



# Associations between blood type and COVID-19 infection, intubation, and death

Michael Zietz <sup>1</sup>, Jason Zucker<sup>2</sup> & Nicholas P. Tatonetti <sup>1,2,3,4</sup>✉

The rapid global spread of the novel coronavirus SARS-CoV-2 has strained healthcare and testing resources, making the identification and prioritization of individuals most at-risk a critical challenge. Recent evidence suggests blood type may affect risk of severe COVID-19. Here, we use observational healthcare data on 14,112 individuals tested for SARS-CoV-2 with known blood type in the New York Presbyterian (NYP) hospital system to assess the association between ABO and Rh blood types and infection, intubation, and death. We find slightly increased infection prevalence among non-O types. Risk of intubation was decreased among A and increased among AB and B types, compared with type O, while risk of death was increased for type AB and decreased for types A and B. We estimate Rh-negative blood type to have a protective effect for all three outcomes. Our results add to the growing body of evidence suggesting blood type may play a role in COVID-19.

<sup>1</sup>Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, USA. <sup>2</sup>Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA. <sup>3</sup>Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA. <sup>4</sup>Institute for Genomic Medicine, Columbia University Irving Medical Center, New York, NY, USA. ✉email: [nick.tatonetti@columbia.edu](mailto:nick.tatonetti@columbia.edu)

The novel Coronavirus disease (COVID-19, caused by the SARS-CoV-2 virus) has spread rapidly across the globe and has caused over 21.1 million confirmed infections and over 761,000 deaths worldwide as of August 17, 2020<sup>1</sup>. Within the United States, New York suffered among the worst outbreaks during the early phases of the pandemic. As of August 22, New York City has recorded 228,144 confirmed infections and 19,014 deaths<sup>2</sup>. A number of risk factors for COVID-19 morbidity and mortality are known, including age, sex, smoking, hypertension, diabetes, and chronic cardiovascular and respiratory diseases<sup>3,4</sup>.

Recent work has demonstrated an association between ABO blood types and COVID-19 risk. Using data from Wuhan and Shenzhen, Zhao et al. found a greater proportion of A and a lower proportion of O blood types among COVID-19 patients, relative to the general populations of Wuhan and Shenzhen<sup>5</sup>. Similarly, using a meta-analysis of data from Italy and Spain, Ellinghaus et al. found<sup>6</sup> a higher risk of COVID-19 among A and a lower risk among O blood types. Conversely, however, they estimated lower odds of mechanical ventilation for all non-O types, though the estimated odds ratios were not statistically significant at the 5% level for this outcome.

The ABO blood type trait reflects polymorphisms within the ABO gene. This gene is associated with a number of other traits, including risk factors for COVID-19 morbidity and mortality. For example, genome-wide association studies have associated variants within ABO to activity of the angiotensin converting enzyme<sup>7</sup>, red blood cell count, hemoglobin concentration, hematocrit<sup>8–11</sup>, von Willebrand factor<sup>12–15</sup>, myocardial infarction<sup>16,17</sup>, coronary artery disease<sup>17–21</sup>, ischemic stroke<sup>13,19,22</sup>, type 2 diabetes<sup>23–25</sup>, and venous thromboembolism<sup>26–33</sup>. A 2012 meta-analysis found that, in addition to individual variants, a non-O blood type is among the most important genetic risk factors for venous thromboembolism<sup>34</sup>. These conditions are also relevant for COVID-19. For example, coagulopathy is a common issue for COVID-19 patients<sup>35–41</sup>, and risk of venous thromboembolism must be carefully managed<sup>42</sup>.

The numerous associations between conditions and both blood type and COVID-19 provide reason to believe that true associations may exist between blood type and morbidity and mortality due to COVID-19. In addition, previous work has identified associations between ABO blood groups and a number of different infections or disease severity following infections, including SARS-CoV-1<sup>43</sup>, *P. falciparum*<sup>44</sup>, *H. pylori*<sup>45</sup>, Norwalk virus<sup>46</sup>, hepatitis B virus<sup>47</sup>, and *N. gonorrhoeae*<sup>48</sup>.

Rh(D) phenotypes (positive and negative Rh blood types) are associated with very few diseases compared to ABO<sup>49</sup>. Like ABO, Rh type is important for type compatibility and immune response. For example, hemolytic disease of the newborn is a concern when Rh(D) is mismatched between mother and offspring<sup>50</sup>. Other studies have found evidence that Rh-positive individuals are protected against the effects of latent toxoplasmosis<sup>51</sup>, though *Toxoplasma gondii* is a eukaryotic parasite<sup>52</sup>, not a virus like SARS-CoV-2.

In this study, we sought to understand the association between SARS-CoV-2 infection/COVID-19 and blood type using electronic health record (EHR) data from NewYork-Presbyterian/Columbia University Irving Medical Center (NYP/CUIMC) hospital in New York City, USA. We compared both ABO and Rh (D) blood types, and we investigated initial infection status and two severe COVID-19 outcomes: intubation and death. We evaluated potential confounding due to population stratification using a multivariate analysis, and we report clinically meaningful measures of effect.

## Results

**Data collection and cohort selection.** We determined blood types using laboratory measurements recorded in the NYP/CUIMC EHR system. After removing likely errors, such as individuals with contradictory blood-type results, we identified 14,112 adult individuals with known blood types who received at least one SARS-CoV-2 swab test (Table 1 and Supplementary Fig. 1). We performed chi-squared tests of independence and found insufficient evidence to conclude that the blood-group frequencies differ between SARS-CoV-2-tested and non-tested groups (Supplementary Table 1). Individuals were considered initially SARS-CoV-2-positive (COV+) if they tested positive on their first recorded test or within the following 96 h. We evaluated associations between blood types and outcomes using three comparisons: infection prevalence among initial tests and survival analysis for intubation and death among individuals with infections confirmed by the swab test. We report on clinical data as of August 1, 2020.

**Infection prevalence.** The unadjusted prevalence of initial infection was higher among A and B blood types and lower among AB types, compared with type O (Table 2 and Fig. 1). To avoid bias with respect to healthcare utilization, length of hospital

**Table 1 Summary demographics for SARS-CoV-2-tested individuals at NYP/CUIMC, stratified by blood type.**

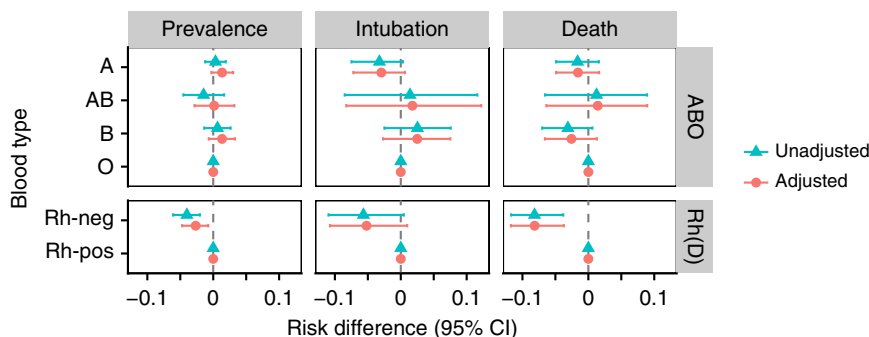
	A	AB	B	O	Rh neg	Rh pos
N	4298 (32.9)	559 (4.3)	2033 (15.6)	6161 (47.2)	1195 (9.2)	11,856 (90.8)
Age (IQR)	58 (37–72)	57 (37–71)	57 (37–72)	55 (36–71)	56 (37–70)	56 (37–71)
Male (%)	1676 (39.0)	231 (41.3)	778 (38.3)	2339 (38.0)	430 (36.0)	4594 (38.7)
Hispanic (%)	1572 (36.6)	173 (30.9)	666 (32.8)	2583 (41.9)	389 (32.6)	4605 (38.8)
Race						
Asian (%)	71 (1.7)	21 (3.8)	89 (4.4)	123 (2.0)	16 (1.3)	288 (2.4)
Black (%)	629 (14.6)	95 (17.0)	493 (24.2)	1179 (19.1)	151 (12.6)	2245 (18.9)
Missing (%)	728 (16.9)	79 (14.1)	370 (18.2)	1093 (17.7)	192 (16.1)	2078 (17.5)
Other (%)	1085 (25.2)	132 (23.6)	464 (22.8)	1715 (27.8)	263 (22.0)	3133 (26.4)
White (%)	1785 (41.5)	232 (41.5)	617 (30.3)	2051 (33.3)	573 (47.9)	4112 (34.7)
Outcomes						
Initially COV+ (%)	754 (17.5)	88 (15.7)	363 (17.9)	1060 (17.2)	164 (13.7)	2101 (17.7)
COV+ (%)	786 (18.3)	94 (16.8)	392 (19.3)	1122 (18.2)	175 (14.6)	2219 (18.7)
COV+ Intubated (%)	111 (2.6)	17 (3.0)	78 (3.8)	193 (3.1)	24 (2.0)	375 (3.2)
COV+ Died (%)	104 (2.4)	15 (2.7)	46 (2.3)	166 (2.7)	11 (0.9)	320 (2.7)

N is the number of individuals having the given blood type who had at least one recorded test for SARS-CoV-2. Age is reported as the median and interquartile range (25–75). Percents are reported relative to individuals having the blood type, except the N row, where percents are by blood-group type (ABO or Rh) and are relative to all individuals in the study.

**Table 2** Effect size estimates for blood types with and without correction for race and ethnicity.

Outcome	Blood type	Unadjusted			Race and ethnicity adjusted		
		Risk	Risk difference	Risk ratio	Risk	Risk difference	Risk ratio
Prevalence	A	17.5 (16.3-18.8)	0.3 (-1.2-1.9)	1.02 (0.93-1.12)	18.0 (16.8-19.2)	1.3 (-0.3-3.0)	1.08 (0.98-1.19)
	AB	15.7 (12.8-18.7)	-1.5 (-4.5-1.7)	0.91 (0.74-1.10)	16.8 (13.9-19.8)	0.1 (-2.8-3.2)	1.01 (0.83-1.20)
	B	17.9 (16.1-19.5)	0.7 (-1.4-2.6)	1.04 (0.92-1.16)	18.0 (16.3-19.7)	1.3 (-0.7-3.3)	1.08 (0.96-1.20)
	O	17.2 (16.2-18.1)	Reference	Reference	16.7 (15.7-17.6)	Reference	Reference
	Rh pos	17.7 (17.0-18.4)	Reference	Reference	17.6 (16.9-18.3)	Reference	Reference
Intubation	Rh neg	13.7 (11.8-15.5)	-4.0 (-6.1 to -2.0)	0.77 (0.66-0.88)	14.9 (13.0-16.8)	-2.7 (-4.7 to -0.8)	0.85 (0.73-0.96)
	A	17.2 (14.2-20.1)	-3.2 (-7.5-0.3)	0.84 (0.66-1.02)	17.3 (14.3-20.4)	-2.9 (-7.2-0.6)	0.85 (0.68-1.03)
	AB	21.8 (12.3-31.7)	1.4 (-8.5-11.6)	1.07 (0.59-1.57)	22.1 (12.8-32.1)	1.8 (-8.3-12.2)	1.09 (0.60-1.59)
	B	22.9 (18.6-27.6)	2.5 (-2.5-7.6)	1.12 (0.89-1.40)	22.8 (18.6-27.5)	2.5 (-2.7-7.5)	1.12 (0.88-1.40)
	O	20.4 (17.8-23.4)	Reference	Reference	20.3 (17.7-23.3)	Reference	Reference
Death	Rh pos	20.3 (18.4-22.1)	Reference	Reference	20.2 (18.4-22.1)	Reference	Reference
	Rh neg	14.6 (9.7-20.7)	-5.7 (-11.0-0.5)	0.72 (0.47-1.02)	15.0 (9.9-21.0)	-5.2 (-10.7-1.0)	0.74 (0.48-1.05)
	A	13.3 (11.0-15.7)	-1.6 (-4.9-1.6)	0.89 (0.71-1.11)	13.2 (10.9-15.6)	-1.6 (-4.9-1.6)	0.89 (0.71-1.12)
	AB	16.1 (8.5-23.8)	1.2 (-6.6-8.9)	1.08 (0.58-1.62)	16.2 (8.7-23.5)	1.4 (-6.4-8.9)	1.10 (0.59-1.64)
	B	11.8 (8.6-15.0)	-3.1 (-7.0-0.6)	0.79 (0.56-1.05)	12.2 (9.0-15.5)	-2.6 (-6.6-1.3)	0.83 (0.58-1.09)
Death	O	14.9 (12.9-17.1)	Reference	Reference	14.8 (12.7-16.9)	Reference	Reference
	Rh pos	14.5 (13.0-16.0)	Reference	Reference	14.5 (13.0-16.0)	Reference	Reference
	Rh neg	6.3 (3.0-10.1)	-8.2 (-11.7 to -3.8)	0.44 (0.21-0.72)	6.4 (3.0-10.3)	-8.2 (-11.7 to -3.7)	0.44 (0.21-0.74)

Risks were computed using linear regression (for prevalence) or the cumulative incidence from Fine-Gray models (for intubation and death). Risk differences and ratios are reported relative to O ABO blood type and positive Rh(D) type.



**Fig. 1** Estimated risk differences for blood types during the period from March 10 to August 1, 2020. Values represent risk differences for each blood type relative to the reference groups: O for ABO and positive for Rh(D). Prevalence differences were computed using linear regression, while intubation and death were computed using the Fine-Gray model. Estimated differences are represented as points. 95% confidence intervals (CI, represented as bars) were computed using the Austin’s method with  $n = 1000$  bootstrap iterations. Adjusted models include race and ethnicity as covariates.

stay, and potential in-patient infection, we evaluated the prevalence of infection among individuals only during the first encounters in which they were tested. In addition, to account for the considerable risk of false-negative tests<sup>53,54</sup> and the fact that providers would repeat the test in patients with high clinical suspicion for COVID-19<sup>55</sup>, any positive test during the first 96 h of an encounter was considered evidence of initial infection.

Blood-type frequencies vary across ancestry groups<sup>56</sup>, so we evaluated the confounding effect of ancestry by adjusting for race/ethnicity (proxies for ancestry). We compared infection prevalence with linear regression, using reference groups O for ABO and Rh-positive for Rh(D) and using bootstrap to compute 95% confidence intervals for each estimate<sup>57</sup>. With adjustment for patient race and ethnicity, prevalences among types A, AB, and B were higher than type O. Rh(D)-negative individuals had a 2.7% lower risk of initial infection after adjustment for ancestry, consistent with a lower risk before adjustment.

**Survival analysis for intubation and death.** Next, we examined intubation and death using a survival framework to understand

how blood type affects progression to disease outcomes over time. Specifically, we used the Fine-Gray model<sup>58</sup> to estimate cumulative incidence functions by blood type while accounting for competing risks and adjusting covariates. Death and recovery were competing events for intubation, and recovery was a competing event for death. Cohort entry was defined as the time of a patient’s first positive test result or the start of a hospital encounter if the first positive test occurred during the first 96 h of the hospitalization. In accordance with CDC guidelines for returning to work<sup>59</sup>, we defined a patient as having recovered only when 10 days had passed since the patient entered the cohort and only once the patient had been discharged. Patients appearing before July 30 were considered, and outcomes beyond August 1, 2020 were censored.

Blood type A was at decreased risk of both intubation and death relative to type O, while type AB was at increased risk of both outcomes (Fig. 1 and Table 2). Conversely, we found that type-B individuals were at higher risk of intubation but at lower risk of death, compared with type O. Individuals negative for Rh(D) were at decreased risk for both intubation and death,

consistent with a lower risk of initial infection. Overall, we estimate between 0.1 and 8.2% absolute risk differences between blood groups, after adjusting for race and ethnicity.

## Discussion

Better understanding COVID-19 is imperative given the current pandemic's toll. We investigated whether blood type is relevant for risk of infection, intubation, and death. Overall, we found modest but consistent risk differences between blood types. After adjusting for ancestry (the relevant confounder for this analysis), estimated risk differences were larger for intubation and death outcomes than for initial infection. We estimate larger risk differences between Rh blood types than between ABO types, with Rh-negative individuals being at lower risk of all three outcomes. Type A had lower risk of intubation and death compared with types AB and O. Only type B had inconsistent effects between intubation and death—type B increased risk of intubation and decreased risk of death compared to type O. We also found consistent evidence for protective associations between Rh-negative blood groups and SARS-CoV-2 infection, intubation, death. Overall, blood type appears to have a consistent effect, though the magnitudes of these effects on risk of intubation or death are modest, and our estimates have large uncertainties relative to their magnitudes. The relatively large estimated errors in our analysis also suggest modest effect sizes and that greater sample sizes or meta-analyses are needed to estimate these effects more precisely.

After adjusting for ancestry by proxies of race and ethnicity, we found that types A and B conferred greater risk of an initial positive test compared to type O, while type AB (the rarest), conferred a very-small risk decrease (0.2%). These results are consistent with an association discovered for SARS-CoV-1, in which O blood groups were less common among SARS patients<sup>43</sup>. Our results are also mostly consistent with the results reported by Zhao et al.<sup>5</sup>, where non-O types appear to be at greater risk of infection, and with Ellinghaus et al.<sup>6</sup>, where non-O appears to be at greater risk of infection but at lesser risk of mechanical ventilation, though the authors note that this decreased risk is not statistically significant at the 5% level. Unlike Ellinghaus et al., though, we estimate slightly higher risk for types B and AB relative to O for intubation.

Our results are based on data collected as part of hospital care during the early course of the pandemic, where outpatient testing was severely limited due to testing capacity and supply limitations. As such, our data are highly enriched for severely ill patients, and the absolute risk values we report are not generalizable to all SARS-CoV-2-infected individuals. A considerable fraction of infections is mild or asymptomatic<sup>60–63</sup>, while our data represent predominantly the most severe cases. Selection bias is a fundamental limitation of our study, so all our effect estimates are conditional on presentation to the hospital. Nonetheless, we minimized additional selection bias by making cohort criteria for cases and controls differ only with respect to the outcome of interest. Moreover, we found concordance between SARS-CoV-2-tested individuals and the general population at NYP/CUIMC in terms of blood type (Supplementary Table 1). Consequently, our results are not affected by selection bias with respect to blood type, unlike some other blood-type case/control study designs—particularly those using blood donors as controls, where enrichment of type O can be expected<sup>6</sup>.

False negatives and time delay between test administrations and the return of their results both introduce noise to this analysis. We attempted to account for these biases by setting cohort entry at the time of first contact with the hospital when the patient tested positive <96 h thereafter. This definition is

imperfect, as 96 h is sufficient for an individual infected shortly after admission to test positive (albeit with probability roughly 0.33)<sup>64</sup>, but it is necessary to allow sufficient adjustment for the considerable time delay and retesting following false negatives. Another source of noise is the fact that not all intubations and deaths following a confirmed infection are related to COVID-19 (e.g., intubation during unrelated surgery). We defined recovery in an attempt to minimize this issue, though we recognize that our definition is imperfect. Patients may be discharged prematurely and later return following onset of severe symptoms. Moreover, our 10-day cutoff for recovery is based on CDC guidelines for returning to work<sup>59</sup>, which may be refined as additional evidence becomes available. Further work is needed to refine the definition of recovery and to determine which outcomes may be causally linked to COVID-19.

The ABO gene is highly polymorphic<sup>65</sup>, and blood types have considerably different distributions across ancestry groups<sup>56</sup>. Like ABO, Rh groups are not distributed equally across race/ethnicity groups, with enrichment of Rh-negative among white and non-Hispanic individuals (Table 1). In addition, negative Rh blood groups are less common, representing only 9% of individuals in our data. While genetic data were not available for the patients included in our study, we used self-reported race and ethnicity as imperfect proxies for genetic ancestry. Adjusting for these covariates had a noticeable effect on our comparison of infection prevalence, but did not have an equally relevant effect on intubation or death (Fig. 1 and Table 2). This suggests that blood type may have a lesser, more confounded effect on infection prevalence than on intubation or death following confirmed infection. Nonetheless, race and ethnicity cannot fully capture ancestry, so the associations between blood types and COVID-19 that we report may still be confounded by ancestry, even after adjustment. Further work is needed to better understand any potential residual confounding due to ancestry, not captured by race and ethnicity.

In this study we found evidence for associations between ABO and Rh blood groups and COVID-19. Using data from NYP/CUIMC, we found moderately increased infection prevalence among non-O blood types and among Rh-positive individuals. Intubation risk was increased among AB and B types, and decreased among A and Rh-negative types. Risk of death was slightly increased among type AB individuals and was decreased among types A, B, and Rh-negative types. All estimates were adjusted for patient ancestry using self-reported race and ethnicity. Our results add further evidence to the previously discovered associations between blood types and COVID-19.

## Methods

**Data collection and cohort selection.** We identified the cohort for this study by filtering the NYP/CUIMC data warehouse for patients with a recorded SARS-CoV-2 test and those having a recorded blood type. Next, we removed any individual with multiple, contradictory blood-type measurements, reflecting likely errors in the data. Finally, we excluded individuals below age 18 from our analysis.

Blood group was determined using laboratory measurements coded using descendant concepts of LOINC LP36683-8 (ABO and Rh group). Intubation was assessed using completed procedures having the procedure description, “intubation”. We grouped race into five categories and ethnicity into two. 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 grouped as either Hispanic or non-Hispanic. This study is approved by the Columbia University IRB (#AAAL0601).

**Covariate adjustment.** We sought to estimate total effects of blood type on COVID-19 outcomes. Using a graphical model (Supplementary Fig. 2), we identified ancestry as the only confounding variable for an estimate of total effect, since blood type is genetic and varies across ancestry groups. As genetic data were not available, we used self-reported race and ethnicity as proxies for ancestry. We were

unable to identify a method to alleviate selection bias in our data, so the effects we report are conditional on presence at NYP/CUIMC.

**Infection prevalence.** We considered three outcomes: initial infection, intubation, and death. Our evaluation of initial infection sought to assess the infection prevalence differences among individuals presenting to the hospital, not those potentially infected at the hospital or long after their first test. Due to the high risk for false negatives<sup>53,54</sup>, we considered any positive test <96 h after the start of an encounter as evidence of initial infection. Initial infection risk differences between blood types were assessed using linear regression, and race/ethnicity were adjusted by including them as covariates. We used Austin's bootstrap method to compute 95% confidence intervals for all risk estimates<sup>57</sup>, using 1000 bootstrap iterations.

**Survival analysis for severe outcomes.** We assessed intubation and death as severe outcomes of COVID-19, and evaluated blood-type effects using survival analysis. Individuals entered the at-risk cohort either at the time of their first positive test result, or at the time of first contact with the hospital when the first positive test occurred within 96 h of the start of a hospital encounter. Patients with Do-Not-Intubate orders were excluded from consideration for the intubation outcome. We defined a patient as recovered only after being discharged from the hospital and only once 10 days have passed since cohort entry. Death and recovery are competing risks for intubation, and recovery is a competing risk for death. Finally, outcomes beyond August 1 were censored, as this was the last date for which we have data available. Intubation and death were assessed using Fine-Gray models, which can estimate cumulative incidences. As before, race and ethnicity were adjusted by including them as covariates, and confidence intervals were computed with 1000 bootstrap iterations.

**Software.** We conducted our analyses using the R language (version 4.0.1), the *cmprsk* (version 2.2-10) package<sup>66</sup> implementation of the Fine-Gray model, MySQL version 5.6, and tidyverse meta-package version 1.1.2. The manuscript was written openly on GitHub using Manubot<sup>67</sup>.

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

## Data availability

We used patient clinical data, which is protected in the United States under 1996 Public Law 104-191 (HIPAA) against public release. Requests for additional materials can be made via email to the corresponding author.

## Code availability

Code used for our analysis is available on GitHub<sup>68</sup> at [https://github.com/zietzm/abo\\_covid\\_analysis](https://github.com/zietzm/abo_covid_analysis).

Received: 15 April 2020; Accepted: 16 October 2020;

Published online: 13 November 2020

## References

- World Health Organization. *Coronavirus Disease 2019 (COVID-19) Weekly Epidemiological Update 1*. <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200817-weekly-epi-update-1.pdf> (2020).
- New York City Department of Health and Mental Hygiene. NYC Coronavirus Disease 2019 (COVID-19) data. GitHub. <https://doi.org/10.5281/zenodo.4073817> (2020).
- 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).
- Zheng, Z. et al. Risk factors of critical & mortal COVID-19 cases: a systematic literature review and meta-analysis. *J. Infect.* **81**, e16–e25 (2020).
- Zhao, J. et al. Relationship between the ABO blood group and the COVID-19 susceptibility. *medRxiv*. <https://doi.org/10.1101/2020.03.11.20031096> (2020).
- Ellinghaus, D. et al. Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa2020283> (2020).
- Chung, C.-M. et al. A genome-wide association study identifies new loci for ACE activity: potential implications for response to ACE inhibitor. *Pharmacogenomics J.* **10**, 537–544 (2010).
- Kamatani, Y. et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet.* **42**, 210–215 (2010).
- Astle, W. J. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429 (2016).
- van Rooij, F. J. A. et al. Genome-wide trans-ethnic meta-analysis identifies seven genetic loci influencing erythrocyte traits and a role for RBPMS in erythropoiesis. *Am. J. Hum. Genet.* **100**, 51–63 (2017).
- Kanai, M. et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
- Smith, N. L. et al. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor. *Circulation* **121**, 1382–1392 (2010).
- Williams, F. M. K. et al. Ischemic stroke is associated with the ABO locus: The EuroCLOT study. *Ann. Neurol.* **73**, 16–31 (2013).
- van Loon, J. et al. Genome-wide association studies identify genetic loci for low von Willebrand factor levels. *Eur. J. Hum. Genet.* **24**, 1035–1040 (2015).
- Sabater-Lleal, M. et al. Genome-wide association transethnic meta-analyses identifies novel associations regulating coagulation factor VIII and von Willebrand factor plasma levels. *Circulation* **139**, 620–635 (2019).
- Reilly, M. P. et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet* **377**, 383–392 (2011).
- Nikpay, Majid et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
- Schunkert, H. et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
- Dichgans, M. et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease. *Stroke* **45**, 24–36 (2014).
- Nelson, C. P. et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385–1391 (2017).
- van der Harst, P. & Verweij, N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ. Res.* **122**, 433–443 (2018).
- Malik, R. et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
- Scott, R. A. et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* **66**, 2888–2902 (2017).
- Xue, A. et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9**, 2941 (2018).
- Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
- Trégouët, D.-A. et al. Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. *Blood* **113**, 5298–5303 (2009).
- Germain, M. et al. Genetics of venous thrombosis: insights from a new genome wide association study. *PLoS ONE* **6**, e25581 (2011).
- Heit, J. A. et al. A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. *J. Thrombosis Haemost.* **10**, 1521–1531 (2012).
- Tang, W. et al. A genome-wide association study for venous thromboembolism: the extended Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. *Genet. Epidemiol.* **37**, 512–521 (2013).
- Germain, M. et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *Am. J. Hum. Genet.* **96**, 532–542 (2015).
- Klarin, D., Emdin, C. A., Natarajan, P., Conrad, M. F. & Kathiresan, S. Genetic analysis of venous thromboembolism in UK Biobank identifies the ZFPM2 locus and implicates obesity as a causal risk factor. *Circ. Cardiovasc. Genet.* **10**, e001643 (2017).
- Lindström, S. et al. Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. *Blood* **134**, 1645–1657 (2019).
- Klarin, D. et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat. Genet.* **51**, 1574–1579 (2019).
- Dentali, F. et al. Non-O blood type is the commonest genetic risk factor for VTE: results from a meta-analysis of the literature. *Semin. Thrombosis Hemost.* **38**, 535–548 (2012).
- Tang, N. et al. Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy. *J. Thrombosis Haemost.* **18**, 1094–1099 (2020).
- Llitijs, J. et al. High incidence of venous thromboembolic events in anticoagulated severe COVID-19 patients. *J. Thrombosis Haemost.* **18**, 1743–1746 (2020).

37. Middeldorp, S. et al. Incidence of venous thromboembolism in hospitalized patients with COVID-19. *J. Thrombosis Haemost.* **18**, 1995–2002 (2020).
38. Lodigiani, C. et al. Venous and arterial thromboembolic complications in COVID-19 patients admitted to an academic hospital in Milan, Italy. *Thrombosis Res.* **191**, 9–14 (2020).
39. Levi, M., Thachil, J., Iba, T. & Levy, J. H. Coagulation abnormalities and thrombosis in patients with COVID-19. *Lancet Haematol.* **7**, e438–e440 (2020).
40. Spiezia, L. et al. COVID-19-related severe hypercoagulability in patients admitted to intensive care unit for acute respiratory failure. *Thrombosis Haemost.* **120**, 998–1000 (2020).
41. Wichmann, D. et al. Autopsy findings and venous thromboembolism in patients with COVID-19. *Ann. Intern. Med.* **173**, 268–277 (2020).
42. Khan, I. H., Savarimuthu, S., Leung, M. S. T. & Harky, A. The need to manage the risk of thromboembolism in COVID-19 patients. *J. Vasc. Surg.* **72**, 799–804 (2020).
43. Cheng, Y. et al. ABO blood group and susceptibility to severe acute respiratory syndrome. *JAMA* **293**, 1447 (2005).
44. Loscertales, M. et al. ABO blood group phenotypes and Plasmodium falciparum malaria: unlocking a pivotal mechanism. *Adv. Parasitol.* **65**, 1–50 (2007).
45. Boren, T., Falk, P., Roth, K., Larson, G. & Normark, S. Attachment of Helicobacter pylori to human gastric epithelium mediated by blood group antigens. *Science* **262**, 1892–1895 (1993).
46. Lindesmith, L. et al. Human susceptibility and resistance to Norwalk virus infection. *Nat. Med.* **9**, 548–553 (2003).
47. Wang, D.-S. et al. ABO blood group, hepatitis B viral infection and risk of pancreatic cancer. *Int. J. Cancer* **131**, 461–468 (2012).
48. Foster, M. T. & Labrum, A. H. Relation of infection with Neisseria gonorrhoeae to ABO blood groups. *J. Infect. Dis.* **133**, 329–330 (1976).
49. Anstee, D. J. The relationship between blood groups and disease. *Blood* **115**, 4635–4643 (2010).
50. Sandler, S. G., Chen, L. N. & Flegel, W. A. Serological weak D phenotypes: a review and guidance for interpreting the RhD blood type using the RHD genotype. *Br. J. Haematol.* **179**, 10–19 (2017).
51. Flegel, J. Influence of latent Toxoplasma infection on human personality, physiology and morphology: pros and cons of the Toxoplasma-human model in studying the manipulation hypothesis. *J. Exp. Biol.* **216**, 127–133 (2012).
52. Dardé, M. L., Ajzenberg, D. & Smith, J. Population structure and epidemiology of Toxoplasma gondii. In: *Toxoplasma gondii the Model Apicomplexan: Perspectives and Methods*, (eds Weiss, L. M. & Kim, K.) (Academic Press, London, 2007).
53. Xiao, A. T., Tong, Y. X. & Zhang, S. False negative of RT-PCR and prolonged nucleic acid conversion in COVID-19: Rather than recurrence. *J. Med. Virol.* **92**, 1755–1756 (2020).
54. West, C. P., Montori, V. M. & Sampathkumar, P. COVID-19 testing. *Mayo Clin. Proc.* **95**, 1127–1129 (2020).
55. Winichakoon, P. et al. Negative nasopharyngeal and oropharyngeal swabs do not rule out COVID-19. *J. Clin. Microbiol.* **58**, e00297–20 (2020).
56. 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).
57. Austin, P. C. Absolute risk reductions and numbers needed to treat can be obtained from adjusted survival models for time-to-event outcomes. *J. Clin. Epidemiol.* **63**, 46–55 (2010).
58. Fine, J. P. & Gray, R. J. A proportional hazards model for the subdistribution of a competing risk. *J. Am. Stat. Assoc.* **94**, 496–509 (1999).
59. Centers for Disease Control and Prevention. Duration of isolation and precautions for adults with COVID-19. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html> (2020).
60. Nishiura, H. et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int. J. Infect. Dis.* **94**, 154–155 (2020).
61. He, X. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
62. Yang, R., Gui, X. & Xiong, Y. Comparison of clinical characteristics of patients with asymptomatic vs symptomatic Coronavirus Disease 2019 in Wuhan, China. *JAMA Netw. Open* **3**, e2010182 (2020).
63. Kim, G.-u et al. Clinical characteristics of asymptomatic and symptomatic patients with mild COVID-19. *Clin. Microbiol. Infect.* **26**, 948.e1–948.e3 (2020).
64. Kucirka, L. M., Lauer, S. A., Laeyendecker, O., Boon, D. & Lessler, J. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. *Ann. Intern. Med.* **173**, 262–267 (2020).
65. Yip, S. Sequence variation at the human ABO locus. *Ann. Hum. Genet.* **66**, 1–27 (2002).
66. Gray, R. J. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Ann. Stat.* **16**, 1141–1154 (1988).
67. Himmelstein, D. S. et al. Open collaborative writing with Manubot. *PLOS Comput. Biol.* **15**, e1007128 (2019).
68. Zietz, M. zietzm/abo\_covid\_analysis. GitHub. <https://doi.org/10.5281/zenodo.4072059> (2020).

## Acknowledgements

We thank Ben May and Vijendra Ramlall for assistance with data collection and periodic updates to the patient data and Nicholas Giangreco, Undina Gisladdottir, Kathleen LaRow Brown, and Dr Phyllis Thangaraj for helpful discussions regarding clinical definitions. M.Z. is funded by T15LM007079, and N.P.T. is funded by R35GM131905.

## Author contributions

M.Z. and N.P.T. conceived and designed the study. M.Z. carried out the statistical analysis with advice from N.P.T. J.Z. and M.Z. created the cohort entry and exit criteria. M.Z. and N.P.T. wrote, revised, and approved the final version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-19623-x>.

**Correspondence** and requests for materials should be addressed to N.P.T.

**Peer review information** *Nature Communications* thanks Rachel Graham, Hamish Nair, Charles Poole, and Peng George Wang for their contribution to the peer review of this work. Peer review reports are available.

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

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020