Received: 16 August 2021 | Accepted: 2 February 2022

DOI: 10.1111/add.15852

RESEARCH REPORT

ADDICTION SSA

Association of smoking, lung function and COPD in COVID-19 risk: a two-step Mendelian randomization study

Shiu Lun Au Yeung[1] | Albert Martin Li[2] | Baoting He[1] | Kin On Kwok[3,4,5] | C Mary Schooling[1,6]

[1]School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China

[2]Department of Pediatrics, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China

[3]Jockey Club School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China

[4]Stanley Ho Centre for Emerging Infectious Diseases, The Chinese University of Hong Kong, Hong Kong, China

[5]Shenzhen Research Institute of the Chinese University of Hong Kong, Shenzhen, China

[6]School of Public Health and Health Policy, City University of New York, New York, NY, USA

Correspondence
Shiu Lun Au Yeung, School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China.
Email: ayslryan@hku.hk

Funding information
General Research Fund, Research Grants Council, HKSAR Government, Grant/Award Numbers: 14112818, 24104920; Pre-emptive retention/Start up fund, LKS Faculty of Medicine, The University of Hong Kong, Grant/Award Number: N/A; Wellcome Trust Fund, Grant/Award Number: 200861/Z/16/Z; Group Research Scheme of The Chinese University of Hong Kong; Health and Medical Research Fund, Food and Health Bureau, HKSAR Government, Grant/Award Numbers: INF-CUHK-1, 17160302, 18170312

Abstract

Background and Aims: Smoking increases the risk of severe COVID-19, but whether lung function or chronic obstructive pulmonary disease (COPD) mediate the underlying associations is unclear. We conducted the largest Mendelian randomization study to date, to our knowledge, to address these questions.

Design: Mendelian randomization study using summary statistics from genome-wide association studies (GWAS), FinnGen and UK Biobank. The main analysis was the inverse variance weighted method, and we included a range of sensitivity analyses to assess the robustness of the findings.

Setting: GWAS which included international consortia, FinnGen and UK Biobank.

Participants: The sample size ranged from 193 638 to 2 586 691.

Measurements: Genetic determinants of life-time smoking index, lung function [e.g. forced expiratory volume in 1 sec (FEV_1)], chronic obstructive pulmonary disease (COPD) and different severities of COID-19.

Results: Smoking increased the risk of COVID-19 compared with population controls for overall COVID-19 [odds ratio (OR) = 1.19 per standard deviation (SD) of life-time smoking index, 95% confidence interval (CI) = 1.11–1.27], hospitalized COVID-19 (OR = 1.67, 95% CI = 1.42–1.97) or severe COVID-19 (OR = 1.48, 95% CI = 1.10–1.98), with directionally consistent effects from sensitivity analyses. Lung function and COPD liability did not appear to mediate these associations.

Conclusion: There is genetic evidence that smoking probably increases the risk of severe COVID-19 and possibly also milder forms of COVID-19. Decreased lung function and increased risk of chronic obstructive pulmonary disease do not seem to mediate the effect of smoking on COVID-19 risk.

KEYWORDS
Chronic obstructive pulmonary disease, COVID-19, genetics, lung function, Mendelian randomization, smoking

Kin On Kwok and C Mary Schooling are joint senior authors.

INTRODUCTION

The coronavirus disease 19 (COVID-19) pandemic has had a substantial impact on global morbidity and mortality. As of June 2021 there were 176 million confirmed cases, of which 3.8 million (2.16%) had died [1]. Previous observational studies, such as the OpenSAFELY study, showed that people with comorbidities had a higher risk of a COVID-19-related death [2]. However, the association of smoking with COVID-19-related deaths was inconsistent among analytical models, while other studies suggested a possible role of smoking in the development of symptomatic COVID-19 [3]. These discrepancies may imply issues in observational studies, such as confounding, selection bias and problems with interpretation of mutually adjusted statistical models [4,5]. Mendelian randomization studies, a design more robust to confounding than other observational studies due to the use of genetic variants randomly allocated at conception, have suggested that smoking and obesity are probably causes of increased susceptibility to COVID-19 and more severe illness [6]. Although the effect of smoking on increasing risk of severe COVID-19 is probably causal, the underlying mechanisms remain unclear. Given that smoking is strongly linked to poorer lung function and higher risk of chronic obstructive pulmonary disease (COPD), these could be potential mediators underlying the detrimental effect of smoking on COVID-19 and hence be targets of intervention. A recent study suggested that idiopathic pulmonary fibrosis may have a role in severe COVID-19, although the results were sensitive to the choice of analysis [7]. More importantly, whether smoking increases vulnerability to milder COVID-19, which accounts for the majority of COVID-19 cases, is still unascertained but is important to clarify, given the issue of long COVID-19 [8]. Moreover, smoking is a modifiable life-style factor and smoking cessation services are readily available. To clarify these interlocking questions, we conducted a Mendelian randomization study to assess the role of smoking in risk of COVID-19 by severity, as well as assessing the mediating role of lung function and COPD in the association of smoking with risk of COVID-19 using the most up-to-date genome-wide association study (GWAS) from the COVID-19 Host Genetic Initiatives (release 6, 15 June 2021) [9].

METHODS

We used a two-step Mendelian randomization study to assess mediation in a Mendelian randomization framework [10], where we explored separately the association of the exposure with the outcome, exposure with mediators, and mediators with the outcome, using univariable Mendelian randomization [10]. We only considered multivariable Mendelian randomization (MVMR) when there was evidence of mediation from the two-step Mendelian randomization, given the possibility of weak instrument bias and the direction of bias being unpredictable [11]. Similar to a conventional Mendelian randomization study, the instruments for exposures and potential mediators should fulfil three main assumptions [12]. First, the instruments should be related to the corresponding phenotypes (e.g. exposures/mediators). Secondly, the instruments should not be confounded. Thirdly, the instruments' association with mediators and outcomes should occur only via their effects on exposure and mediators. Supporting information, Fig. S1 shows the schematic diagram of this study.

EXPOSURE: LIFE-TIME SMOKING INDEX

Genetic determinants of a life-time smoking index [standard deviation (SD)] was obtained from a GWAS in UK Biobank (n = 462 690 of European ancestry) by Wootton *et al* [13]. Smoking heaviness, duration and smoking initiation from a self-reported questionnaire were used to derive a life-time smoking index using the method proposed by Leffondre *et al*., which could more clearly capture smoking behaviours [14]. The study gave 124 instruments for the life-time smoking index reaching genome-wide significance (P-value < 5×10^{-8}) and not in high linkage disequilibrium (LD) ($r^2 < 0.001$). The GWAS was adjusted for sex and genotype array, and accounted for population stratification and relatedness using a linear mixed model (BOLT-LMM) by the original GWAS investigators.

MEDIATORS: LUNG FUNCTION AND LIABILITY TO DEVELOP COPD

Forced expiratory volume in 1 sec (FEV_1) and forced vital capacity (FVC)

Genetic determinants of FEV_1 (SD) and FVC (SD) were obtained from publicly available UK Biobank GWAS summary statistics (https://gwas.mrcieu.ac.uk/) by Medical Research Council Integrative Epidemiology Unit (MRC IEU), University of Bristol [15,16], as per our previous study (n = 421 986 of European ancestry) [17]. Pre-bronchodilation lung function testing was performed by trained health-care staff using a Vitalograph Pneumotrac 6800 spirometer (Maids Moreton, UK). The GWAS provided 260 instruments for FEV_1 and 320 for FVC which reached genome-wide significance, and not in high LD, although many of these instruments predicted both FEV_1 and FVC. The GWAS was adjusted for sex and genotype array, and accounted for population stratification and relatedness using BOLT-LMM by MRC IEU investigators. We also used the same GWAS as the outcome when assessing the exposure–mediator association.

FEV_1/FVC

Genetic determinants of FEV_1/FVC (SD) were obtained from a GWAS of lung function, which was comprised of the UK Biobank and SpiroMeta (n = 400 102 of European ancestry) [18]. Measurement of FEV_1/FVC in the UK Biobank was described previously and measurement for SpiroMeta varied throughout studies. FEV_1/FVC was inverse

normal-transformed [18]. According to the original GWAS investigators, the UK Biobank GWAS was adjusted for age, age^2, sex, height, smoking status and genotyping array, and accounted for population stratification and relatedness using BOLT-LMM. The studies in SpiroMeta were adjusted for age, age^2, sex, height and principal components using linear regression, and smoking status was controlled for via stratification. A different analytical model was used for studies with related participants. The study gave 97 instruments for FEV_1/FVC reaching genome-wide significance and not in high LD. We also used the same GWAS (n = 321 047) as the outcome when assessing the exposure–mediator association.

Genetic liability to COPD

Genetic determinants of liability to COPD were obtained from a GWAS comprised of the UK Biobank and International COPD Genetics Consortium (35 735 cases and 222 076 controls, mixed ancestry) [19]. COPD was defined using pre-bronchodilator spirometry according to modified GOLD criteria in both studies. The GWAS were adjusted for age, sex, genotyping array (UK Biobank only), smoking pack-years, smoking status and principal components using logistic regression by the original GWAS investigators. The study gave 71 instruments for the liability of COPD reaching genome-wide significance and not in high LD. When assessing the exposure–mediator association, GWAS summary statistics for COPD were required, which were not readily available for the GWAS from which the genetic instruments for COPD were extracted [19]. As such, we used summary statistics for COPD (R5_J10_COPD cases: 6915; controls: 186 723) in FinnGen, a consortium of Biobanks in Finland (Freeze 5), when assessing the exposure–mediator association. COPD was defined based on International Classification of Diseases (ICD) codes retrieved from nation-wide registries in Finland. The two data sources (GWAS and FinnGen) were largely comparable in terms of genetic associations, given that they were both of people of European ancestry. The analyses in FinnGen were adjusted for age, sex, 10 principal components and genotype batch using mixed-model logistic regression [Scalable and Accurate Implementation of GEneralized mixed model (SAIGE)] by the FinnGen investigators.

OUTCOMES: COVID-19 PHENOTYPES

Genetic associations with COVID-19 phenotypes were obtained from the COVID-19 Host Genetics Initiative (https://www.covid19hg.org/, release 6, 15 June 2021), excluding results from 23andMe [9]. COVID-19 Host Genetics Initiative is an ongoing international collaborative initiative focusing upon the genetic determinants of SARS-CoV-2 infection and COVID-19 phenotypes (e.g. severity). The majority of the studies were from people of European descent. COVID-19 phenotypes included all COVID-19 (cases: 112 612; population controls: 2 474 079, 89% Europeans), hospitalized COVID-19 (cases: 24 274; population controls: 2 061 529, 88% Europeans) and very severe respiratory confirmed COVID-19 (cases: 8779; population controls: 1 001 875, 95% Europeans). (https://www.covid19hg.org/). In brief, COVID-19 status was identified differently in each study, including laboratory-confirmed SARS-CoV-2 infection (based on RNA and/or serology tests) or hospital/physician-confirmed COVID-19 via self-reports. Hospitalized COVID-19 was defined as hospitalized laboratory-confirmed SARS-CoV-2 infection where hospitalization was due to corona-related symptoms. Severe COVID-19 was defined as the condition of hospitalized COVID-19 and death/respiratory support (e.g. intubation, continuous airway pressure). The GWAS was adjusted for age, age^2, sex, age × sex, principal components and study-specific covariates by the original GWAS investigators, and SAIGE was recommended as the analysis approach.

A more detailed description of the above-mentioned GWAS, such as mean age (if reported), quality controls and exclusion criteria, can be found in the Supporting information.

Exposure

The primary exposure was genetically predicted life-time smoking index (SD).

Mediators

The mediators were genetically predicted FEV_1, FVC, FEV_1/FVC and genetic liability to COPD.

Outcomes

The primary outcome was overall COVID-19. The secondary outcomes were hospitalized COVID-19 and severe COVID-19.

Confounders

To assess whether the instruments were related to potential confounders of the exposure-outcome relation and hence constituted possible horizontal pleiotropic effects [20], we assessed the relation of life-time smoking instruments (aligned to risk alleles) with possible confounders using relevant GWAS, i.e. body mass index [n = 681 275 from Genetic Investigation of ANthropometic Traits (GIANT) consortium] [21], educational attainment [n = 328 917 from Social Science Genetic Association Consortium (SSAGC), measured as years of education] [22] and alcohol use [n = 537 349 from GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN), measured as drinks per week] [23] using standard meta-analysis with additive random effects, as per our previous study [17]. Afterwards, we conducted bi-directional Mendelian randomization to verify whether the pleiotropy was vertical (consequences of smoking and hence unbiased pathways) or horizontal (causes of smoking and hence biased pathways).

Statistical analyses

To verify whether the instruments were of adequate strength for valid analyses, we assessed the instruments' strength by calculating the total variance explained (R^2) by the instruments and derived the overall *F*-statistic using summary statistics, as per previous studies [17,24]. We also approximated the *F*-statistics for each instrument for genetic liability to COPD using the instrument association with exposure and its standard error [25]. We aligned all genetic associations on the same effect allele, and also used effect allele frequencies for palindromic instruments.

We used inverse variance weighted (IVW) with multiplicative random effects to assess the association of the life-time smoking index (SD) with COVID-19 risk (exposure to outcome) in the main analyses, where we corrected for multiple comparisons using a Bonferroni correction (i.e. 0.05/3 = 0.016). The IVW method regressed instrument–outcome associations on instrument–exposure associations for the instruments, weighted by the inverse of the variance of the instrument–outcome association, with the intercept constrained to 0. We also used IVW to assess the association of the association of life-time smoking index (SD) with lung function and COPD (exposure to mediators), as well as the association of lung function (FEV_1, FVC, FEV_1/FVC in SD) and liability to COPD (log-odds) with COVID-19 risk (mediators to outcome). IVW assumes balanced pleiotropy [26]. We assessed the heterogeneity of instruments using Cochran's *Q*-statistics, where high heterogeneity indicates the presence of invalid instruments, and Mendelian randomization Egger (MR-Egger) intercept and I^2_{GX} to assess evidence of overall horizontal pleiotropy and possible regression dilution in MR-Egger, respectively.[25 27] We also included other sensitivity analyses which rely upon different assumptions (e.g. majority valid; plurality valid; balanced pleiotropy) for valid inferences (as detailed in the Supporting information), such as MR-Egger, weighted median, weighted mode and MR-robust adjusted profile scores (MR-RAPS) [27–30]. Consistency in estimate direction across analyses strengthens the certainty of any observed associations [31].

As our previous study showed that height might bias the Mendelian randomization estimates for FEV_1 and FVC [17], we additionally adjusted for height using multivariable MR (MVMR) when FEV_1, FVC and FEV_1/FVC were used as exposures. This is preferable to older approaches, such as removal of instruments associated with height, which would reduce statistical power. We also assessed the robustness of findings by using different GWAS for the height [Genetic Investigation of ANthropometic Traits (GIANT) and UK Biobank] [32]. We calculated the conditional *F*-statistic as an indicator of weak instrument bias and repeated the analyses using MVMR-Egger, which is more robust than MVMR [17,33,34].

All analyses were performed using R version 4.0.4 (R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; https://www.R-project.org/) and the R packages ('TwoSampleMR') and ('MVMR').[11,16] As the primary research question and analysis plan were not pre-registered, the results should be considered exploratory.

Ethics approval

This study only used publicly available data and hence ethics approval was waived. Details of ethical approval and participant consent for each of the studies that contributed to the GWAS can be found in the original publications.

RESULTS

The instruments for life-time smoking index had an overall *F*-statistic of 49.2 and an R^2 of 1.2%, indicating low evidence for weak instrument bias (Supporting information, Tables S1–S2). There was also no evidence for weak instruments concerning lung function-related traits or COPD (Supporting information, Table S2), and Supporting information, Tables S3–S6 shows the instruments for lung function-related traits and COPD. Smoking-increasing alleles were positively associated with body mass index (BMI) and alcohol use and inversely associated with educational attainment (Supporting information, Table S7). However, the associations were bi-directional and hence indicated a mixture of both horizontal and vertical pleiotropy, i.e. both biasing and downstream pathways. For the main analyses we included up to 118 instruments for life-time smoking index, 254 instruments for FEV_1, 309 instruments for FVC, 94 instruments for FEV_1/FVC and 69 instruments for genetic liability to COPD.

Figure 1a shows that a higher life-time smoking index was associated with a higher risk of all COVID-19 phenotypes for all analyses except MR-Egger. The IVW estimates for all three phenotypes remained after considering multiple comparisons ($OR_{COVID\text{-}19}$: 1.19 per SD, 95%CI 1.11 to 1.27, $P_{COVID\text{-}19}$: 3.5×10^{-7}; $OR_{hospitalized\ COVID\text{-}19}$: 1.67 per SD, 95%CI 1.42 to 1.97, $P_{hospitalized\ COVID\text{-}19}$: 9.7×10^{-10}; $OR_{severe\ COVID\text{-}19}$: 1.48 per SD, 95%CI 1.10 to 1.98, $P_{severe\ COVID\text{-}19}$: 0.009). The most consistent findings were observed for severe COVID-19 where all analyses suggested a positive association. Although there was evidence for heterogeneity based on Cochran's *Q*-test (*P*-value: 0.0019), the MR-Egger intercept test did not indicate the presence of horizontal pleiotropy. Figure 1b shows that a higher life-time smoking index was associated with lower lung function, with consistent evidence across all analyses. Similarly, a higher life-time smoking index was associated with a higher risk of COPD, although there was horizontal pleiotropy (Fig. 1c).

Figure 2a,b shows that FEV_1 and FVC were not associated with COVID-19. While Cochran's *Q*-test indicated the presence of heterogeneity (ranging from 1.4×10^{-10} to 0.0007), the MR-Egger intercept did not indicate the presence of overall horizontal pleiotropy. Additional adjustment for height using multivariable Mendelian randomization (IVW and MR-Egger) did not change the conclusion, regardless of GWAS used for height, although the conditional *F*-statistics were generally low (Supporting information, Table S8).

Figure 3a shows that higher FEV_1/FVC was associated with higher hospitalized and severe COVID-19 risk. Although there was evidence for substantial heterogeneity among instruments in IVW

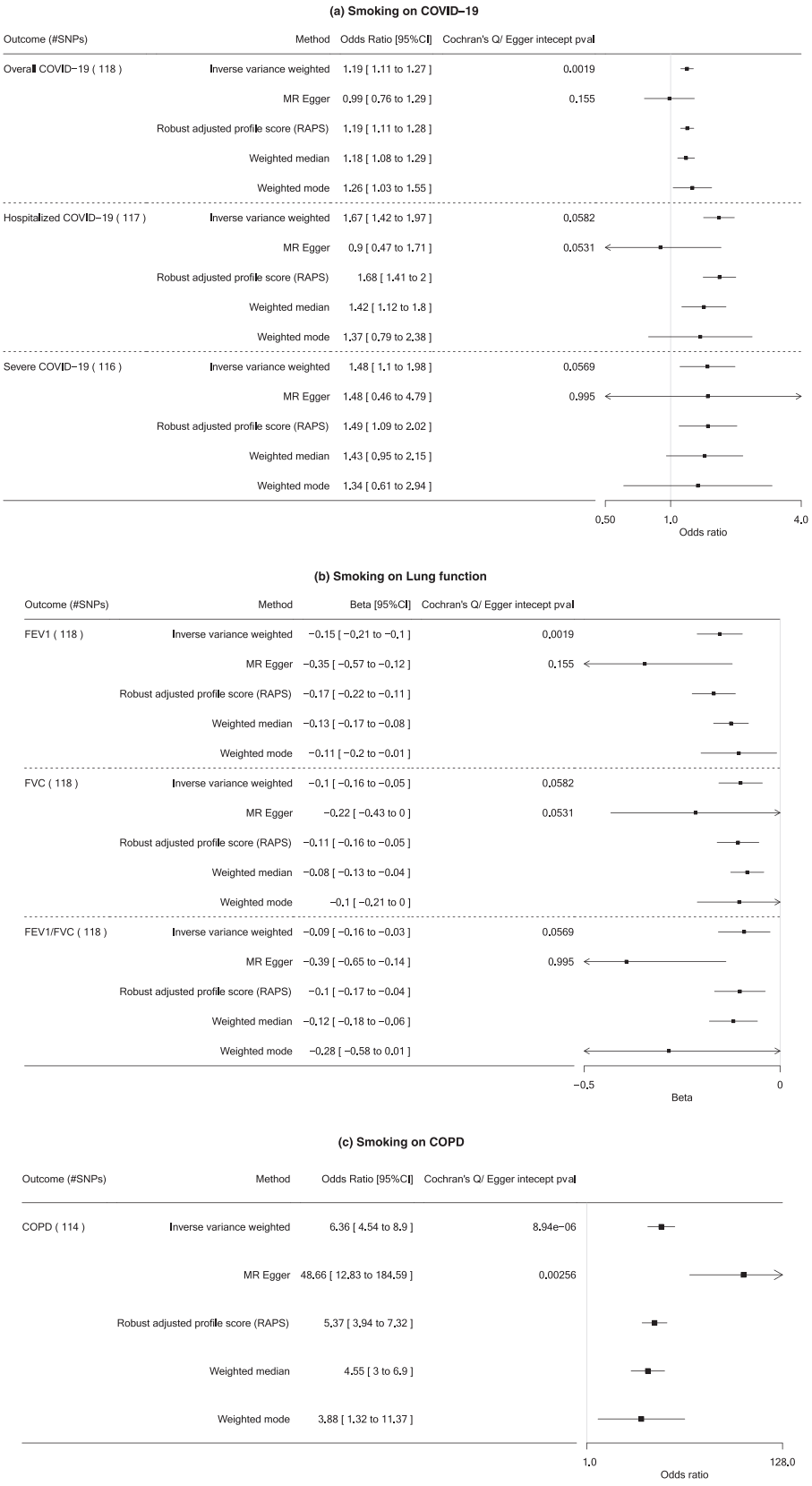

FIGURE 1 Association of life-time smoking index [standard deviation (SD)] in: (a) COVID-19 risk; (b) lung function [forced expiratory volume in 1 second (FEV_1), forced vital capacity (FVC) and FEV_1/FVC)]; and (c) chronic obstructive pulmonary disease (COPD) risk using Mendelian randomization

(Cochran's *Q* *P*-value ranged from 3.74×10^{-17} to 0.00598), the MR-Egger intercept did not suggest directional horizontal pleiotropy. These associations were not present in sensitivity analyses, including the multivariable Mendelian randomization analyses (Supporting information, Table S8). Similar trends were observed for liability to COPD (Fig. 3b), which was associated with lower risk of hospitalized and severe COVID-19, but these associations were not evident in other sensitivity analyses.

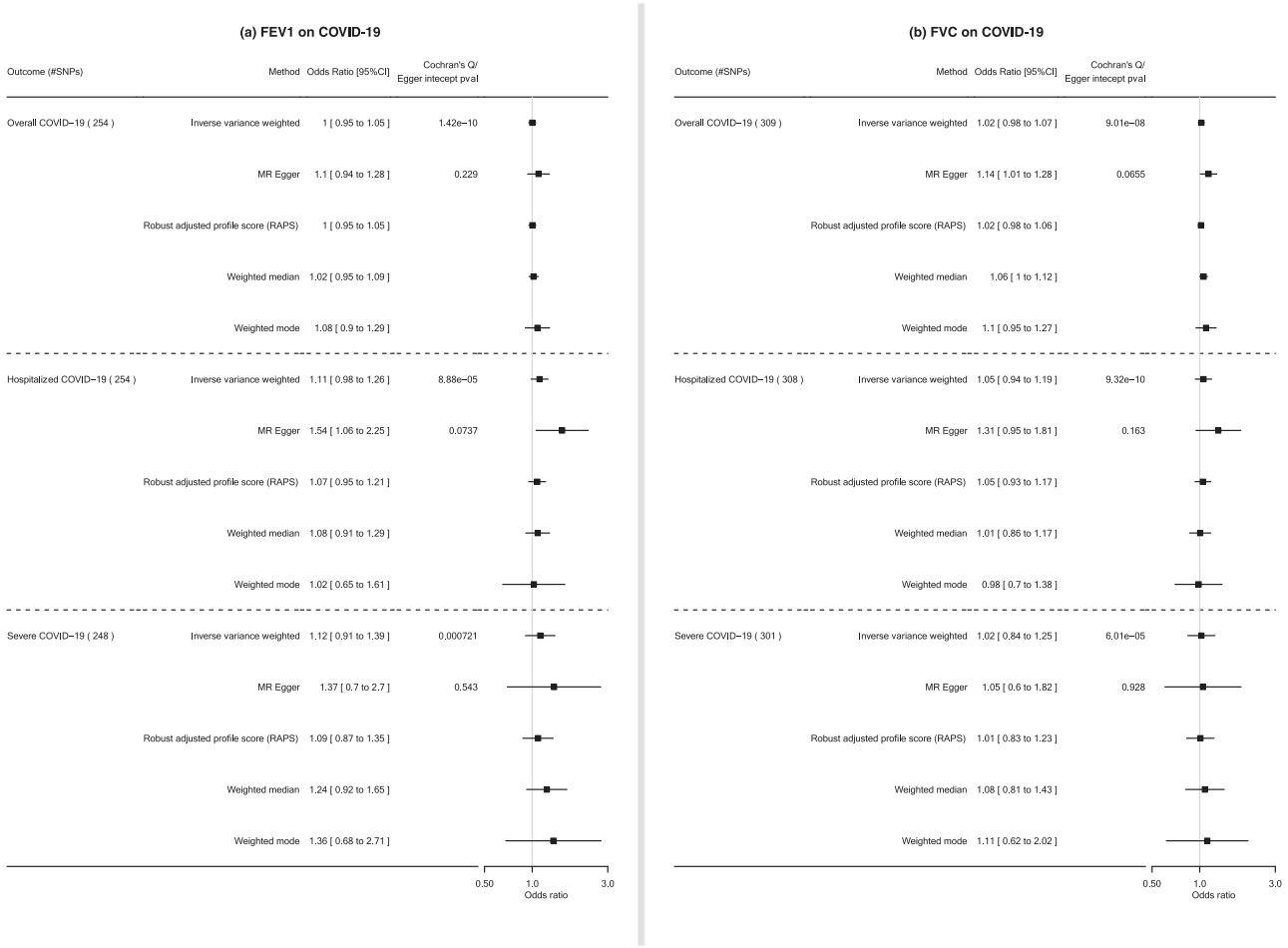

FIGURE 2 Association of: (a) forced expiratory volume in 1 second [FEV_1, standard deviation (SD)] and (b) forced vital capacity [FVC, SD] in COVID-19 risk using Mendelian randomization

Most I^2_{GX} were close to 100%, indicating that there is little evidence of regression dilution (Supporting information, Table S9). However, I^2_{GX} indicated possible regression dilution for analyses of COPD liability in COVID-19 risk (83%) (Fig. 3b).

DISCUSSION

In this large Mendelian randomization study, we confirmed previous studies showing the detrimental role of smoking in severe COVID-19 risk and have added to these studies by showing that smoking probably increases the risk of hospitalized or overall COVID-19 [6]. Whether or not smoking cessation can reduce the burden of COVID-19 among smokers should be considered. As expected, smoking decreased lung function and increased COPD risk. However, our study did not provide strong evidence that these mediate the effect of smoking on COVID-19 risk.

In this study we focused primarily upon the possible mediating role of traits related to lung function. Although previous studies suggest that COPD is a risk factor for poorer COVID-19 prognosis [35], this is not consistent with our findings. We could not rule out the possibility that our study only considered liability to COPD rather than the effect of COPD diagnosis, which is one issue with instruments for diseases. As the COPD GWAS was adjusted for smoking, it is also possible that some genetic variants identified from the GWAS were due to collider bias [36]. Inconsistencies could also be due to confounding by smoking and obesity in previous observational studies, or to selection bias due to inevitably only selecting survivors of COPD and competing risk of COVID-19 to which Mendelian randomization studies are vulnerable [37]. Similarly, we were not able to show strong evidence for an effect of lung function on COVID-19 risk, but Mendelian randomization studies of lung function could be challenged, given the strong pleiotropic effects of height. However, we accounted for this issue using multivariable Mendelian randomization.

As smoking impacts upon a wide spectrum of health outcomes, it is possible that these could be unexplored mediating pathways [38]. Inflammation, in particular, has received much attention regarding COVID-19 because of its relevance to the immune system, and may explain why certain people have a poorer COVID-19 prognosis [39]. Genetic evidence for associations of inflammatory markers with COVID-19 is limited [40]. Nevertheless, randomized controlled trials

FIGURE 3 Association of: (a) forced expiratory volume in 1 second (FEV_1)/ forced vital capacity (FVC); and (b) liability to chronic obstructive pulmonary disease (COPD, log-odds) in COVID-19 risk using Mendelian randomization

of interleukin (IL)-6R antagonists, such as tocilizumab, suggest that it may be an effective treatment [41]. Apart from the physiological role of smoking in COVID-19, the higher risk in observational studies could be due to other factors associated with smoking, such as lower socio-economic position, limited living space and access to facilities hindering adherence to social distancing policies or how cigarettes were being used (repeated removal of masks), which exacerbates the higher risk of COVID-19 among smokers independently of its biological mechanisms [42–44].

Our study suggests that smoking probably increases susceptibility of COVID-19 risk, with the most convincing evidence regarding COVID-19 severity. This study substantiates the importance of tobacco control in mitigating the disease burden due to COVID-19, which could narrow COVID-19-induced inequalities [43]. The use of more innovative means, such as emphasizing vulnerability to COVID-19 as another hazard of smoking, to improve messaging of smoking cessation could be further explored [45].

Although we used a design which is more robust to residual confounding than observational studies, there are some limitations. First, Mendelian randomization requires three main assumptions. We chose strong instruments which were derived from a large GWAS to reduce the risk of weak instrument bias. A low R^2 of the instruments can lead to false negatives, although our main analyses indicated detrimental effects of smoking on COVID-19 risk and severity (Fig. 1) [46]. We could not rule out the possibility of biases due to violation of the instrumental variable assumptions, such as association of instruments with confounders, where smoking-increasing alleles were associated with confounders which could increase the risk of COVID-19, such as lower educational attainment and higher BMI. These may exaggerate the overall detrimental effect of smoking. The majority of the sensitivity analyses are consistent with the main analyses regarding the detrimental effect of smoking in COVID-19 except MR-Egger. While the underlying reasons are not clear, the MR-Egger method is also particularly sensitive to outliers and has lower statistical power. Furthermore, MR-Egger requires instrument strength independent of direct effect assumption which could be violated, as our instruments were related to confounders. These may explain the different results compared to results from other sensitivity analyses, such as the weighted median and MR-RAPS which rely upon other assumptions. Secondly, we used a two-step Mendelian randomization design to explore possible mediation, and as there was no strong evidence for an effect of lung function or COPD in COVID-19 risk we did not assess mediation using other approaches as verification, such as multivariable Mendelian randomization, which were more prone to weak instrument bias, as indicated by

the conditional *F*-statistics. As we used two-sample Mendelian randomization we were not able to explore other phenotypes, such as smoking heaviness, which requires stratification by smoking status and hence access to individual level data. Nevertheless, a recent Mendelian randomization study in the UK Biobank suggested that smoking heaviness among smokers was also related to a higher risk of COVID-19. Taking into account existing evidence from previous Mendelian randomization studies, it is likely that smoking increases COVID-19 risk [47]. Thirdly, our study was based on European populations. However, many previous observational studies of smoking have shown consistent effects globally, so it is unlikely that the harmful effect of smoking on COVID-19 differs by ethnicity [48], and similar associations were seen in observational studies among Chinese people [49]. Nevertheless, additional studies in non-European populations would be valuable to further confirm the harm of smoking in COVID-19. Fourthly, we could not exclude the possibility that some of the controls were cases who were asymptomatic or who only had mild symptoms, which would bias our estimates towards null due to non-differential misclassification. The current definition of severe COVID-19 excluded people who died of COVID-19 without being hospitalized (e.g. frail older people living in elderly facilities), which may introduce bias. Although smoking probably increased the risk of severe COVID-19, it would be worthwhile exploring its role in overall COVID-19 mortality in future studies. Lastly, smoking is difficult to measure precisely, although we demonstrated the known effects of smoking on reduced lung function and increased COPD risk.

Our Mendelian randomization study provides genetic evidence that smoking probably increases the risk of severe COVID-19 and possibly milder forms of COVID-19. Our study provides credible evidence on the harmful effect of smoking in COVID-19 risk and supports the importance of tobacco control in reducing the associated disease burden. Future studies should be conducted to explore the mechanistic pathways in multiple dimensions (i.e. social/biological/behavioural) underpinning these associations to reduce the burden of diseases arising from COVID-19 in smokers.

ACKNOWLEDGEMENTS

This study was supported by the pre-emptive retention/start up fund, LKS Faculty of Medicine, The University of Hong Kong (S.L.A.Y.). Summary data on COVID-19 were contributed by the COVID-19 Host Genetics Initiative and were downloaded from https://www.covid19hg.org. Summary data on life-time smoking were contributed by Wootton *et al*. and were downloaded from https://data.bris.ac.uk/data/dataset/10i96zb8gm0j81yz0q6ztei23d. Summary data on FEV_1 and FVC were extracted from the genome-wide association study results in the UK Biobank, available in the IEU GWAS database (https://gwas.mrcieu.ac.uk/). Summary data on FEV_1/FVC (contributed by Shrine *et al*.) and BMI (contributed by GIANT consortium) were available in the IEU GWAS database. Summary data on educational attainment was contributed by the Social Science Genetic Association Consortium and were downloaded from https://www.thessgac.org/data. Summary data on alcohol use were contributed by the GWAS and Sequencing Consortium of Alcohol and Nicotine use and were downloaded from https://genome.psych.umn.edu/index.php/GSCAN. We would also like to acknowledge the participants and investigators of the FinnGen study. K.O.K. acknowledges support from Health and Medical Research Fund (reference numbers INF-CUHK-1, 17160302, 18170312), General Research Fund (reference numbers 14112818, 24104920), Wellcome Trust Fund (United Kingdom, 200861/Z/16/Z) and Group Research Scheme of The Chinese University of Hong Kong.

DECLARATION OF INTERESTS

None.

AUTHOR CONTRIBUTIONS

Shiu Lun Au Yeung: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; software; supervision; visualization. **Albert Martin Li:** Supervision. **Baoting He:** Methodology. **Kin On Kwok:** Conceptualization; formal analysis; funding acquisition; investigation; methodology; visualization. **C Mary Schooling:** Investigation; methodology; supervision.

ORCID

Shiu Lun Au Yeung https://orcid.org/0000-0001-6136-1836
Albert Martin Li https://orcid.org/0000-0002-9314-477X
Baoting He https://orcid.org/0000-0002-9976-7775
Kin On Kwok https://orcid.org/0000-0002-2804-5433
C Mary Schooling https://orcid.org/0000-0001-9933-5887

REFERENCES

1. World Health Organization (WHO) WHO Coronavirus (COVID-19) dashboard (accessed 16 June 2021). https://covid19.who.int
2. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. OpenSAFELY: factors associated with COVID-19 death in 17 million patients. Nature. 2020. 584(7821):430–6. https://doi.org/10.1038/s41586-020-2521-4
3. Hopkinson NS, Rossi N, El-Sayed Moustafa J, Laverty AA, Quint JK, Freidin M, et al. Current smoking and COVID-19 risk: results from a population symptom app in over 2.4 million people. Thorax. 2021. 76:714–22. https://doi.org/10.1136/thoraxjnl-2020-216422
4. Westreich D, Edwards JK, van Smeden M. Comment on Williamson *et al*. (OpenSAFELY): the table 2 fallacy in a study of COVID-19 mortality risk factors. Epidemiology. 2021;32:e1–2.
5. Griffith GJ, Morris TT, Tudball MJ, Herbert A, Mancano G, Pike L, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. Nat Commun. 2020;11(1):5749.
6. Ponsford MJ, Gkatzionis A, Walker VM, Grant AJ, Wootton RE, Moore LSP, et al. Cardiometabolic traits, sepsis, and severe COVID-19: a Mendelian randomization investigation. Circulation. 2020;142:1791–3.
7. Fadista J, Kraven LM, Karjalainen J, Andrew SJ, Geller F, COVID-19 Host Genetics Initiative, et al. Shared genetic etiology between idiopathic pulmonary fibrosis and COVID-19 severity. EBioMedicine. 2021;65:103277.
8. Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS, et al. Post-acute COVID-19 syndrome. Nat Med. 2021;27:601–15.
9. COVID-19 Host Genetics Initiative. 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. 2020;28:715–8.
10. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol. 2012;41:161–76.
11. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. Int J Epidemiol. 2019;48(3):713–27.
12. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.
13. Wootton RE, Richmond RC, Stuijfzand BG, Lawn RB, Sallis HM, Taylor GMJ, et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. Psychol Med. 2020;50:2435–43.
14. Leffondré K, Abrahamowicz M, Xiao Y, Siemiatycki J. Modelling smoking history using a comprehensive smoking index: application to lung cancer. Stat Med. 2006;25(24):4132–46. https://doi.org/10.1002/sim.2680
15. Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, et al. The MRC IEU OpenGWAS data infrastructure. bioRxiv. 2020. https://doi.org/10.1101/2020.08.10.244293
16. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base platform supports systematic causal inference across the human phenome. eLife. 2018;7:e34408. https://doi.org/10.7554/eLife.34408
17. Au Yeung SL, Borges MC, Lawlor DA, Schooling CM. Impact of lung function on cardiovascular diseases and cardiovascular risk factors: a two sample bidirectional Mendelian randomisation study. Thorax. 2022;77(2):164–71. https://doi.org/10.1136/thoraxjnl-2020-215600
18. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nat Genet. 2019;51:481–93.
19. Sakornsakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL, Jackson VE, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. Nat Genet. 2019;51:494–505.
20. Yang Q, Sanderson E, Tilling K, Borges MC, Lawlor DA. Exploring and mitigating potential bias when genetic instrumental variables are associated with multiple non-exposure traits in Mendelian randomization. medRxiv. 2019. https://doi.org/10.1101/19009605
21. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. Hum Mol Genet. 2018;27:3641–9.
22. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016;533:539–42.
23. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet. 2019;51:237–44.
24. Yarmolinsky J, Bonilla C, Haycock PC, Langdon RJQ, Lotta LA, Langenberg C, et al. Circulating selenium and prostate cancer risk: a Mendelian randomization analysis. J Natl Cancer Inst. 2018;110:1035–8.
25. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. Int J Epidemiol. 2016;45(6):1961–74.
26. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Stat Med. 2017;36(11):1783–802. https://doi.org/10.1002/sim.7221
27. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44:512–25.
28. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304–14. https://doi.org/10.1002/gepi.21965
29. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. Ann Stat. 2020;48(3). https://doi.org/10.1214/19-aos1866
30. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46:1985–98.
31. Lawlor DA, Tilling K, Davey SG. Triangulation in aetiological epidemiology. Int J Epidemiol. 2016;45:1866–86.
32. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014;46:1173–86.
33. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol. 2015;181:251–60.
34. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. Stat Med. 2017;36:4705–18.
35. Gerayeli FV, Milne S, Cheung C, Li X, Yang CWT, Tam A, et al. COPD and the risk of poor outcomes in COVID-19: a systematic review and meta-analysis. EClinicalMedicine. 2021;33:100789.
36. Hartwig FP, Tilling K, Davey Smith G, Lawlor DA, Borges MC. Bias in two-sample Mendelian randomization when using heritable covariable-adjusted summary associations. Int J Epidemiol. 2021;50(5):1639–50. https://doi.org/10.1093/ije/dyaa266
37. Mary Schooling C, Zhao JV, Au Yeung SL, Kwok MK. Letter in response to 'Bias in two-sample Mendelian randomization when using heritable covariable-adjusted summary associations'—'Interpreting Mendelian randomization studies pre-adjusted for the heritable covariable survival to recruitment'. Int J Epidemiol. 2021;50(5):1744–1745. https://doi.org/10.1093/ije/dyab126
38. Çolak Y, Afzal S, Lange P, Nordestgaard BG. Smoking, Systemic Inflammation, and Airflow Limitation: A Mendelian Randomization Analysis of 98 085 Individuals From the General Population. Nicotine Tob Res. 2019;21(8):1036–44. https://doi.org/10.1093/ntr/nty077
39. Del Valle DM, Kim-Schulze S, Huang HH, Beckmann ND, Nirenberg S, Wang B, et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. Nat Med. 2020;26:1636–43.
40. Li M, Yeung CHC, Schooling CM. Circulating cytokines and coronavirus disease: a bi-directional Mendelian randomization study. Front Genet. 2021;12:680646.
41. Angriman F, Ferreyro BL, Burry L, Fan E, Ferguson ND, Husain S, et al. Interleukin-6 receptor blockade in patients with COVID-19: Placing clinical trials into context. Lancet Respir Med. 2021;9:655–64.
42. Sun Y, Lam TH, Cheung YTD, Wang MP, Wu Y, Chen J, et al. First report on smoking and infection control behaviours at outdoor hotspots during the COVID-19 pandemic: an unobtrusive observational study. Int J Environ Res Public Health. 2021;18:1031.
43. Jackson SE, Brown J, Shahab L, Steptoe A, Fancourt D. COVID-19, smoking and inequalities: a study of 53 002 adults in the UK. Tob Control. 2021;30(e2):e111–e121. https://doi.org/10.1136/tobaccocontrol-2020-055933

44. Kawachi I. COVID-19 and the 'rediscovery' of health inequities. Int J Epidemiol. 2020;49:1415–8.
45. Grummon AH, Hall MG, Mitchell CG, Pulido M, Sheldon JM, Noar SM, et al. Reactions to messages about smoking, vaping and COVID-19: two national experiments. Tob Control. 2020. https://doi.org/10.1136/tobaccocontrol-2020-055956
46. Freeman G, Cowling BJ, Schooling CM. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. Int J Epidemiol. 2013;42:1157–63.
47. Clift AK, von Ende A, Tan PS, Sallis HM, Lindson N, Coupland CAC, et al. Smoking and COVID-19 outcomes: an observational and Mendelian randomisation study using the UK biobank cohort. Thorax. 2021. 77(1):65–73. https://doi.org/10.1136/thoraxjnl-2021-217080
48. GBD 2019 Tobacco Collaborators. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: A systematic analysis from the global burden of disease study 2019. Lancet. 2021;397:2337–60.
49. Zhao Q, Meng M, Kumar R, Wu Y, Huang J, Lian N, et al. The impact of COPD and smoking history on the severity of COVID-19: a systemic review and meta-analysis. J Med Virol. 2020;92(10):1915–21.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Au Yeung SL, Li AM, He B, Kwok KO, Schooling CM. Association of smoking, lung function and COPD in COVID-19 risk: a two-step Mendelian randomization study. Addiction. 2022;117:2027–36. https://doi.org/10.1111/add.15852