medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

# Testing the association between blood type and COVID-19 infection, intubation, and death

## Authors


- **Michael Zietz**
  0000-0003-0539-630X
  Department of Biomedical Informatics, Columbia University Irving Medical Center

- **Nicholas P. Tatonetti ***
  0000-0002-2700-2597
  Department of Biomedical Informatics, Columbia University Irving Medical Center; Department of Systems Biology, Columbia University Irving Medical Center; Department of Medicine, Columbia University Irving Medical Center;
  Institute for Genomic Medicine, Columbia University Irving Medical Center

* — Corresponding author: nick.tatonetti@columbia.edu (Nicholas P. Tatonetti)




NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .


# Abstract

The rapid global spread of the novel coronavirus SARS-CoV-2 has strained healthcare systems, making the identification of individuals most at-risk a critical challenge. Recent studies identified associations between ABO blood groups and COVID-19. Using observational data on 7,770 SARS-CoV-2-tested individuals at New York Presbyterian (NYP) hospital, we find evidence of overall association with ABO blood groups and a beneficial association between Rh-negative blood groups and both infection status and death. We estimate pooled effect sizes using our data with previously-reported data from China and the UK, finding enrichment of B and depletion of O blood groups among infected individuals. Finally, we show that blood type's effects are not explained by other risk factors (age, sex, race, ethnicity, hypertension, diabetes mellitus, obesity, and cardiovascular and respiratory diseases). To our knowledge, this is the first report of an independent Rh(D) association with COVID-19.


# Background

The novel Coronavirus disease (COVID-19, caused by the SARS-CoV-2 virus) has spread rapidly across the globe and has caused over 10,000,000 confirmed infections and over 500,000 deaths worldwide as of June 30, 2020[1]. A number of risk factors for COVID-19 infection, morbidity, and mortality are known, including age, sex, and several chronic conditions and laboratory findings[2]. Recently, a study on COVID-19 patients in Wuhan and Shenzhen, China discovered associations between ABO blood types and infection[3]. They found that the odds of having COVID-19 were increased among A and decreased among O blood groups relative to the general populations of Wuhan and Shenzhen. Previous work has identified similar

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

associations between ABO blood groups and a number of different infections or disease severity following infections, including for SARS-CoV-1[4], *P. falciparum*[5], *H. pylori*[6], Norwalk virus[7], hepatitis B virus[8], and *N. gonorrhoeae*[9].

Within the United States, New York City has become the epicenter of the pandemic, with over 212,000 cases and over 18,000 deaths as of June 30, 2020[10]. We sought to understand the association between SARS-CoV-2 infection/COVID-19 and blood type using electronic health record (EHR) data from New York-Presbyterian/Columbia University Irving Medical Center (NYP/CUIMC) hospital in New York, USA. We compared both ABO and Rh(D) blood types, and we investigated infection status and two severe COVID-19 outcomes: intubation and death. We performed a multivariate analysis of our results to evaluate potential confounding due to population stratification and risk factors, and we meta-analyzed our results in combination with data from the UK Biobank and previously-reported data from China. To the best of our knowledge, this is the first study to evaluate associations between both ABO and Rh(D) blood groups and COVID-19 infection, morbidity, and mortality.

## Results

We determined blood groups for SARS-CoV-2-tested individuals using laboratory measurements recorded in the NYP/CUIMC EHR system. Excluding individuals with contradictory blood group measurements, we found 7,770 individuals (Table 1) with known blood groups who received a SARS-CoV-2 swab test (either positive or negative result). Individuals with a single positive SARS-CoV-2 lab test were considered COV+, even if they had previous or subsequent negative tests. We evaluated associations between blood groups and

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

outcomes using four pairs of populations: COV+ vs COV-, COV+ vs general population (excluding those tested for SARS-CoV-2), COV+/Intubated vs COV+/Not intubated, and COV+/Deceased vs COV+/Alive. We report data as of June 15, 2020 and make the most recent data available on GitHub (Methods).

## Overall association

We found significant associations between SARS-CoV-2 test results and both Rh (p=0.00041) and ABO/Rh (p=0.048) blood groups, though not for ABO alone (p=0.34, Supplementary Table 2). Intubation following confirmed infection was significantly associated with ABO (p=0.016), Rh (p=0.021), and ABO/Rh (p=0.0064), while death following confirmed infection was significantly associated with Rh (p=0.0044) and ABO/Rh (p=0.0087), but not ABO (p=0.15) blood groups (Supplementary Table 2). For each comparison cohort pair, we performed Pearson's chi-squared tests using ABO, Rh, and ABO/Rh blood types. Since there were few AB-negative individuals tested for SARS-CoV-2 infection, we excluded AB-negative from all ABO/Rh analyses. Additionally, we found insufficient evidence to conclude that the blood group distribution among all individuals tested for SARS-CoV-2 is different from the general population at NYP/CUIMC (ABO: p=0.64, Rh: p=0.36, Supplementary Table 2).

## Individual blood group associations

Next, we tested each individual blood type against all others (within the same ABO, Rh, or ABO/Rh system) for association with each outcome using logistic regression with and without adjustments for demographics and clinical risk factors. Without adjustments, Rh(D) groups were significantly associated with SARS-CoV-2 test result, intubation, and death (Figure 1,

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Supplementary Table 3), while the only significant ABO blood group association was between blood group A and intubation (OR 0.762, 95% CI [0.620-0.937], p=0.0099). Adjusting for demographics and comorbidities did not significantly change effect size estimates (Figure 1, Supplementary Table 3), though Rh(D) associations were slightly attenuated, and Rh(D) was no longer significantly associated with intubation (p=0.084). Meanwhile, estimates for associations between B and positive test result and between A and AB and intubation shifted slightly to reach significance at the 5% level (Figure 1, Supplementary Table 3). Several ABO/Rh groups were associated with SARS-CoV-2/COVID-19 outcomes (Figure 1, Supplementary Table 3). However, the effect sizes appear to be independent combinations of ABO and Rh effects. Moreover, our data provide insufficient evidence to conclude that an association exists between ABO and Rh groups (Pearson's chi-squared test, p=0.088, Supplementary Table 2), which would have evidenced confounding.

## Multivariate analyses

To better understand the relationships among blood groups, demographics, comorbidities, and SARS-CoV-2/COVID-19 outcomes, we performed two additional analyses. First, we found a number of significant associations between blood groups and risk factors using blood group ~ risk factors logistic regressions for each blood group (Supplementary Table 4, Supplementary Figure 2). Second, we evaluated whether blood groups provide significant additional information on outcomes beyond risk factors. We verified that the risk factors predict outcomes (Supplementary Table 5) and compared logistic regression model fit when adding blood groups (outcome ~ blood group + risk factors) using analysis of deviance. We found that test results are

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

significantly better predicted with Rh(D) information alongside demographics and comorbidities (p=0.019), but not ABO (p=0.11) or ABO/Rh (p=0.12). All blood group types improved model fits for intubation and death following infection at values reaching or nearly reaching significance at the 5% level (Supplementary Table 5). These results are consistent with our univariate association tests between blood group types (ABO, Rh, ABO/Rh) and outcomes (Supplementary Table 2). We also estimated effect sizes within racial and ethnic strata, finding no significant differences in estimated effect sizes (Supplementary Figure 3). In summary, we find little evidence for conditional independence between outcomes and blood groups given risk factors.

## Meta-analysis

We performed a meta-analysis, comparing our data from New York City to data from the UK Biobank and the data from Wuhan and Shenzhen presented by Zhao et al. [3]. Since Zhao et al. do not report Rh(D) blood groups or negative test results, our meta-analysis analyzed only ABO blood group distributions between COV+ individuals and the general population of each source. We found significant heterogeneity among the meta-analysis sites (Supplementary Figure 1), and the distribution of blood groups in the general population at NYP/CUIMC differs significantly from the distributions in Shenzhen (p-value=1.2e-441), Wuhan (p-value=7.9e-126), and the UK Biobank (p-value=6.1e-1148, Supplementary Table 6).

We used a random effects model to weight and pool effects between four data sources: NYP/CUIMC, Wuhan Jinyintan, Renmin Hospital in Wuhan, Shenzhen Third People's Hospital, and the UK Biobank. We find COV+ odds significantly increased among B (OR 1.11, 95% CI

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[1.03-1.19], p=0.0059) and significantly decreased among O blood groups (OR 0.77, 95% CI [0.65-0.92], p=0.0038). Our meta-analysis finds similar pooled effect sizes as reported by Zhao et al., though heterogeneity among sites precludes rejection of the null hypothesis for A and AB blood groups (p=0.083 and p=0.11, respectively; Table 2, Figure 2, Supplementary Figure 1).

## Discussion

Since both blood groups and risk factors vary across populations, we evaluated associations in the context of demographics and comorbidities. We found that adding blood groups significantly improved the fit of models predicting SARS-CoV-2/COVID-19 outcomes. No effect size estimates were significantly changed by covariate adjustment, suggesting blood groups have an independent effect not captured by other risk factors.

The NYP/CUIMC patient cohort used in our comparisons consists of patients visiting the hospital during the COVID-19 pandemic. This cohort is enriched for SARS-CoV-2-infected and otherwise ill patients compared to the general population of New York, thus representing a better comparison than the entire population. We found concordance between SARS-CoV-2-tested individuals and the general population at NYP/CUIMC in terms of blood type (Supplementary Table 2), though we cannot rule out other differences or biases, especially across meta-analysis sites.

We found significant heterogeneity in blood group distributions between meta-analysis sites (Table 2, Supplementary Figure 1), consistent with previous work indicating differences in blood group distributions between the United States, the United Kingdom, and China[11–13]. However, since only aggregate data were available for Wuhan and Shenzhen, our meta-analysis

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

was unable to evaluate additional differences between sites or among patients. Differences in testing practice through time and between meta-analysis sites introduce additional heterogeneity. Further work is needed to understand how the population of patients tested for infection differs from the general population and whether the effects of blood group on COVID-19 depend on other factors.

Our meta-analysis found evidence for a protective association between O blood groups and SARS-CoV-2 infection, consistent with a similar association discovered for SARS-CoV-1[4]. Guillon et al. provide evidence suggesting human anti-A antibodies may interfere with interactions between the SARS-CoV-1 spike protein and the human ACE2 receptor [14]. Since anti-A antibodies are present in individuals with both B and O blood groups, this result suggests that B and O blood groups could be at lower risk. However, our meta-analysis associations for B and O blood groups are significant in opposite directions, with enrichment of B blood groups among SARS-CoV-2 infected individuals. Further work is needed to understand the mechanistic basis for associations between blood groups and COVID-19.

While Rh(D) information was not available from the other meta-analysis sites, we found consistent evidence for protective associations between Rh negative blood groups and SARS-CoV-2 infection and death in NYP/CUIMC data. Negative Rh blood groups are less common, representing only 9% of individuals in our data, and Rh group associations were consistently moderated by adjustment for covariates (Figure 1), suggesting the potential for confounding due to population stratification or selection bias. Further work is needed to better understand the associations between Rh(D) blood groups and COVID-19.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Multiple comparisons present an important consideration for hypothesis testing, and our analysis involved a number of statistical tests performed in parallel. The significant associations from our meta-analysis (B and O) have p-values less than 0.006, thus remaining less than 0.05 upon Bonferroni's correction. However, covariate-adjusted COV+ B group association from NYP/CUIMC data has a p-value of 0.03, meaning Bonferroni correction moves it above the 5% significance threshold. While these associations are unlikely to have precisely zero true effect, p-values > 0.05 given large sample sizes suggest that the true effect may be small, potentially inconsequential compared to other risk factors.

Our data are preliminary, and we will be better able to assess the relationship between blood group and intubation or death when more patients become tested, intubated, and recovered. In particular, since only a fraction of individuals experience severe disease (e.g. intubation or death), greater sample sizes are necessary to understand these outcomes. As an observational study using EHR data collected during the care of patients—not necessarily with research intent—our results, on their own, should be considered preliminary and should not inform clinical practice or policy.

## Conclusion

In this study we found evidence for association between ABO and Rh blood groups and COVID-19. Using data from NYP/CUIMC, the UK Biobank, and previously-reported data from China, we found evidence for enrichment of B and depletion of O blood groups among SARS-CoV-2 positive patients. Rh(D) positive blood types were associated with both SARS-CoV-2 infection and death following infection. We demonstrated that the associations we

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

found were not explained by confounding due to demographics or several known risk factors. Our results add further evidence to the previously-discovered COVID-19 protective association for O blood type, and we show evidence for additional associations between B and Rh(D) blood groups.

## Acknowledgements

We would like to thank Ben May and Vijendra Ramlall for assistance with data collection and daily updates to the patient data. We would also like to thank Nicholas Giangreco, Undina Gisladottir, Vijendra Ramlall, and Dr. Phyllis Thangaraj for helpful discussions regarding risk factor definitions and Dr. Jason Zucker for helpful clinical insights. MZ is funded by NIH T15 LM007079, and NPT is funded by R35GM131905.

## Author Contributions

MZ and NPT conceived and designed the study. MZ carried out the statistical analysis with advice from NPT. MZ and NPT wrote, revised, and approved the final version of the manuscript.

## Competing Interests

The authors have no competing interests to disclose.

## Methods

Throughout our analysis, individuals with a single positive SARS-CoV-2 lab test are considered COV+, even if they had previous or subsequent negative tests. Blood type at NYP/CUIMC was identified using a number of laboratory measurements (Supplementary

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Materials). We excluded individuals with multiple contradictory blood group measurements. As outcomes we considered confirmed infection, intubation, and death. Because intubation and death occur also for non-COVID-19 reasons (e.g. intubation during surgical anesthesia), we restrict our evaluation of these outcomes to COV+ patients.

We compared blood groups (defined as ABO, Rh, and ABO/Rh) and COVID-19 outcomes using four pairs of populations: COV+ vs COV-, COV+ vs general population (excluding those tested for COVID-19), COV+/Intubated vs COV+/Not intubated, and COV+/Deceased vs COV+/Alive. For each of the test conditions, we performed a Pearson's Chi-squared test to evaluate whether blood group distributions differ between the compared populations. As a test of individual blood groups, we compared each blood group against all others using logistic regression to determine effect sizes for each blood group. For these individual comparisons, we report odds ratios (OR), p-values (two-sided), and 95% odds ratio confidence intervals. Each effect size is reported as both a raw (univariate) and covariate-adjusted (multivariate) estimate.

We evaluated the confounding effects of known risk factors (age, sex, self-reported race and ethnicity, hypertension, diabetes mellitus, obesity, and cardiovascular and respiratory diseases) on associations between blood group and COVID-19 outcomes. Since these analyses were performed at the individual level, we only considered COV+ vs COV-, COV+/Intubated vs COV+/Not intubated, and COV+/Died vs COV+/Alive, leaving out the COV+ vs general population comparison. Risk factor phenotypes were assigned using diagnosis codes, laboratory measurements, and other data available in the EHR (Supplementary Materials).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

First, we evaluated associations between risk factors and blood groups using logistic regressions of risk factors on blood groups (blood group ~ risk factors). Second, we verified that risk factors are collectively predictive of COVID-19 outcomes by comparing the fit of a logistic regression model using risk factors to a null model using only an intercept term. Third, we tested whether blood groups provide additional information on outcomes beyond risk factors by comparing the deviances of a full model (outcome ~ blood group + risk factors) to a nested model using only risk factors (outcome ~ risk factors). Fourth, we tested whether the effects of blood groups are modulated by risk factors by comparing logistic regression coefficients for blood groups between nested (outcome ~ blood group) and full (outcome ~ blood group + risk factors) logistic regression models. In this comparison, the magnitude of blood group coefficients greatly shrinking when risk factors are added would be evidence that outcome is conditionally independent of blood group given risk factors.

We performed a meta-analysis using our data in combination with data from the UK Biobank[15] and from Wuhan and Shenzhen reported by Zhao et al.[3]. These analyses used a random effects model to create pooled estimates of odds ratios for each ABO blood group in comparisons between COV+ individuals and the general populations of New York, Wuhan, Shenzhen, and the UK Biobank without a recorded SARS-CoV-2 test. The NYP/CUIMC distribution of blood groups in the general population was estimated using blood group lab results on 106,528 individuals recorded in the NYP/CUIMC electronic health record (EHR) system between May 2011 and June 2019, excluding results for any individuals later tested for COVID-19 (regardless of result). We then compared the general population blood group

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

distributions between New York, the UK Biobank, and Wuhan and Shenzhen and evaluated the heterogeneity between sites.

Blood type data is not available directly through the UK Biobank. However, ABO blood type can be determined using available genotype data[16,17]. We first removed individuals with more than 10% missing genotypes or mismatched listed and genetic sex, then determined blood groups using the variant-blood-type mapping described by Melzer et al., converted into the variant coding used by the UK Biobank (Supplementary Table 6). We applied the mapping to determine ABO blood types (Supplementary Table 7), and removed individuals whose blood genotype did not correspond to one of the mappings we used. Supplementary Table 8 gives the ABO blood type distribution we determined. More information is available in the Supplementary Materials.

This study is approved by the IRB (#AAAL0601). We use EHR data up to June 15, 2020 and data from the UK Biobank under project ID 41039. We conducted our analyses in the R language, using the meta package [18] for meta-analysis.

## Data availability

While our data from NYP/CUIMC are protected by HIPAA and cannot be released, we have made longer summary statistics available at https://github.com/zietzm/abo_covid_analysis. The source data underlying Figure 1 and Supplementary Figures 2 and 3 are provided as a Source Data file. In addition, we are updating online summary data as additional patient data become available. The manuscript was written openly on GitHub using Manubot [19].

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

# Code availability

All code used in our analysis is available at https://github.com/zietzm/abo_covid_analysis.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

# References


1. World Health Organization. *Coronavirus disease (COVID-19) Situation Report – 162* . (World Health Organization, 2020).
2. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395,** 1054–1062 (2020).
3. Zhao, J. *et al.* Relationship between the ABO Blood Group and the COVID-19 Susceptibility. *medRxiv* (2020). doi:10.1101/2020.03.11.20031096
4. Cheng, Y. *et al.* ABO blood group and susceptibility to severe acute respiratory syndrome. *JAMA* **293,** 1450–1451 (2005).
5. Loscertales, M.-P. *et al.* ABO blood group phenotypes and Plasmodium falciparum malaria: unlocking a pivotal mechanism. *Adv. Parasitol.* **65,** 1–50 (2007).
6. Borén, T., Falk, P., Roth, K. A., Larson, G. & Normark, S. Attachment of Helicobacter pylori to human gastric epithelium mediated by blood group antigens. *Science* **262,** 1892–1895 (1993).
7. Lindesmith, L. *et al.* Human susceptibility and resistance to Norwalk virus infection. *Nat. Med.* **9,** 548–553 (2003).
8. Wang, D.-S. *et al.* ABO blood group, hepatitis B viral infection and risk of pancreatic cancer. *Int. J. Cancer* **131,** 461–468 (2012).
9. Foster, M. T. & Labrum, A. H. Relation of infection with Neisseria gonorrhoeae to ABO blood groups. *J. Infect. Dis.* **133,** 329–330 (1976).
10. New York City Department of Health and Mental Hygiene. COVID-19: Data Summary - NYC Health. *COVID-19: Data* (2020). at

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .


<https://www1.nyc.gov/site/doh/covid/covid-19-data.page>

11. Garratty, G., Glynn, S. A., McEntire, R. & Retrovirus Epidemiology Donor Study. ABO and Rh(D) phenotype frequencies of different racial/ethnic groups in the United States. *Transfusion* **44,** 703–706 (2004).
12. Blood types - NHS Blood Donation. at <https://www.blood.co.uk/why-give-blood/blood-types/>
13. Liu, J. *et al.* Frequencies and ethnic distribution of ABO and RhD blood groups in China: a population-based cross-sectional study. *BMJ Open* **7,** e018476 (2017).
14. Guillon, P. *et al.* Inhibition of the interaction between the SARS-CoV spike protein and its cellular receptor by anti-histo-blood group antibodies. *Glycobiology* **18,** 1085–1093 (2008).
15. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562,** 203–209 (2018).
16. Melzer, D. *et al.* A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet.* **4,** e1000072 (2008).
17. Groot, H. E. *et al.* Genetically determined ABO blood group and its associations with health and disease. *Arterioscler. Thromb. Vasc. Biol.* **40,** 830–838 (2020).
18. Schwarzer, G., Carpenter, J. R. & Rücker, G. *Meta-Analysis with R*. (Springer International Publishing, 2015). doi:10.1007/978-3-319-21416-0
19. Himmelstein, D. S. *et al.* Open collaborative writing with Manubot. *PLoS Comput. Biol.* **15,** e1007128 (2019).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

# Tables

*Table 1:* ***Summary demographics for cohort, stratified by blood group.*** *N is the number of individuals having the given blood type who have a recorded test (positive or negative) for SARS-CoV-2. COV+ gives the number and percent of individuals with a recorded positive test result. COV+/Intubated and COV+/Died report percentages relative to COV+ individuals. Stratification by ABO/Rh is available in Supplementary Table 1.*

| Characteristic | A | AB | B | O | Rh-negative | Rh-positive |
|---|---|---|---|---|---|---|
| N | 2537 | 334 | 1198 | 3701 | 696 | 7074 |
| Median age (IQR) | 58 (37-73) | 58 (37-71) | 57 (37-72) | 56 (36-71) | 56 (36-71) | 57 (37-72) |
| Male sex (%) | 983 (38.7) | 141 (42.2) | 473 (39.5) | 1429 (38.6) | 252 (36.2) | 2774 (39.2) |
| Race - Asian (%) | 40 (1.6) | 10 (3.0) | 59 (4.9) | 66 (1.8) | 9 (1.3) | 166 (2.3) |
| Race - Black/AA (%) | 425 (16.8) | 68 (20.4) | 325 (27.1) | 738 (19.9) | 92 (13.2) | 1464 (20.7) |
| Race - White (%) | 1015 (40.0) | 125 (37.4) | 345 (28.8) | 1166 (31.5) | 325 (46.7) | 2326 (32.9) |
| Race - Other (%) | 634 (25.0) | 77 (23.1) | 253 (21.1) | 1063 (28.7) | 147 (21.1) | 1880 (26.6) |
| Race - Missing (%) | 423 (16.7) | 54 (16.2) | 216 (18.0) | 668 (18.0) | 123 (17.7) | 1238 (17.5) |
| Ethnicity - Hispanic (%) | 1008 (39.7) | 110 (32.9) | 398 (33.2) | 1690 (45.7) | 245 (35.2) | 2961 (41.9) |
| Ethnicity - Non-Hispanic (%) | 1041 (41.0) | 149 (44.6) | 549 (45.8) | 1277 (34.5) | 295 (42.4) | 2721 (38.5) |
| Ethnicity - Other (%) | 27 (1.1) | 7 (2.1) | 12 (1.0) | 58 (1.6) | 10 (1.4) | 94 (1.3) |
| Ethnicity - Missing (%) | 461 (18.2) | 68 (20.4) | 239 (19.9) | 676 (18.3) | 146 (21.0) | 1298 (18.3) |
| Hypertension (%) | 1258 (49.6) | 155 (46.4) | 582 (48.6) | 1788 (48.3) | 334 (48.0) | 3449 (48.8) |
| Cardiovascular diseases (%) | 1723 (67.9) | 226 (67.7) | 771 (64.4) | 2455 (66.3) | 456 (65.5) | 4719 (66.7) |
| Respiratory diseases (%) | 1560 (61.5) | 200 (59.9) | 717 (59.8) | 2269 (61.3) | 422 (60.6) | 4324 (61.1) |
| Diabetes mellitus (%) | 829 (32.7) | 101 (30.2) | 403 (33.6) | 1189 (32.1) | 205 (29.5) | 2317 (32.8) |
| Obesity (%) | 1073 (42.3) | 133 (39.8) | 468 (39.1) | 1591 (43.0) | 284 (40.8) | 2981 (42.1) |
| COV+ (%) | 721 (28.4) | 87 (26.0) | 363 (30.3) | 1035 (28.0) | 157 (22.6) | 2049 (29.0) |
| COV+/Intubated (%) | 167 (6.6) | 31 (9.3) | 109 (9.1) | 281 (7.6) | 29 (4.2) | 559 (7.9) |
| COV+/Died (%) | 161 (6.3) | 26 (7.8) | 71 (5.9) | 247 (6.7) | 21 (3.0) | 484 (6.8) |

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

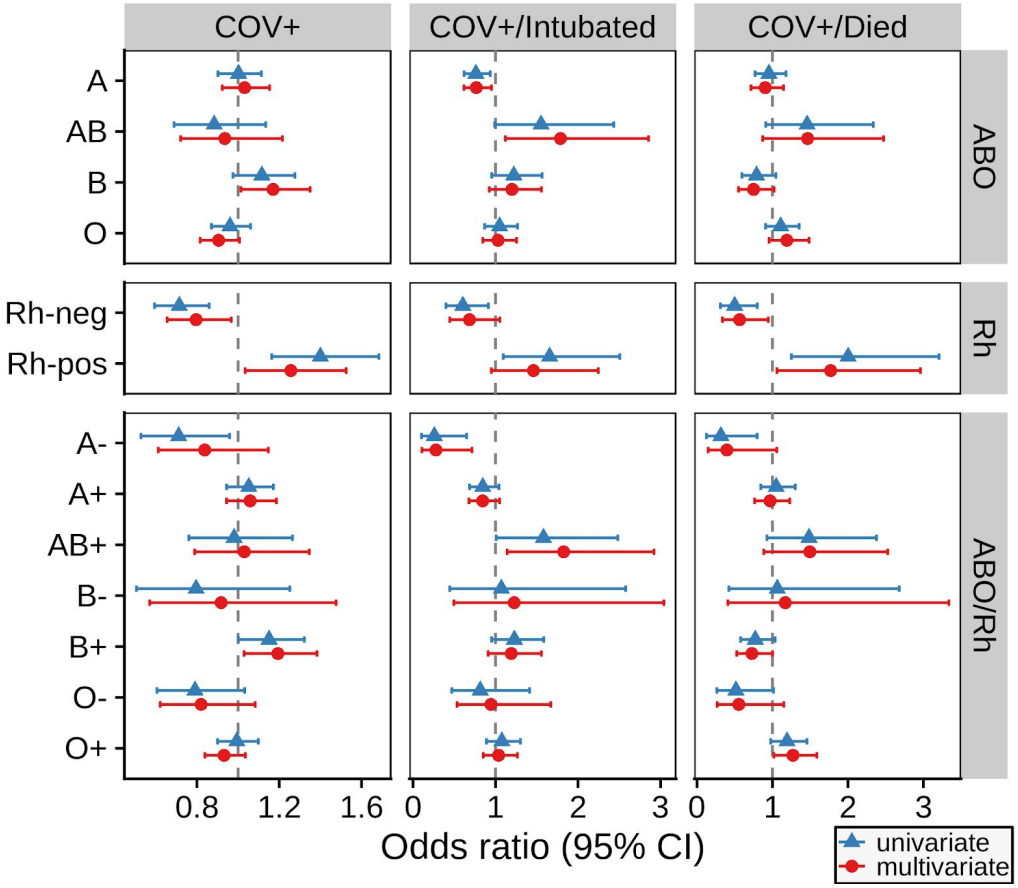
*Table 2:* ***Meta-analysis of data from Wuhan, Shenzhen, and NYP/CUIMC.*** *Distributions of blood groups between New York City data from the NYP/CUIMC EHR system and individuals from Shenzhen (cases from Shenzhen Third People's Hospital, controls from Shenzhen general population), Wuhan (cases from Wuhan Jinyintan Hospital and Renmin Hospital of Wuhan University, controls from Wuhan general population), and the UK Biobank (UKB). Shenzhen and Wuhan data reported by Zhao et al. [3]. Meta-analysis associations are shown for individual ABO blood groups (eg. AB vs not AB) in comparisons of COV+ vs general population using a random effects model. OR, 95% confidence interval (CI), and p-value refer to the pooled effect size estimate (COV+ vs general population) from the random effects model.*

| Site | A | AB | B | O |
|---|---|---|---|---|
| NYP/CUIMC COV+ | 32.7% (721) | 3.9% (87) | 16.5% (363) | 46.9% (1035) |
| NYP/CUIMC controls | 32.7% (34831) | 4.2% (4492) | 14.9% (15904) | 48.2% (51301) |
| Wuhan Jinyintan COV+ | 37.7% (670) | 10.0% (178) | 26.4% (469) | 25.8% (458) |
| Wuhan Renmin COV+ | 39.8% (45) | 13.3% (15) | 22.1% (25) | 24.8% (28) |
| Wuhan controls | 32.2% (1188) | 9.1% (336) | 24.9% (920) | 33.8% (1250) |
| Shenzhen COV+ | 28.8% (82) | 13.7% (39) | 29.1% (83) | 28.4% (81) |
| Shenzhen controls | 28.8% (6728) | 7.3% (1712) | 25.1% (5880) | 38.8% (9066) |
| UK Biobank COV+ | 45.3% (466) | 3.9% (40) | 10.7% (110) | 40.1% (412) |
| UK Biobank controls | 43.4% (210213) | 3.6% (17561) | 9.6% (46576) | 43.3% (209777) |
| Pooled OR, 95% CI, p-value | 1.11, [0.99-1.26], p=0.083 | 1.23, [0.96-1.59], p=0.11 | 1.11, [1.03-1.19], p=0.0059 | 0.77, [0.65-0.92], p=0.0038 |

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

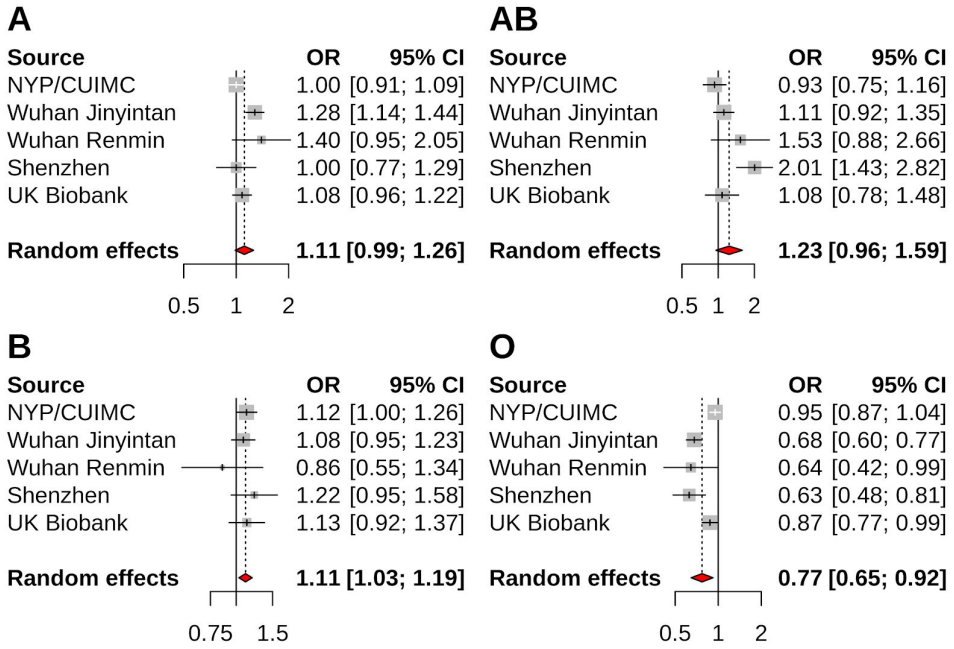
# Figures

*Figure 1:* ***Blood group effect size estimates.*** *Effect size estimates (odds ratios and conditional odds ratios) for univariate (outcome ~ blood group; blue) and multivariate (outcome ~ blood group + demographics + comorbidities; red) logistic regressions. These regressions were run for each blood group separately, considering all other groups as the reference group. 95% confidence intervals (CI) computed using Wald's normal approximation.*



medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

*Figure 2:* ***Meta-analysis effect size estimates.*** *The meta-analysis considered each blood group separately, and individual source effect sizes were pooled using a random effects model to estimate a combined effect size estimate. Boxes in the forest plot are sized according to their weight in the random effects model. For example, the NYP/CUIMC estimate for blood group A received higher weight than Wuhan Renmin. Supplementary Figure 1 shows more detailed forest plots, including explicit weights and tests of heterogeneity.*



medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

# Supplementary information

*Supplementary Table 1:* ***Cohort breakdown by ABO/Rh blood group***

| Characteristic | A- | A+ | AB-* | AB+ | B- | B+ | O- | O+ |
|---|---|---|---|---|---|---|---|---|
| N | 257 | 2280 | 27 | 307 | 104 | 1094 | 308 | 3393 |
| Median age (IQR) | 52 (35-70) | 59 (37-73) | 61 (44-75) | 58 (36-71) | 55 (36-73) | 57 (37-72) | 59 (38-71) | 56 (36-71) |
| Male sex (%) | 94 (36.6) | 889 (39) | 10 (37) | 131 (42.7) | 30 (28.8) | 443 (40.5) | 118 (38.3) | 1311 (38.6) |
| Race - Asian (%) | 3 (1.2) | 37 (1.6) | 2 (7.4) | 8 (2.6) | 1 (1) | 58 (5.3) | 3 (1) | 63 (1.9) |
| Race - Black/AA (%) | 34 (13.2) | 391 (17.1) | 1 (3.7) | 67 (21.8) | 19 (18.3) | 306 (28) | 38 (12.3) | 700 (20.6) |
| Race - White (%) | 126 (49) | 889 (39) | 16 (59.3) | 109 (35.5) | 46 (44.2) | 299 (27.3) | 137 (44.5) | 1029 (30.3) |
| Race - Other (%) | 52 (20.2) | 582 (25.5) | 3 (11.1) | 74 (24.1) | 17 (16.3) | 236 (21.6) | 75 (24.4) | 988 (29.1) |
| Race - Missing (%) | 42 (16.3) | 381 (16.7) | 5 (18.5) | 49 (16) | 21 (20.2) | 195 (17.8) | 55 (17.9) | 613 (18.1) |
| Ethnicity - Hispanic (%) | 91 (35.4) | 917 (40.2) | 7 (25.9) | 103 (33.6) | 32 (30.8) | 366 (33.5) | 115 (37.3) | 1575 (46.4) |
| Ethnicity - Non-Hispanic (%) | 120 (46.7) | 921 (40.4) | 11 (40.7) | 138 (45) | 46 (44.2) | 503 (46) | 118 (38.3) | 1159 (34.2) |
| Ethnicity - Other (%) | 3 (1.2) | 24 (1.1) | 0 (0) | 7 (2.3) | 3 (2.9) | 9 (0.8) | 4 (1.3) | 54 (1.6) |
| Ethnicity - Missing (%) | 43 (16.7) | 418 (18.3) | 9 (33.3) | 59 (19.2) | 23 (22.1) | 216 (19.7) | 71 (23.1) | 605 (17.8) |
| Hypertension (%) | 114 (44.4) | 1144 (50.2) | 11 (40.7) | 144 (46.9) | 44 (42.3) | 538 (49.2) | 165 (53.6) | 1623 (47.8) |
| Cardiovascular diseases (%) | 172 (66.9) | 1551 (68) | 20 (74.1) | 206 (67.1) | 58 (55.8) | 713 (65.2) | 206 (66.9) | 2249 (66.3) |
| Respiratory diseases (%) | 152 (59.1) | 1408 (61.8) | 13 (48.1) | 187 (60.9) | 58 (55.8) | 659 (60.2) | 199 (64.6) | 2070 (61) |
| Diabetes mellitus (%) | 69 (26.8) | 760 (33.3) | 5 (18.5) | 96 (31.3) | 27 (26) | 376 (34.4) | 104 (33.8) | 1085 (32) |
| Obesity (%) | 96 (37.4) | 977 (42.9) | 17 (63) | 116 (37.8) | 39 (37.5) | 429 (39.2) | 132 (42.9) | 1459 (43) |
| COV+ (%) | 57 (22.2) | 664 (29.1) | 1 (3.7) | 86 (28) | 25 (24) | 338 (30.9) | 74 (24) | 961 (28.3) |
| COV+/Intubated (%) | 5 (1.9) | 162 (7.1) | 0 (0) | 31 (10.1) | 7 (6.7) | 102 (9.3) | 17 (5.5) | 264 (7.8) |
| COV+/Died (%) | 5 (1.9) | 156 (6.8) | 0 (0) | 26 (8.5) | 6 (5.8) | 65 (5.9) | 10 (3.2) | 237 (7) |

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

*Supplementary Table 2:* ***Tests of overall associations.*** *Counts for groups 1 and 2 are the individual group counts for the former and latter groups in the comparison. For example, in the 'COV+ vs COV-' comparison, Group 1 counts gives counts and percentages for 'COV+' and Group 2 counts gives counts for 'COV-'. ABO used a 4x2 table for each test, Rh used a 2x2 table for each test, while ABO/Rh used a 6x2 table for each test, resulting in 3, 1, and 5 degrees of freedom, respectively.*

| Comparison | Blood group type | Group 1 counts | Group 2 counts | p-value |
|---|---|---|---|---|
| COV+ vs COV- | ABO | A: 721 (32.7%), AB: 87 (3.9%), B: 363 (16.5%), O: 1035 (46.9%) | A: 1816 (32.6%), AB: 247 (4.4%), B: 835 (15%), O: 2666 (47.9%) | 0.34 |
| COV+ vs general population | ABO | A: 721 (32.7%), AB: 87 (3.9%), B: 363 (16.5%), O: 1035 (46.9%) | A: 34831 (32.7%), AB: 4492 (4.2%), B: 15904 (14.9%), O: 51301 (48.2%) | 0.22 |
| COV+/Intubated vs COV+/Not intubated | ABO | A: 167 (28.4%), AB: 31 (5.3%), B: 109 (18.5%), O: 281 (47.8%) | A: 554 (34.2%), AB: 56 (3.5%), B: 254 (15.7%), O: 754 (46.6%) | 0.016 |
| COV+/Intubated vs COV+ | ABO | A: 167 (28.4%), AB: 31 (5.3%), B: 109 (18.5%), O: 281 (47.8%) | A: 721 (32.7%), AB: 87 (3.9%), B: 363 (16.5%), O: 1035 (46.9%) | 0.12 |
| COV+/Died vs COV+/Alive | ABO | A: 161 (31.9%), AB: 26 (5.1%), B: 71 (14.1%), O: 247 (48.9%) | A: 560 (32.9%), AB: 61 (3.6%), B: 292 (17.2%), O: 788 (46.3%) | 0.15 |
| COV+/Died vs COV+ | ABO | A: 161 (31.9%), AB: 26 (5.1%), B: 71 (14.1%), O: 247 (48.9%) | A: 721 (32.7%), AB: 87 (3.9%), B: 363 (16.5%), O: 1035 (46.9%) | 0.34 |
| SARS-CoV-2 tested vs general population | ABO | A: 2537 (32.7%), AB: 334 (4.3%), B: 1198 (15.4%), O: 3701 (47.6%) | A: 34831 (32.7%), AB: 4492 (4.2%), B: 15904 (14.9%), O: 51301 (48.2%) | 0.64 |
| COV+ vs COV- | ABO/Rh | A-: 57 (2.6%), A+: 664 (30.1%), AB+: 86 (3.9%), B-: 25 (1.1%), B+: 338 (15.3%), O-: 74 (3.4%), O+: 961 (43.6%) | A-: 200 (3.6%), A+: 1616 (29.2%), AB+: 221 (4%), B-: 79 (1.4%), B+: 756 (13.7%), O-: 234 (4.2%), O+: 2432 (43.9%) | 0.048 |

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

| | | | | |
|---|---|---|---|---|
| COV+ vs general population | ABO/Rh | A-: 57 (2.6%), A+: 664 (30.1%), AB+: 86 (3.9%), B-: 25 (1.1%), B+: 338 (15.3%), O-: 74 (3.4%), O+: 961 (43.6%) | A-: 3378 (3.2%), A+: 31453 (29.6%), AB+: 4103 (3.9%), B-: 1399 (1.3%), B+: 14505 (13.7%), O-: 4718 (4.4%), O+: 46583 (43.9%) | 0.038 |
| COV+/Intubated vs COV+/Not intubated | ABO/Rh | A-: 5 (0.9%), A+: 162 (27.6%), AB+: 31 (5.3%), B-: 7 (1.2%), B+: 102 (17.3%), O-: 17 (2.9%), O+: 264 (44.9%) | A-: 52 (3.2%), A+: 502 (31%), AB+: 55 (3.4%), B-: 18 (1.1%), B+: 236 (14.6%), O-: 57 (3.5%), O+: 697 (43.1%) | 0.0064 |
| COV+/Intubated vs COV+ | ABO/Rh | A-: 5 (0.9%), A+: 162 (27.6%), AB+: 31 (5.3%), B-: 7 (1.2%), B+: 102 (17.3%), O-: 17 (2.9%), O+: 264 (44.9%) | A-: 57 (2.6%), A+: 664 (30.1%), AB+: 86 (3.9%), B-: 25 (1.1%), B+: 338 (15.3%), O-: 74 (3.4%), O+: 961 (43.6%) | 0.085 |
| COV+/Died vs COV+/Alive | ABO/Rh | A-: 5 (1%), A+: 156 (30.9%), AB+: 26 (5.1%), B-: 6 (1.2%), B+: 65 (12.9%), O-: 10 (2%), O+: 237 (46.9%) | A-: 52 (3.1%), A+: 508 (29.9%), AB+: 60 (3.5%), B-: 19 (1.1%), B+: 273 (16.1%), O-: 64 (3.8%), O+: 724 (42.6%) | 0.0087 |
| COV+/Died vs COV+ | ABO/Rh | A-: 5 (1%), A+: 156 (30.9%), AB+: 26 (5.1%), B-: 6 (1.2%), B+: 65 (12.9%), O-: 10 (2%), O+: 237 (46.9%) | A-: 57 (2.6%), A+: 664 (30.1%), AB+: 86 (3.9%), B-: 25 (1.1%), B+: 338 (15.3%), O-: 74 (3.4%), O+: 961 (43.6%) | 0.076 |
| SARS-CoV-2 tested vs general population | ABO/Rh | A-: 257 (3.3%), A+: 2280 (29.4%), AB+: 307 (4%), B-: 104 (1.3%), B+: 1094 (14.1%), O-: 308 (4%), O+: 3393 (43.8%) | A-: 3378 (3.2%), A+: 31453 (29.6%), AB+: 4103 (3.9%), B-: 1399 (1.3%), B+: 14505 (13.7%), O-: 4718 (4.4%), O+: 46583 (43.9%) | 0.49 |
| COV+ vs COV- | Rh | Rh-neg: 157 (7.1%), Rh-pos: 2049 (92.9%) | Rh-neg: 539 (9.7%), Rh-pos: 5025 (90.3%) | 0.00041 |
| COV+ vs general population | Rh | Rh-neg: 157 (7.1%), Rh-pos: 2049 (92.9%) | Rh-neg: 9884 (9.3%), Rh-pos: 96644 (90.7%) | 0.00060 |
| COV+/Intubated vs COV+/Not intubated | Rh | Rh-neg: 29 (4.9%), Rh-pos: 559 (95.1%) | Rh-neg: 128 (7.9%), Rh-pos: 1490 (92.1%) | 0.021 |
| COV+/Intubated vs COV+ | Rh | Rh-neg: 29 (4.9%), Rh-pos: 559 (95.1%) | Rh-neg: 157 (7.1%), Rh-pos: 2049 (92.9%) | 0.073 |
| COV+/Died vs COV+/Alive | Rh | Rh-neg: 21 (4.2%), Rh-pos: 484 (95.8%) | Rh-neg: 136 (8%), Rh-pos: 1565 (92%) | 0.0044 |

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

| | | | | |
|---|---|---|---|---|
| COV+/Died vs COV+ | Rh | Rh-neg: 21 (4.2%), Rh-pos: 484 (95.8%) | Rh-neg: 157 (7.1%), Rh-pos: 2049 (92.9%) | 0.020 |
| SARS-CoV-2 tested vs general population | Rh | Rh-neg: 696 (9%), Rh-pos: 7074 (91%) | Rh-neg: 9884 (9.3%), Rh-pos: 96644 (90.7%) | 0.36 |
| Rh-neg vs Rh-pos | ABO | A: 257 (36.9%), AB: 27 (3.9%), B: 104 (14.9%), O: 308 (44.3%) | A: 2280 (32.2%), AB: 307 (4.3%), B: 1094 (15.5%), O: 3393 (48%) | 0.088 |

*Supplementary Table 3:* ***Individual blood group effect size estimates.*** *Each estimate made using logistic regression of outcome ~ blood group, with or without risk factor covariates. Raw estimate refers to a univariate estimate. Adjusted estimate refers to an estimate using regression including risk factor (demographics and comorbidities) covariates. Cell values represent odds ratio, 95% confidence interval, and p-value.*

| Outcome | Blood group | Raw estimate | Adjusted estimate |
|---|---|---|---|
| COV+ | A | 1.002, [0.902-1.113], p=0.97 | 1.032, [0.923-1.153], p=0.58 |
| COV+ | AB | 0.884, [0.689-1.134], p=0.33 | 0.936, [0.720-1.215], p=0.62 |
| COV+ | B | 1.115, [0.975-1.276], p=0.11 | 1.170, [1.014-1.350], p=0.032 |
| COV+ | O | 0.961, [0.870-1.061], p=0.43 | 0.906, [0.816-1.006], p=0.064 |
| COV+ | A- | 0.711, [0.528-0.959], p=0.025 | 0.838, [0.612-1.147], p=0.27 |
| COV+ | A+ | 1.052, [0.944-1.172], p=0.36 | 1.059, [0.945-1.187], p=0.33 |
| COV+ | AB+ | 0.981, [0.761-1.265], p=0.88 | 1.030, [0.789-1.346], p=0.83 |
| COV+ | B- | 0.796, [0.506-1.251], p=0.32 | 0.918, [0.570-1.476], p=0.72 |
| COV+ | B+ | 1.151, [1.001-1.322], p=0.048 | 1.193, [1.029-1.384], p=0.019 |
| COV+ | O- | 0.791, [0.606-1.032], p=0.084 | 0.820, [0.621-1.083], p=0.16 |
| COV+ | O+ | 0.994, [0.900-1.098], p=0.91 | 0.932, [0.838-1.035], p=0.19 |
| COV+ | Rh-neg | 0.714, [0.594-0.860], p=0.00037 | 0.796, [0.655-0.966], p=0.021 |
| COV+ | Rh-pos | 1.400, [1.163-1.684], p=0.00037 | 1.256, [1.035-1.526], p=0.021 |
| COV+/Intubated | A | 0.762, [0.620-0.937], p=0.0099 | 0.767, [0.618-0.952], p=0.016 |
| COV+/Intubated | AB | 1.552, [0.991-2.433], p=0.055 | 1.788, [1.120-2.854], p=0.015 |
| COV+/Intubated | B | 1.222, [0.954-1.565], p=0.11 | 1.200, [0.925-1.557], p=0.17 |
| COV+/Intubated | O | 1.049, [0.868-1.267], p=0.62 | 1.030, [0.846-1.255], p=0.77 |
| COV+/Intubated | A- | 0.258, [0.103-0.650], p=0.004 | 0.278, [0.108-0.714], p=0.0078 |
| COV+/Intubated | A+ | 0.845, [0.686-1.042], p=0.12 | 0.844, [0.679-1.050], p=0.13 |
| COV+/Intubated | AB+ | 1.582, [1.008-2.482], p=0.046 | 1.826, [1.142-2.920], p=0.012 |
| COV+/Intubated | B- | 1.071, [0.445-2.577], p=0.88 | 1.227, [0.495-3.041], p=0.66 |

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

| | | | |
|---|---|---|---|
| COV+/Intubated | B+ | 1.229, [0.953-1.585], p=0.11 | 1.191, [0.911-1.556], p=0.2 |
| COV+/Intubated | O- | 0.815, [0.470-1.413], p=0.47 | 0.945, [0.535-1.671], p=0.85 |
| COV+/Intubated | O+ | 1.077, [0.890-1.302], p=0.45 | 1.038, [0.851-1.266], p=0.71 |
| COV+/Intubated | Rh-neg | 0.604, [0.399-0.914], p=0.017 | 0.685, [0.445-1.053], p=0.084 |
| COV+/Intubated | Rh-pos | 1.656, [1.094-2.507], p=0.017 | 1.460, [0.950-2.245], p=0.084 |
| COV+/Died | A | 0.954, [0.771-1.180], p=0.66 | 0.905, [0.715-1.145], p=0.4 |
| COV+/Died | AB | 1.459, [0.912-2.335], p=0.12 | 1.466, [0.870-2.472], p=0.15 |
| COV+/Died | B | 0.789, [0.596-1.045], p=0.099 | 0.750, [0.550-1.023], p=0.069 |
| COV+/Died | O | 1.109, [0.909-1.353], p=0.31 | 1.191, [0.955-1.486], p=0.12 |
| COV+/Died | A- | 0.317, [0.126-0.798], p=0.015 | 0.396, [0.149-1.057], p=0.064 |
| COV+/Died | A+ | 1.050, [0.846-1.302], p=0.66 | 0.969, [0.763-1.229], p=0.79 |
| COV+/Died | AB+ | 1.485, [0.927-2.378], p=0.1 | 1.495, [0.885-2.526], p=0.13 |
| COV+/Died | B- | 1.064, [0.423-2.680], p=0.89 | 1.170, [0.410-3.336], p=0.77 |
| COV+/Died | B+ | 0.773, [0.578-1.034], p=0.082 | 0.727, [0.528-1.002], p=0.051 |
| COV+/Died | O- | 0.517, [0.263-1.014], p=0.055 | 0.555, [0.268-1.150], p=0.11 |
| COV+/Died | O+ | 1.193, [0.978-1.457], p=0.082 | 1.272, [1.019-1.588], p=0.034 |
| COV+/Died | Rh-neg | 0.499, [0.312-0.799], p=0.0038 | 0.565, [0.338-0.944], p=0.029 |
| COV+/Died | Rh-pos | 2.003, [1.251-3.207], p=0.0038 | 1.770, [1.059-2.959], p=0.029 |

*Supplementary Table 4:* ***Associations between blood groups and risk factors.*** *Shown are associations reaching statistical significance at the 5% level. Full associations are available online (https://git.io/JJfRC)*

| Blood group | Term | OR | 95% CI | p-value |
|---|---|---|---|---|
| A | ethnicity (Hispanic) | 0.838 | 0.740-0.950 | 0.0057 |
| A | race (Asian) | 0.467 | 0.325-0.672 | 4.1e-05 |
| A | race (Black) | 0.592 | 0.516-0.680 | 1.2e-13 |
| A | race (Other) | 0.800 | 0.698-0.915 | 0.0012 |
| A | race (Missing) | 0.769 | 0.655-0.902 | 0.0013 |
| AB | ethnicity (Hispanic) | 0.725 | 0.540-0.974 | 0.033 |
| B | ethnicity (Hispanic) | 0.727 | 0.616-0.858 | 0.00016 |
| B | race (Asian) | 3.098 | 2.212-4.338 | 4.7e-11 |
| B | race (Black) | 1.719 | 1.451-2.036 | 3.5e-10 |
| B | race (Missing) | 1.279 | 1.035-1.581 | 0.023 |
| B | cardiovascular disorders | 0.782 | 0.643-0.952 | 0.014 |

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

| | | | | |
|---|---|---|---|---|
| B | obesity | 0.872 | 0.765-0.995 | 0.041 |
| O | ethnicity (Hispanic) | 1.446 | 1.285-1.626 | 8.1e-10 |
| O | ethnicity (Other) | 1.660 | 1.114-2.473 | 0.013 |
| O | race (Black) | 1.193 | 1.050-1.356 | 0.0068 |
| O | race (Other) | 1.181 | 1.039-1.342 | 0.011 |
| O | race (Missing) | 1.171 | 1.007-1.361 | 0.04 |
| Rh-neg | ethnicity (Hispanic) | 0.758 | 0.617-0.932 | 0.0085 |
| Rh-neg | race (Asian) | 0.356 | 0.180-0.705 | 0.0031 |
| Rh-neg | race (Black) | 0.434 | 0.340-0.554 | 2e-11 |
| Rh-neg | race (Other) | 0.633 | 0.505-0.793 | 6.8e-05 |
| Rh-neg | race (Missing) | 0.671 | 0.520-0.866 | 0.0021 |

*Supplementary Table 5:* ***Regression model fits using blood groups and risk factors.*** *The deviance column gives the deviance reduced by the addition of the first term in the comparison. Similarly, df indicates the degrees of freedom reduced by the addition. For both, the "Resid." column indicates the remaining deviance and degrees of freedom for the full model. P-values are computed using a chi-squared distribution with df degrees of freedom.*

| Outcome | Comparison | df | Resid. df | Deviance | Resid. deviance | p-value |
|---|---|---|---|---|---|---|
| COV+ | Risk factors vs Null | 14 | 7755 | 774.7 | 8496.6 | 2.9e-156 |
| COV+ | ABO + Risk factors vs Risk factors | 3 | 7752 | 6.1 | 8490.6 | 0.11 |
| COV+ | ABO/Rh + Risk factors vs Risk factors | 6 | 7749 | 10.1 | 8486.6 | 0.12 |
| COV+ | Rh(D) + Risk factors vs Risk factors | 1 | 7754 | 5.5 | 8491.2 | 0.019 |
| COV+/Intubated | Risk factors vs Null | 14 | 2191 | 148.6 | 2409.5 | 1.4e-24 |
| COV+/Intubated | ABO + Risk factors vs Risk factors | 3 | 2188 | 11.1 | 2398.4 | 0.011 |
| COV+/Intubated | ABO/Rh + Risk factors vs Risk factors | 6 | 2185 | 18.5 | 2390.9 | 0.005 |
| COV+/Intubated | Rh(D) + Risk factors vs Risk factors | 1 | 2190 | 3.2 | 2406.3 | 0.076 |
| COV+/Died | Risk factors vs Null | 14 | 2191 | 401.6 | 1971.9 | 5.8e-77 |
| COV+/Died | ABO + Risk factors vs Risk factors | 3 | 2188 | 6.5 | 1965.4 | 0.089 |
| COV+/Died | ABO/Rh + Risk factors vs Risk factors | 6 | 2185 | 14.8 | 1957.1 | 0.022 |
| COV+/Died | Rh(D) + Risk factors vs Risk factors | 1 | 2190 | 5.2 | 1966.8 | 0.023 |

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

*Supplementary Table 6:* ***Chi-squared tests between meta-analysis site references.*** *Each comparison is between NYP/CUIMC and the reference population for the site listed in 'Site'.*

| Site | Chi-squared | Degrees of freedom | p-value |
|---|---|---|---|
| UK Biobank | 5289.85 | 3 | 6.1e-1148 |
| Wuhan | 580.64 | 3 | 7.9e-126 |
| Shenzhen | 2036.29 | 3 | 1.2e-441 |

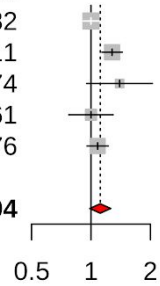
medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

*Supplementary Figure 1:* ***Expanded forest plots for meta-analysis.*** *Each meta-analysis considered a single ABO blood type versus all other types. Meta-analyses compared blood group distribution between COV+ individuals and general population (i.e. not tested for SARS-CoV-2 infection) estimates from each source or city. Effect size estimates are identical to those in Figure 2.*

A

| Source | ABO: A COV+ | ABO: A Total | ABO: ¬A COV+ | ABO: ¬A Total | OR | 95%-CI | Weight |
|---|---|---|---|---|---|---|---|
| NYP/CUIMC | 721 | 35552 | 1485 | 73182 | 1.00 | [0.91; 1.09] | 28.3% |
| Wuhan Jinyintan | 670 | 1858 | 1105 | 3611 | 1.28 | [1.14; 1.44] | 25.5% |
| Wuhan Renmin | 45 | 1233 | 68 | 2574 | 1.40 | [0.95; 2.05] | 7.8% |
| Shenzhen | 82 | 6810 | 203 | 16861 | 1.00 | [0.77; 1.29] | 13.4% |
| UK Biobank | 466 | 210679 | 562 | 274476 | 1.08 | [0.96; 1.22] | 25.0% |
| **Random effects** | | **256132** | | **370704** | **1.11** | **[0.99; 1.26]** | **100.0%** |

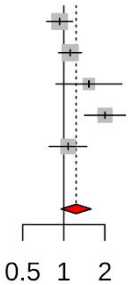
Heterogeneity: $I^2$ = 68%, $\tau^2$ = 0.0116, $p$ = 0.01
Test for overall effect: $p$ = 0.08




AB

| Source | ABO: AB COV+ | ABO: AB Total | ABO: ¬AB COV+ | ABO: ¬AB Total | OR | 95%-CI | Weight |
|---|---|---|---|---|---|---|---|
| NYP/CUIMC | 87 | 4579 | 2119 | 104155 | 0.93 | [0.75; 1.16] | 24.0% |
| Wuhan Jinyintan | 178 | 514 | 1597 | 4955 | 1.11 | [0.92; 1.35] | 24.9% |
| Wuhan Renmin | 15 | 351 | 98 | 3456 | 1.53 | [0.88; 2.66] | 12.1% |
| Shenzhen | 39 | 1751 | 246 | 21920 | 2.01 | [1.43; 2.82] | 19.0% |
| UK Biobank | 40 | 17601 | 988 | 467554 | 1.08 | [0.78; 1.48] | 20.0% |
| **Random effects** | | **24796** | | **602040** | **1.23** | **[0.96; 1.59]** | **100.0%** |

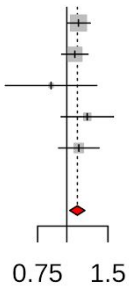
Heterogeneity: $I^2$ = 74%, $\tau^2$ = 0.0574, $p$ = 0.004
Test for overall effect: $p$ = 0.11




B

| Source | ABO: B COV+ | ABO: B Total | ABO: ¬B COV+ | ABO: ¬B Total | OR | 95%-CI | Weight |
|---|---|---|---|---|---|---|---|
| NYP/CUIMC | 363 | 16267 | 1843 | 92467 | 1.12 | [1.00; 1.26] | 42.3% |
| Wuhan Jinyintan | 469 | 1389 | 1306 | 4080 | 1.08 | [0.95; 1.23] | 32.8% |
| Wuhan Renmin | 25 | 945 | 88 | 2862 | 0.86 | [0.55; 1.34] | 2.7% |
| Shenzhen | 83 | 5963 | 202 | 17708 | 1.22 | [0.95; 1.58] | 8.3% |
| UK Biobank | 110 | 46686 | 918 | 438469 | 1.13 | [0.92; 1.37] | 14.0% |
| **Random effects** | | **71250** | | **555586** | **1.11** | **[1.03; 1.19]** | **100.0%** |

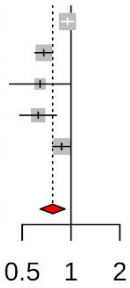
Heterogeneity: $I^2$ = 0%, $\tau^2$ = 0, $p$ = 0.73
Test for overall effect: $p$ = 0.006




O

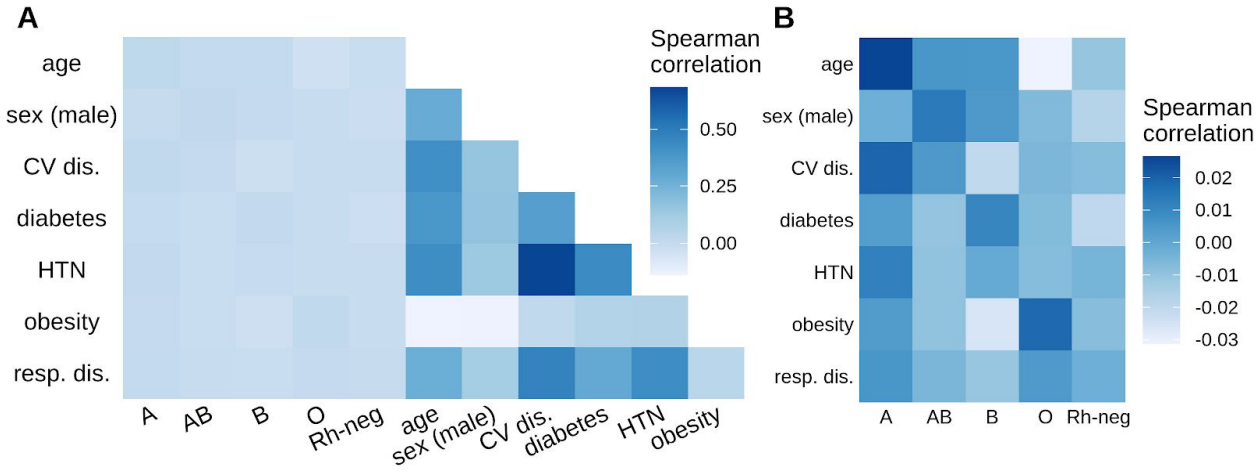
| Source | ABO: O COV+ | ABO: O Total | ABO: ¬O COV+ | ABO: ¬O Total | OR | 95%-CI | Weight |
|---|---|---|---|---|---|---|---|
| NYP/CUIMC | 1035 | 52336 | 1171 | 56398 | 0.95 | [0.87; 1.04] | 25.4% |
| Wuhan Jinyintan | 458 | 1708 | 1317 | 3761 | 0.68 | [0.60; 0.77] | 23.6% |
| Wuhan Renmin | 28 | 1278 | 85 | 2529 | 0.64 | [0.42; 0.99] | 10.3% |
| Shenzhen | 81 | 9147 | 204 | 14524 | 0.63 | [0.48; 0.81] | 17.0% |
| UK Biobank | 412 | 210189 | 616 | 274966 | 0.87 | [0.77; 0.99] | 23.7% |
| **Random effects** | | **274658** | | **352178** | **0.77** | **[0.65; 0.92]** | **100.0%** |

Heterogeneity: $I^2$ = 85%, $\tau^2$ = 0.0300, $p$ < 0.001
Test for overall effect: $p$ = 0.004



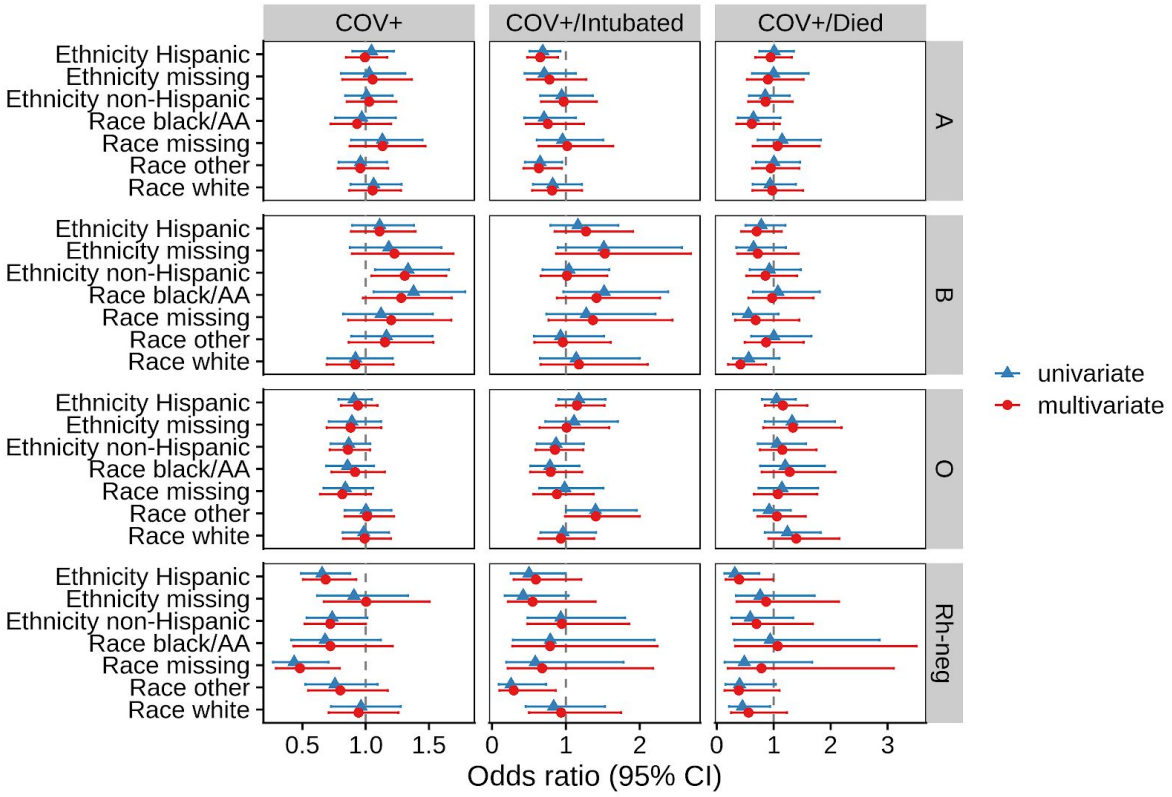
medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

*Supplementary Figure 2:* ***Correlation matrices between blood groups, demographics, and comorbidities.*** *A. shows the correlations among blood groups, risk factors, and risk factors. B. shows just the correlations between blood groups and risk factors, rescaled so the relatively smaller correlations become more discernible.*



medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

*Supplementary Figure 3:* ***Race/ethnicity stratified associations.*** *Each point represents an odds ratio and 95% confidence interval estimated using logistic regression. We excluded race and ethnicity from the multivariate regression form and computed estimates on one race/ethnicity at a time. Asian race and AB blood groups were excluded due to low sample sizes resulting in a lack of model convergence.*




## Electronic health record definitions

Blood group was determined using laboratory measurements coded using descendant concepts of LOINC LP36683-8 (ABO and Rh group). Individuals with contradictory measurements were excluded. Intubation was assessed using completed procedures having the procedure description, "Intubation."

Age was computed using a patient's birth date and either the patient's death date or the current date, for currently living patients. Sex was ascertained using a patient's self report, and we considered only individuals who specified either male or female. Race and ethnicity were reported by patients themselves. We grouped race into five categories and ethnicity into four. Specifically, we considered only Asian, Black/African-American, and White, categorizing other listed races (all of which were small minorities) as 'Other', and missing or declined race as 'Missing'. Ethnicity was categorized into Hispanic, non-Hispanic (explicitly specified), Other (specified but not Hispanic or non-Hispanic), and Missing.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

We considered five clinical risk factors—cardiovascular diseases, diabetes mellitus, hypertension, obesity, and respiratory diseases.

- Cardiovascular diseases refer to one or more diagnoses of any descendant concept of SNOMED 49601007 (Disorder of cardiovascular system).
- Diabetes mellitus refers to an individual having either: 1. one or more diagnoses of descendants of SNOMED 73211009 (Diabetes mellitus), 2. one or more diagnosis codes in the range ICD 10 CM E08-E13, or 3. HbA1c ≥ 6.5%, using any of the following laboratory measurements: LOINC 4548-4 (Hemoglobin A1c/Hemoglobin.total in Blood), LOINC 17856-6 (Hemoglobin A1c/Hemoglobin.total in Blood by HPLC), LOINC 55454-3 (Deprecated Hemoglobin A1c in Blood).
- We defined hypertension as a previous diagnosis using any descendant concept of SNOMED 38341003 (Hypertensive disorder), including term mappings to ICD 10 CM.
- Obesity was defined using either a measured BMI greater ≥ 30 or a BMI percentile ≥ 95 since January 1, 2019 or one or more occurrences of a descendant concept of SNOMED 414915002 (Obese).
- Respiratory diseases refer to one or more diagnoses of any descendant concept of SNOMED 50043002 (Disorder of respiratory system), including term mappings to ICD 10 CM.

## UK Biobank

*Supplementary Table 6:* ***Variant map between Melzer et al. and the UK Biobank****.*

| Variant rs ID | Melzer et al. [16] | UK Biobank |
|---|---|---|
| rs8176746 | C > A | G > T |
| rs8176719 | del G | T > TC |

*Supplementary Table 7:* ***Genotype to phenotype map used for ABO blood type in the UK Biobank.*** *The third column gives allele dosages for the listed variants, which corresponds to the data available.*

| Melzer et al. [16] | UK Biobank | rs8176746_G / rs8176719_T | ABO blood type |
|---|---|---|---|
| C/C del/del | G/G T/T | 2 / 2 | O |
| C/C del/G | G/G T/TC | 2 / 1 | A |
| C/C G/G | G/G TC/TC | 2 / 0 | A |
| C/A del/del | G/T T/T | 1 / 2 | O |
| A/A G/G | T/T TC/TC | 0 / 0 | B |

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058073; this version posted July 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

| | | | |
|---|---|---|---|
| C/A del/G | G/T T/TC | 1 / 1 | B |
| C/A G/G | G/T TC/TC | 1 / 0 | AB |

*Supplementary Table 8:* ***Genetically-determined ABO distribution in the UK Biobank.*** *'Other' indicates that an individual's blood type could not be determined using the mapping we adapted from Melzer et al.* [16].

| Blood group | Number of individuals |
|---|---|
| A | 211478 |
| AB | 17642 |
| B | 46860 |
| O | 210925 |
| Other | 137 |

## SARS-CoV-2 infection

We used the 'covid19_result' table to determine COV+ and COV-tested status in the UK Biobank. Following the same procedure as we used for NYP/CUIMC data, we considered an individual COV+ if any recorded SARS-CoV-2 test was positive and COV— only if the individual received at least one test, but every recorded test was negative for infection. No distinction was made by type of specimen, origin, or laboratory.