 |
|                  | AUD | 11 1.05 1.00 1.11 | 0.056 134 1.01 0.97 1.05 |
| Viral Pneumonia | Tobacco smoking | 117 1.66 0.53 5.17 | 0.386 138 1.70 0.44 6.52 |
|                  | Cannabis Use | 36 1.07 0.81 1.41 | 0.638 142 1.34 0.84 2.15 |
|                  | CUD | 27 1.00 0.78 1.28 | 0.993 138 0.90 0.67 1.20 |
|                  | Drinks Per Week | 33 2.24 0.60 8.43 | 0.232 142 1.27 0.32 5.04 |
|                  | AUD | 11 1.13 0.85 1.51 | 0.409 138 1.09 0.85 1.39 |

Notes: Results from two sample SVMR inverse-variance weighted MR analysis; outliers identified by MR PRESSO global test and, for MVMR, MR Lasso penalization were removed; estimated associations reported as odds ratios with 95% confidence intervals. Boldface indicates statistical significance after correction for multiple comparisons ($P < 0.0000714$). Genetic instruments selected from 5 GWASs, selection threshold $P < 10^{-8}$ or $P < 5 \times 10^{-6}$ (AUD and AUD), clumped at linkage disequilibrium (LD) $r^2 = 0.01$ (100 kib base pair window); N SNPs differs across outcomes depending on number of genetic instruments found in outcome GWASs. AUD, cannabis use disorder, AUD, alcohol use disorder, MR Mendelian randomization, GWAS genome wide association study, N SNPs number of single-nucleotide polymorphism (genetic instruments). OR odds ratio, CI confidence interval.
**Methods**

**Data sources and genetic instruments.** Summary-level data for both modifiable risk factor instrument and infectious disease outcome data were derived from publicly available GWAS in populations of predominantly European ancestry (Fig. 1; Table 6; Supplementary Data 1). All GWAS have existing ethical permissions from their respective institutional review boards and include participant informed consent with rigorous quality control. For this study, we included all exposure SNPs associated at conventional genome-wide significance (GWAS) P $<$ 5 x 10^-8 for smoking, alcohol and cannabis use, and 5 x 10^-6 for AUD and CUD due to the relatively low number of SNPs at GWAS, clumped at linkage disequilibrium (LD) $r^2$ = 0.001 and a distance of 10,000 kb, using reference samples comprised of participants of European ancestry 34.

**Substance use exposures:**
- Lifetime smoking (N=462,690)
- Cannabis use (N=162,062)
- Cannabis use disorder (CUD) (N=358,534)
- Drinks per week (N=941,280)
- Alcohol use disorder (AUD) (N=28,757)

**Outcomes:**
**COVID-19:** Very severe respiratory confirmed COVID-19 v population (N=707,407); Hospitalized v not hospitalized (N=16,645); Hospitalized v population (N=1,206,629); Very severe hospitalized v population (N=689,689); COVID-19 v population (European only) (N=1,209,010); COVID-19 v population (N=1,586,783)

**FinnGen respiratory infections**
Acute Upper Respiratory Infections (N=218,792); Asthma ARI (N=142,793); Asthma COPD (N=208,167); Asthma-related infections (N=218,792); Asthma-related pneumonia and sepsis (N=140,994); Asthma-related pneumonia (N=140,991); Bacterial pneumonia (N=190,655); Bronchitis (N=218,792); Chronic Lower Respiratory Diseases (N=218,792); Common Cold (N=165,198); Influenza (N=120,672); Influenza and pneumonia (N=190,130); All pneumonia (N=218,792); Viral Pneumonia (N=189,568)

**SNP extraction/instrument selection**
- SNP-smoking associations
- SNP-cannabis associations
- SNP-alcohol associations

**Two-sample MR**
- Single variable Mendelian randomization (MR):
  - IVW, MR Egger, Weighted Median, Weighted Mode;
  - Heterogeneity testing: MR Egger Intercept, MR PRESSO, Cochran Q

**Multivariable MR (MVMR):**
- IVW, MR Egger
- Heterogeneity testing: MR Egger Intercept, MR PRESSO, Cochran Q

![Fig. 1 Study overview](image)

**Exposure and Outcome Data**

**Tobacco smoking.** We included lifetime smoking instruments from the recent GWAS of a lifetime smoking index/score (which combined smoking initiation, duration, heaviness and cessation), conducted in a sample of 462,690 current, former and never smokers in the UKB (mean score value 0.359 (standard deviation (SD) = 0.694); sample: 54% female, mean age 56.7 years, 54% never smokers, 36% former smokers, and 11% current smokers53,54. (An SD increase in lifetime smoking index score would be equivalent to smoking 20 cigarettes per day for 15 years and stopping 17 years previously or 60 cigarettes per day for 13 years and stopping 22 years previously)53 (Supplementary Data 2).

**Cannabis use.** We included two cannabis-related instrument sets: cannabis use and AUD. Summary statistics for lifetime cannabis use (a yes/no variable of whether participants reported using cannabis during their lifetime) were obtained from the PGC meta-analysis GWAS of 3 cohorts (International Cannabis Consortium (35,297 respondents, 55.5 percent female, ages 16–87, mean 35.7 years; 42.8 percent had used cannabis); UKB (126,785 respondents, 56.3 percent female, ages 39–72, mean age 55.0 years, 22.3 percent had used cannabis); and 23andMe (22,683 respondents, 55.3 percent female, ages 18–94, mean age 54.0 years, 43.2% had used cannabis))55,56. CUD instruments were obtained from a recent PGC meta-analysis of three cohorts of predominantly European ancestry (PGC, Lundbeck Foundation Initiative for Integrative Psychiatric Research (IPsyCH), and deCODE cohorts, excluding related individuals from PGC family-based cohorts; demographics not available), including 14,808 cases of cannabis abuse or dependence defined as meeting DSM-IIIIR, DSM-IV, DSM-5, or ICD110 codes (depending on study cohort) criteria; the 358,534 controls were defined as anyone not meeting the criteria57,58 (Supplementary Data 3).

**Alcohol consumption.** We included two instrument sets related to alcohol use: drinks per week69, and AUD. Drinks per week instruments were obtained from the GSCAN GWAS meta-analysis of 29 cohorts (941,280 individuals; demographics not available) of predominantly white European ancestry69,70. Given the varied cohort methods used to measure alcohol consumption (binned, normalized, etc.), the data was log transformed: thus, the effect estimate is measured in log transformed drinks per week49 (Supplementary Data 4). For the AUD instrument set, we used the Psychiatric Genomics Consortium (PGC) GWAS meta-analysis of 28 cohorts (51.6% female, 8485 cases, 20,657 controls) of predominantly European ancestry71,72. AUD was diagnosed by either clinician rating or semi-structured interview using DSM-IV criteria including the presence of at least three of seven alcohol-related symptoms (withdrawal, drinking larger amounts/drinking for longer time, tolerance, desire or attempts to cut down drinking, giving up important activities to drink, time related to drinking, or continued alcohol consumption despite psychological and/or physical problems)73 (Supplementary Data 4).

For the multivariable MR (MVMR) analyses, we concatenated independent instrument sets for alcohol use, cannabis use and lifetime smoking, and also AUD, CUD, and lifetime smoking, clumping the resulting two multivariable (MV)
Table 6 Study data sources.

| Phenotypes: | Consortium | First author (Year) | Sample size | N cases | N controls | Population |
|-------------|------------|---------------------|-------------|---------|------------|------------|
| Exposures:  |            |                     |             |         |            |            |
| Drinks per week | GSCAN       | Liu (2019)           | 941,280     | NA      | NA         | European   |
| Alcohol use disorder | PGC         | Walters (2019)       | 28,757      | 8485    | 20,272     |            |
| Lifetime smoking | MRC-IEU     | Wootton (2019)       | 462,690     | NA      | NA         |            |
| Cannabis Use | PGC         | Pasman (2018)        | 162,082     | 43,380  | 118,702    |            |
| Cannabis Use Disorder | PGC       | Johnson (2020)       | 358,534     | 14,804  | 343,726    |            |
| Coronary Artery Disease | CARDIoGRAMplusC4D | van der Harst (2017) | 547,261     | 122,733 | 424,528    |            |
| Type 2 Diabetes | DIAGRAM     | Xue (2018)           | 590,283     | 50,721  | 539,562    |            |
| Obesity Class I | GIANT       | Berndt (2013)        | 98,697      | 32,858  | 65,839     |            |
| Obesity Class II | GIANT       | Berndt (2013)        | 725,46      | 9889    | 62,657     |            |
| Obesity Class III | GIANT      | Berndt (2013)        | 50,364      | 2896    | 47,468     |            |
| COVID-19 outcomes (Round 5): |            |                     |             |         |            |            |
| Very severe respiratory confirmed COVID-19 vs. population | COVID 19-hg | —                   | 707,407     | 4606    | 702,801    | European   |
| Very severe respiratory confirmed COVID-19 vs. population (excluding UKB) | COVID 19-hg | —                   | 378,521     | 4297    | 374,224    | European   |
| Hospitalized vs. not hospitalized COVID-19 | COVID 19-hg | —                   | 16,645      | 4829    | 11,816     | European   |
| Hospitalized vs. not hospitalized COVID-19 (excluding UKB) | COVID 19-hg | —                   | 10,365      | 3159    | 7,206      | European   |
| Hospitalized COVID-19 vs. population | COVID 19-hg | —                   | 1,206,629   | 9373    | 1,197,256  | European   |
| Hospitalized COVID-19 vs. population (excluding UKB) | COVID 19-hg | —                   | 876,382     | 7703    | 868,679    | European   |
| COVID-19 vs. population | COVID 19-hg | —                   | 1,588,783   | 29,071  | 1,559,712  | European   |
| COVID-19 vs. population (excluding UKB) | COVID 19-hg | —                   | 1,257,316   | 22,581  | 1,231,135  | European   |
| FinnGen outcomes (Release 5): |            |                     |             |         |            |            |
| Acute Upper Respiratory Infections | FinnGen   | —                   | 218,792     | 35,847  | 182,945    | European   |
| Asthma related acute respiratory infections | FinnGen   | —                   | 142,793     | 7348    | 135,445    | European   |
| Asthma/COPD (Kela code 203) | FinnGen   | —                   | 208,167     | 21,444  | 186,723    | European   |
| Asthma related infections | FinnGen | —                   | 218,792     | 58,925  | 159,867    | European   |
| Asthma-related pneumonia or sepsis | FinnGen | —                   | 140,994     | 5545    | 135,449    | European   |
| Asthma-related pneumonia | FinnGen | —                   | 140,981     | 5532    | 135,449    | European   |
| Bacterial pneumonia | FinnGen | —                   | 196,855     | 7987    | 188,868    | European   |
| Bronchiitis | FinnGen | —                   | 218,792     | 27,361  | 191,431    | European   |
| Chronic Lower Respiratory Diseases | FinnGen | —                   | 218,792     | 32,069  | 186,723    | European   |
| Acute nasopharyngitis (common cold) | FinnGen | —                   | 185,198     | 2253    | 162,945    | European   |
| Influenza | FinnGen | —                   | 193,130     | 4262    | 188,868    | European   |
| Influenza and Pneumonia | FinnGen | —                   | 218,792     | 29,924  | 188,868    | European   |
| All Pneumonia | FinnGen | —                   | 218,798     | 27,376  | 191,422    | European   |
| Viral Pneumonia | FinnGen | —                   | 189,568     | 700     | 188,868    | European   |

COVID-19 outcomes. We used summary GWAS statistics from the COVID-19 Host Genetics Initiative (COVID-19 hg) meta-analysis round 5a (18 January 2021 release date) (https://www.covid19hg.org/results) for four COVID-19 phenotypes in cohorts of European ancestry, both including and excluding the UKB cohorts for sensitivity analyses (N cases; N controls): very severe respiratory confirmed COVID-19 versus population (4606; 702,801); very severe respiratory confirmed COVID-19 versus population excluding UKB cohorts (4297; 374,224); hospitalized versus not hospitalized COVID-19 (4829; 11,816); hospitalized versus not hospitalized COVID-19 excluding UKB cohorts (3159; 7206); hospitalized COVID-19 versus population excluding UKB cohorts (7703; 868,679); COVID-19 versus population (29,071; 1,559,712); COVID-19 versus population excluding UKB cohorts (22,581; 1,231,135) (demographics not available) (Fig. 1; Table 6; Supplementary Data 1).

Other respiratory infection and disease outcomes. We used data from FinnGen Release 5 (released to public, 11 May 2021) for additional respiratory-related outcomes45, including acute upper respiratory infections, asthma related acute respiratory infections, pneumonia, influenza, bronchitis, chronic lower respiratory diseases, and acute nasopharyngitis (common cold) (N ≤ 218,792) (Fig. 1; Table 6; Supplementary Data 1). FinnGen is a public-private partnership incorporating genetic data for disease endpoints from Finnish biobanks and Finnish health registry EHR86. Detailed documentation is provided on the FinnGen website (https://finngen.giiboo.fi/documentation/).

Sample independence. Participant overlap in samples used to estimate genetic associations between exposures and outcomes can increase weak instrument bias (WIB) in MR analyses86,87, but to a lesser extent with large biobank samples (including UKB and deCODE). Given the large size of the overlapping cohorts (e.g., UKB, deCode) (Supplementary Data 1) and the strength of the instruments in

GSCAN GWAS & Sequencing Consortium of Alcohol and Nicotine, MRC-IEU Medical Research Council Integrative Epidemiology Unit, PGC Psychiatric Genomics Consortium, GIANT Genetic Investigation of ANthropometric Traits, CARDIoGRAMplusC4D Coronary Artery Disease Genome-wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (CAD) Genetics. DIAGRAM, DiAbeetes Genetics Replication and Meta-analysis, COVID-19 coronavirus disease 2019.
both directions (F statistics > 10; Supplementary Data 2–4), considerable WIB would not be expected. We have conducted analyses for COVID-19 outcomes using COVID-19 GWAS performed both including and excluding UKB cohorts.

Statistics and reproducibility. For SVMR analyses, we used inverse-variance weighted MR (MR IVW) as the main analyses, supplemented by MR Egger, weighted median, and weighted mode methods. These are complementary robust methods developed to estimate consistent causal effects under weaker assumptions than MR IVW to assess evidence of causal effects for each of alcohol, cannabis and tobacco use and use disorders on infectious disease outcomes, and evaluate the sensitivity of the analyses to different patterns of violations of IV assumptions. Consistency of results across methods strengthens an inference of causality. For MVMR analyses, we used the multivariable extensions of MR IVW, MR Egger, and MR median. We used the MR Egger intercept test, Cochran Q heterogeneity test, and multivariable extensions thereof, to evaluate heterogeneity in instrument effects, as heterogeneity may indicate violations of IV assumptions. The MRpleiotropy residual sum and outlier (MR PRESSO) global test, and multivariable extensions thereof, were used to facilitate identification and removal of outlier instruments to correct potential directional horizontal pleiotropy and resolve detected heterogeneity. For SVMR, we also used, alternatively, Generalized single variable parameter summary-data based MR (GSMR) to identify and remove instruments with heterogeneous causal estimates suspected to be invalid instruments with apparent pleiotropic effects on both exposure and outcome disease (using the recommended default level of heterogeneity in dependent instruments-outlier threshold of 0.01) to retain sufficient power to detect heterogeneity. We used the SVMR Steiger directionality test to test the causal direction between the hypothesized exposure and outcomes. We also performed a leave-one-out analysis to evaluate the potential SNPs within each instrument that may be high influence points.

For MR, in the multivariable extension of the MR PRESSO global test, we used the multivariable extension of the MR Lasso method, which applies lasso-type penalization to the direct effects of the instruments on the outcome disease: the so-called post-lasso estimate is obtained by performing MR IVW using only those instruments identified as valid instruments (tuning parameter specified after default heterogeneity stopping rule). Analyses were carried out using TwoSampleMR, version 0.5.5.58, MendelianRandomization, version 0.5.0, in the R environment, version 4.0.2; the GSMR method was implemented in the GCTA (Genome-wide Complex Trait Analysis) software (https://cnsgenomics.com/software/gcta/#GSMR).

Reported results and interpretation of findings. MR IVW odds ratios (OR) with 95% CI, or unit increase in the exposures (e.g., per unit increase of log-transformed alcoholic drinks per week or lifetime smoking index), with P-values derived from two-sided tests, corrected for outlier or invalid variants, are presented in Tables 1–5. For our COVID-19 analyses, we used a two-sided a of 0.0025 (based on comparing four COVID-19 outcomes and five substance use exposures) and for the other infectious disease outcomes, a threshold of 0.0071 (based on comparing 14 FinnGen infectious respiratory diseases and five substance use exposures) as a heuristic that allows for follow-up analyses for a plausible number of findings. In assessing consistency and robustness, we looked for estimates substantially agreeing in direction and magnitude (overlapping confidence intervals) across then four complementary MR methods used. We evaluate evidence strength based upon the effect magnitude and direction, the 95% confidence interval of that effect, and the P-value.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
All analyses were based upon publicly available data. Single-variable MR and multivariable MR instrument datasets for each substance use behavior required to replicate the findings of this study are available in the Supplemental Data files. Full COVID-19 GWAS summary-level data is available at https://www.covid19gwag.org/results. FinnGen data are available at the Finnish Genomics Consortium database data at https://finn.gen. psychnl.uhn.utoronto/index.php/GSCAN; cannabis use disorder and alcohol use disorder data are available through the Psychiatric Genomics Consortium data portal: https://www.med.uic.edu/pgc/download-results/; and the cannabis use data are available through the International Cannabis Consortium at: https://www.nuhs.edu/biostat/research/group-pages/substance-use-addiction-food-sal/mva-sal/genetics/international-cannabis-consortium-icc/. Coronary artery disease and obesity summary statistics are available through the Cardiovascular Disease Knowledge Portal: https://cvd.hugeamp.org/. Type 2 Diabetes summary-level data is available Type 2 Diabetes Knowledge Portal: https://t2d.hugeamp.org.

Code availability
We used TwoSampleMR and MendelianRandomization R packages to perform Mendelian Randomization analysis. These packages are publicly available at https://github.com/MRCIEU/TwoSampleMR, and https://github.com/cran/MendelianRandomization, respectively.

Received: 2 February 2021; Accepted: 1 September 2021; Published online: 28 October 2021

References
1. Puntmann, V. O. et al. Outcomes of cardiovascular magnetic resonance imaging in patients recently recovered from coronavirus disease 2019 (COVID-19). JAMA Cardiol. 5, 1265–1273 (2020).
2. Shi, S. et al. Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. JAMA Cardiol. 5, 802–810 (2020).
3. Nishiga, M., Wang, D. W., Han, Y., Lewis, D. B. & Wu, J. C. COVID-19 and cardiovascular disease: from basic mechanisms to clinical perspectives. Nat. Rev. Cardiol. 17, 543–559 (2020).
4. Troeger, C. et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Infect. Dis. 18, 1191–1210 (2018).
5. Soriano, J. B. et al. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet. Respiratory Med. 5, 691–706 (2017).
6. Adams, P. F., Hendershot, G. E. & Marano, M. A. Current estimates from the National Health Interview Survey, 1996. Vital Health Stat. 10, 1–203 (1999).
7. Simet, S. M. & Sisson, J. H. Alcohol’s effects on lung health and immunity. Alcohol Res. Curr. Rev. 37, 199–208 (2015).
8. Tashkin, D. P. Effects of Marijuana smoking on the lung. Ann. Am. Thorac. Soc. 10, 239–247 (2013).
9. Jiang, C., Chen, Q. & Xie, M. Smoking increases the risk of infectious diseases: a narrative review. Tob. Induc. Dis. 18, 60–60 (2020).
10. Simou, E., Leonardis-Bee, J. & Britton, J. The effect of alcohol consumption on the risk of ARDS: a systematic review and meta-analysis. Chest 158, 58–68 (2018).
11. Moir, D. et al. A comparison of mainstream and sidestream marijuana and tobacco cigarette smoke produced under two machine smoking conditions. Res. Toxicol. 21, 191–197 (2001).
12. Ribeiro, L. I. G. & Ind, P. W. Effect of cannabis smoking on lung function and respiratory symptoms: a structured literature review. Respiratory Med. 26, 16071 (2016).
13. Tavejo-Nunez, G., Kolls, J. K. & de Wit, M. Alcohol use as a risk factor in infections and healing: a clinician’s perspective. Alcohol Res. 37, 177–184 (2015).
14. Farhoudian, A. et al. A global survey on changes in the supply, price and use of illicit drugs and alcohol, and related complications during the 2020 COVID-19 pandemic. Front. Psychiatry 12, 646206 https://doi.org/10.3389/fpsyg.2021.646206 (2021).
15. Services, T3AAMHSAUSDOHAla. T Key Substance Use and Mental Health Indicators in the United States: results from the 2018 National Survey on Drug Use and Health. https://www.samhsa.gov/data/sites/default/files/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf. Published 2018 Accessed December 5 2020.
16. Wang, Q. Q., Kaelber, D. C., Xu, R. & Volkow, N. D. COVID-19 risk and outcomes in patients with substance use disorders: analyses from electronic health records in the United States. Mol. Psychiatry 26, 30–39 (2020).
17. Smith, G. D. & Ebrahim, S. Epidemiology—is it time to call it a day? Int. J. Epidemiol. 30, 1–11 (2001).
18. Smith, G. D. & Ebrahim, S. ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? Int. J. Epidemiol. 32, 1–22 (2003).
19. Evans, D. M., Davey & Smith, G. Mendelian randomization: new applications in the coming age of hypothesis-free causality. Ann. Rev. Genomics Hum. Genet. 16, 327–350 (2015).
20. Sekula, P., Del Greco, M. F., Pattaro, C. & Kotting, A. Mendelian randomization as an approach to assess causality using observational data. J. Am. Soc. Nephrol. 27, 3253–3265 (2016).
21. Goldstein, C. E. et al. Ethical issues in pragmatic randomized controlled trials: a review of the recent literature identifies gaps in ethical argumentation. BMC Med. Ethic. 19, 14–14 (2018).
22. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum. Mol. Genet. 23, R89–R98 (2014).
23. Yarmolinsky, J. et al. Association between genetically proxied inhibition of HMG-CoA reductase and epithelial ovarian cancer. JAMA323, 646–655 (2020).
24. Pingault, J.-B. et al. Using genetic data to strengthen causal inference in observational research. Nat. Rev. Genet. 19, 566–580 (2018).
Acknowledgements

We want to acknowledge the participants and investigators of the many studies used in this research without whom this effort would not be possible: the COVID-19 Host Genetics Initiative and the contributors thereto specified at http://www.covid19hg.org/acknowledgments.html, the FinnGen study, and the UK Biobank. We also want to acknowledge the Medical Research Council Integrative Epidemiology Unit (MRC-IEU, University of Bristol, UK), especially the developers of the MRC-IEU UK Biobank GWAS Pipeline. We also want to acknowledge the participants and investigators of FinnGen study. This work was supported by the National Institutes of Health (NIH) intramural funding [ZIA-AA000242 to F.W.L.]; Division of Intramural Clinical and Biological Research of the National Institute on Alcohol Abuse and Alcoholism (NIAAA).

Author contributions

D.B.R. and J.Y. generated the data. All authors contributed to the analysis and interpretation. D.B.R. and F.W.L. conceptualized and designed the study. F.W.L. obtained funding, provided supervision, and critical oversight to data collection and study coordination. D.B.R. and J.Y. wrote the manuscript. All authors contributed to critical editing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-02685-y.

Correspondence and requests for materials should be addressed to Falk W. Lohoff.

Peer review information Communications Biology thanks Zhaozhong Zhu and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editors: Chiea Chuen Khor and George Inglis.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.