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

Authors

- **Michael Zietz**
  0000-0003-0539-630X · zietzm · ZietzMichael
  Department of Biomedical Informatics, Columbia University Irving Medical Center ·
  Funded by NIH T15 LM007079

- **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)

Abstract

The rapid global spread of the novel coronavirus SARS-CoV-2 has strained existing healthcare and testing resources, making the identification and prioritization of individuals most at-risk a critical challenge. A recent study of patients in China discovered an association between ABO blood type and SARS-CoV-2 infection status by comparing COVID-19 patients with the general population. Whether blood type is associated with increased COVID-19 morbidity or mortality remains unknown. We used observational healthcare data on 1559 individuals tested for SARS-CoV-2 (682 COV+) with known blood type in the New York Presbyterian (NYP) hospital system to assess the association between ABO+Rh blood type and SARS-CoV-2 infection status, intubation, and death. We found a higher proportion of blood group A and a lower proportion of blood group O among COV+ patients compared to COV-, though in both cases the result is significant only in Rh positive blood types. We show that the effect of blood type is not explained by risk factors we considered (age, sex, hypertension, diabetes mellitus, overweight status, and chronic cardiovascular and lung disorders). In a meta-analysis of NYP data with previously-reported data from China, we find enrichment for A and B and depletion of O blood groups among COVID-19 patients compared to the general population. Our data do not provide strong evidence of associations between blood group and intubation or death among COVID-19 patients. In this preliminary observational study of data currently being collected during the outbreak, we find new evidence of associations between B, AB, and Rh blood groups and COVID-19 and further evidence of recently-discovered associations between A and O blood groups and 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 1,130,000 confirmed infections and over 62,000 deaths worldwide as of April 5, 2020 [1]. A number of risk factors for COVID-19 infection, morbidity, and mortality are known, including age, sex, and a number of 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]. Their analysis compared blood groups between hospitalized COVID-19 patients and the general populations of Wuhan and Shenzhen City, as assessed by previously-published samples of healthy individuals. They found that the odds of testing positive for COVID-19 among A blood groups was increased and among O blood groups was decreased relative to the general population. Similarly, previous work has identified associations between ABO blood groups and a number of different infections or disease severity following infections, including 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 a major center of the pandemic, with over 64,000 cases and over 2,400 deaths as of April 5, 2020 [10]. We sought to replicate and extend the previous investigation into the association between COVID-19 and blood type using electronic health record data from New York Presbyterian/Columbia University Irving Medical Center (NYP/CUIMC) hospital in New York, USA. We compared both ABO and ABO+Rh blood types, and we investigated three COVID-19 outcomes: infection status, intubation, and death. We performed a multivariate analysis of our results to evaluate potential confounding due to risk factors, and we meta-analyzed our results in combination with data from China. This study is approved by the IRB (#AAAL0601).

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 group was identified using either a measurement of LOINC code 34474-7, "ABO and Rh group [Type] in Cord blood," or the results of a procedure identified by one of the names listed in Table 4. We excluded individuals with multiple contradictory blood group measurements.

We compared blood groups (defined both as ABO 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 eight test conditions (2 blood group definitions and 4 outcome comparison population pairs), we performed a Pearson's Chi-squared test to test whether blood group distributions differ between the compared populations. Additionally, we compared each blood group against all others using a 2x2 contingency table to determine effect sizes for each blood group itself. For the one-vs-rest blood group comparisons, we report odds ratios (OR), p-values from Fisher's exact test (two-sided), and odds ratio confidence intervals.

We evaluated the confounding effect of risk factors (age, sex, overweight status, diabetes mellitus, hypertension, pulmonary diseases, and cardiovascular 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 (Table 5).

First, we evaluated associations between risk factors and blood groups using logistic regressions of risk factors on blood groups. 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 going to zero 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 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, and Shenzhen. The distribution of blood groups in the general population was estimated using blood group lab results on 108,860 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 distributions between New York and Wuhan and Shenzhen and evaluated the heterogeneity between sites.

We considered EHR data up to April 5, 2020. We conducted our analyses using the R language, using the `meta` package [11] for meta-analysis. While our data from NYP/CUIMC are protected by HIPAA and cannot be released, we have made all code used for our analysis available at https://github.com/zietzm/abo_covid_analysis. The manuscript was written openly on GitHub using Manubot [12].

Results

We first determined blood groups for SARS-CoV-2-tested individuals using laboratory measurements recorded in the NYP/CUIMC EHR system. One individual with multiple contradictory blood group measurements was excluded, resulting in 1,559 individuals with known blood groups who received a SARS-CoV-2 test (either positive or negative result). Of these, 682 were COV+ (positive in at least one SARS-CoV-2 test) and 877 were COV- (negative in all SARS-CoV-2 tests administered). Among the COV+ individuals, 179 were intubated and 80 had died, while the remaining individuals had not been intubated and were alive as of April 5, 2020. We found that 354 tested COV- individuals were intubated during the same time, though we did not include them in any analysis. Table 1 gives a summary of the cohort we considered.

Table 1: Summary demographics for SARS-CoV-2-tested individuals at NYP/CUIMC, stratified by blood group (BG). N is the number of individuals having the given blood type who have a

recorded test (positive or negative) for SARS-CoV-2, and reports percentages relative to all blood groups. Rh-stratified blood groups are individuals from the ABO blood groups (eg. for N, A-negative + A-positive = A). Med. age gives the median age within each group. COV+ and COV- give percentages relative to all tested individuals for that blood group. COV+/Intubated (intub.) and COV+/Died report percentages relative to COV+ individuals. Risk factors (hypertension (HTN), cardiovascular (CV) diseases mellitus (DM), respiratory (resp) diseases, diabetes, overweight status (OW)) are reported as percentages relative to all tested individuals of that blood group. The general population (gen) column reports counts and percentages by blood group relative to the non-SARS-CoV-2-tested individuals from NYP/CUIMC with recorded blood type. The final row (Σ) gives combined results without stratification by blood type.

BG	N	Male	Med age	COV+	COV+/ Intub.	COV+/ Died	HTN	CV dis.	Resp. dis.	DM	OW	Gen
A	478 (30.7%)	219 (45.8%)	61.4	233 (48.7%)	62 (26.6%)	27 (11.6%)	233 (48.7%)	285 (59.6%)	278 (58.2%)	146 (30.5%)	51 (10.7%)	35643 (32.7%)
A-	43 (2.8%)	20 (46.5%)	51.7	17 (39.5%)	2 (11.8%)	1 (5.9%)	17 (39.5%)	25 (58.1%)	22 (51.2%)	15 (34.9%)	2 (4.7%)	3447 (3.2%)
A+	435 (27.9%)	199 (45.7%)	61.8	216 (49.7%)	60 (27.8%)	26 (12%)	216 (49.7%)	260 (59.8%)	256 (58.9%)	131 (30.1%)	49 (11.3%)	32196 (29.6%)
AB	68 (4.4%)	34 (50%)	58.8	21 (30.9%)	8 (38.1%)	5 (23.8%)	23 (33.8%)	31 (45.6%)	30 (44.1%)	14 (20.6%)	5 (7.4%)	4582 (4.2%)
AB- *	5 (0.3%)	4 (80%)	56.8	0 (0%)	0	0	1 (20%)	2 (40%)	2 (40%)	0 (0%)	1 (20%)	394 (0.4%)
AB+	63 (4%)	30 (47.6%)	58.9	21 (33.3%)	8 (38.1%)	5 (23.8%)	22 (34.9%)	29 (46%)	28 (44.4%)	14 (22.2%)	4 (6.3%)	4188 (3.8%)
B	252 (16.2%)	107 (42.5%)	52.7	116 (46%)	35 (30.2%)	12 (10.3%)	114 (45.2%)	143 (56.7%)	136 (54%)	58 (23%)	31 (12.3%)	16229 (14.9%)
B-	21 (1.3%)	8 (38.1%)	59.1	7 (33.3%)	2 (28.6%)	0 (0%)	7 (33.3%)	9 (42.9%)	8 (38.1%)	3 (14.3%)	1 (4.8%)	1422 (1.3%)
B+	231 (14.8%)	99 (42.9%)	51.8	109 (47.2%)	33 (30.3%)	12 (11%)	107 (46.3%)	134 (58%)	128 (55.4%)	55 (23.8%)	30 (13%)	14807 (13.6%)
O	761 (48.8%)	356 (46.8%)	55.0	312 (41%)	74 (23.7%)	36 (11.5%)	354 (46.5%)	419 (55.1%)	404 (53.1%)	227 (29.8%)	101 (13.3%)	52406 (48.1%)
O-	47 (3%)	23 (48.9%)	58.2	21 (44.7%)	4 (19%)	0 (0%)	28 (59.6%)	29 (61.7%)	26 (55.3%)	14 (29.8%)	2 (4.3%)	4808 (4.4%)
O+	714 (45.8%)	333 (46.6%)	54.8	291 (40.8%)	70 (24.1%)	36 (12.4%)	326 (45.7%)	390 (54.6%)	378 (52.9%)	213 (29.8%)	99 (13.9%)	47598 (43.7%)
Σ	1559 (100%)	716 (45.9%)	56.8	682 (43.7%)	179 (26.2%)	80 (11.7%)	724 (46.4%)	878 (56.3%)	848 (54.4%)	445 (28.5%)	188 (12.1%)	108860 (100%)

* AB-negative was not included in the ABO+Rh analyses as no individuals with that blood type recorded tested positive for COVID-19.

For each comparison cohort pair, we performed chi-squared tests using both ABO and ABO+Rh blood types (Table 6). Since there were no AB-negative individuals testing positive for COVID-19 and only 5 individuals testing negative, we excluded AB-negative from all ABO+Rh analyses. Finally, we conducted individual tests of each blood type against all others (within the

same ABO vs ABO+Rh system) for each of the COVID-19 outcomes we considered (Tables 2, 7).

We found associations between COVID-19 status and both ABO (p=0.006) and ABO+Rh (p=0.031) blood groups in a comparison between individuals testing positive vs testing negative (Tables 2, 6). Blood groups A were associated with increased odds of testing positive for COVID-19 (OR 1.338, 95% CI [1.072-1.672], p=0.009), while O blood groups were associated with decreased odds of testing positive (OR 0.804, 95% CI [0.654-0.987], p=0.036). While few individuals with AB blood groups were included (21 COV+, 47 COV-), we also found AB blood groups to be associated with decreased odds of testing positive (OR 0.561, 95% CI [0.315-0.969], p=0.033). When we tested individual ABO+Rh blood groups against all others, we discovered that strong associations are only found in Rh positive blood groups (Tables 2, 7). We did not find any significant associations between blood group and intubation or death. Finally, we compared the blood group distributions between all individuals tested for SARS-CoV-2 with the general population at NYP/CUIMC, finding insufficient evidence to conclude that tested individuals are not drawn from the general population at random (p=0.2736).

Table 2: Summary of COV+ vs COV- comparison of individual blood types. Each test compares the listed blood group to all other blood groups (combined) between the COV+ and COV- individuals. OR means odds ratio (COV+ vs COV-), and the 95% CI is a confidence interval on the OR. P-values computed using Fisher's exact test.

Blood group	COV+ counts	COV- counts	OR	95% CI	p-value
A	A:233 (34.2%), ¬A:449 (65.8%)	A:245 (27.9%), ¬A:632 (72.1%)	1.338	1.072 - 1.672	**0.009**
A-negative	A-:17 (2.5%), ¬A-:665 (97.5%)	A-:26 (3%), ¬A-:846 (97%)	0.832	0.42 - 1.608	0.641
A-positive	A+:216 (31.7%), ¬A+:466 (68.3%)	A+:219 (25.1%), ¬A+:653 (74.9%)	1.382	1.099 - 1.737	**0.004**
AB	AB:21 (3.1%), ¬AB:661 (96.9%)	AB:47 (5.4%), ¬AB:830 (94.6%)	0.561	0.315 - 0.969	**0.033**
AB-positive	AB+:21 (3.1%), ¬AB+:661 (96.9%)	AB+:42 (4.8%), ¬AB+:830 (95.2%)	0.628	0.35 - 1.097	0.093
B	B:116 (17%), ¬B:566 (83%)	B:136 (15.5%), ¬B:741 (84.5%)	1.117	0.843 - 1.477	0.446
B-negative	B-:7 (1%), ¬B-:675 (99%)	B-:14 (1.6%), ¬B-:858 (98.4%)	0.636	0.216 - 1.695	0.381
B-positive	B+:109 (16%), ¬B+:573 (84%)	B+:122 (14%), ¬B+:750 (86%)	1.169	0.874 - 1.563	0.282
O	O:312 (45.7%), ¬O:370 (54.3%)	O:449 (51.2%), ¬O:428 (48.8%)	0.804	0.654 - 0.987	**0.036**

O-negative	O-:21 (3.1%), ¬O-:661 (96.9%)	O-:26 (3%), ¬O-:846 (97%)	1.034	0.548 - 1.93	1.000
O-positive	O+:291 (42.7%), ¬O+:391 (57.3%)	O+:423 (48.5%), ¬O+:449 (51.5%)	0.790	0.642 - 0.971	**0.024**

Multivariate analysis of blood group associations

We find that the significant associations we report cannot be explained by known risk factors. We considered age, sex, overweight status, diabetes mellitus, hypertension, pulmonary diseases, and cardiovascular diseases. First, we found significant associations between blood groups and risk factors. Using blood group ~ risk factors logistic regressions for each blood group, we found significant associations between hypertension and O- blood groups, between age and A, A+, AB, AB+, O, and O+ blood groups, between diabetes mellitus and B and A- blood groups, as well as between overweight status and O+ blood groups (Table 8).

We found that blood groups provide significant additional information on outcomes beyond risk factors. First, we verified that the risk factors (outcome ~ risk factors) provide predictive power compared to intercept-only null models (Table 9). Then, we evaluated blood group's effect on COVID-19 outcomes beyond risk factors using analysis of deviance. Specifically, we compared models including both risk factors and blood group (outcome ~ blood group + risk factors) to those including only risk factors (outcome ~ risk factors), and we found that only the COVID-19 test result outcome (COV+ vs COV-) is significantly better explained by including blood group in addition to risk factors ($p<0.02$ for both ABO and ABO+Rh, Table 9). This result is consistent with our univariate analysis (Table 6), where only COV+ vs COV- cohorts showed significant differences in blood group distribution.

Finally, we showed that blood group's effects are not modulated by risk factors. To do so, we inspected individual blood group coefficients between nested (outcome ~ blood group) and full (outcome ~ blood group + risk factors) logistic regression models. For ABO blood groups, we found no large coefficient changes, with some coefficients even being more extreme (greater in magnitude and more significant) in full models than in nested models (Table 10). No ABO+Rh blood groups were significant in either the nested or full models for ABO+Rh, though we found no large coefficient changes for these blood groups either. We find little evidence for conditional independence between outcomes and blood groups given risk factors.

Meta-analysis

Finally, we compared our data from New York City to the data from Wuhan and Shenzhen presented by Zhao et al. [3] and conducted a meta-analysis. Zhao et al. used a random effects model to weight and pool effects between three different hospitals (Wuhan Jinyintan, Renmin Hospital in Wuhan, and Shenzhen Third People's Hospital), comparing each hospital's COV+ blood group distribution to the general population distribution for each city. We performed a similar analysis—including NYP/CUIMC data—to assess the effect of blood type in the combined data from all four sources (full counts in Table 3).

We fit a random effects model for each ABO blood type using data from NYP/CUIMC and the three sources for which Zhao et al. report data. The overall associations between ABO blood

groups and COVID-19 status that Zhao et al. identified (significantly increased COV+ odds for blood group A and decreased COV+ odds for blood group O) are replicated in our meta-analysis (Table 3). Using the additional data from NYP/CUIMC, the pooled association between blood group B becomes larger in effect size and significant at the 5% level (original: OR 1.09, p=0.121; with NYP/CUIMC data: OR 1.25, p=0.0361).

Table 3: 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) and Wuhan (cases from Wuhan Jinyintan Hospital and Renmin Hospital of Wuhan University, controls from Wuhan general population). Shenzhen and Wuhan data reported by Zhao et al. [3] (Rh groups not reported). 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 refers to the pooled odds ratio (COV+ vs general population), and the 95% CI is a confidence interval on the OR. P-values are for the pooled association from the random effects model.

Blood group	NYP general pop.	NYP COV+	Shenzhen general pop.	Shenzhen COV+	Wuhan general pop.	Wuhan Jinyintan COV+	Wuhan Renmin COV+	OR	95% CI	p-value
A	32.7% (35643)	34.2% (233)	28.8% (6728)	28.8% (82)	32.2% (1188)	37.7% (670)	39.8% (45)	1.164	1.015 - 1.333	**0.0291**
AB	4.2% (4582)	3.1% (21)	7.3% (1712)	13.7% (39)	9.1% (336)	10% (178)	13.3% (15)	1.2519	0.8384 - 1.8694	0.2721
B	14.9% (16229)	17% (116)	25.1% (5880)	29.1% (83)	24.9% (920)	26.4% (469)	22.1% (25)	1.1101	1.0068 - 1.2240	**0.0361**
O	48.1% (52406)	45.7% (312)	38.8% (9066)	28.4% (81)	33.8% (1250)	25.8% (458)	24.8% (28)	0.7252	0.5971 - 0.8807	**0.0012**

Finally, we found significant heterogeneity among the sites we considered in the meta-analysis. We found that the distribution of blood groups in the general population at NYP/CUIMC differs significantly from both the distributions in Shenzhen (Chi-squared = 2056, p-value < 2.2e-16) and Wuhan (Chi-squared = 583.29, p-value < 2.2e-16) (Table 3). The difference in distributions is reflected in tests of heterogeneity between sites, where we find more heterogeneity between sites in our meta-analysis than Zhao et al.'s meta-analysis (Table 12).

Discussion

We found consistent negative associations between O blood groups and COVID-19. These results are consistent with an association discovered for SARS-CoV-1 [4], in which O blood groups were significantly less common among SARS patients.

Our results from NYP/CUIMC identified significant associations between A, AB, and O blood groups. However, further stratifying by Rh resulted in significant associations for A+ and O+ only. Negative Rh blood groups are less common in our data, representing only 9.25% of individuals, so the lack of evidence for association with negative blood types could be due to lower sample sizes. Yet, odds ratios for ABO groups A and O are less extreme than the associated ABO+Rh blood groups (A+, O+), and the corresponding negative blood groups (A-, O-) have (insignificant) odds in the opposite directions as their positive counterparts. Further work is needed to better understand the associations between Rh negative blood groups and COVID-19.

Since both blood groups and risk factors vary across populations, we thought it was important to evaluate the associations we found in a multivariate context as well. Indeed, we found significant associations between risk factors and blood groups. However, the significant associations between blood group and COVID-19 status were reflected in significant reductions in deviance when adding blood group to a regression of risk factors on COVID-19 status. Moreover, the blood group regression coefficients were largely unchanged when risk factors were present or not, indicating that blood groups have an independent effect on COVID-19 status not captured by the risk factors. These results suggest that the significant associations we discovered are not explained by confounding due to these risk factors.

Our meta-analysis found large heterogeneity in blood group distributions between Wuhan, Shenzhen, and New York City, consistent with previous work indicating large differences in blood group distributions between the United States and China [13,14]. Overall blood group differences introduced heterogeneity in our meta-analysis comparisons of blood group between COV+ individuals and the general population. However, the increased sample size afforded by adding NYP/CUIMC data allowed the B blood group association to be declared significant at the 5% level. Larger sample sizes of COVID-19 patients will allow afford a more detailed picture of the effects of blood type on COVID-19 susceptibility.

The significant associations we found for blood type between COV+ and COV- individuals were far from significant in a comparison between COV+ and the general population at NYP/CUIMC. A possible explanation for this finding is that individuals tested for infection at NYP/CUIMC represent a more homogeneous sub-population of patients at NYP/CUIMC. Increased homogeneity would strengthen the blood-group-COVID-19 association signal as it would reduce the influence of overall population differences in blood-type distribution. We did not find sufficient evidence to conclude that SARS-CoV-2-tested individuals have a significantly different blood group distribution than the general population at NYP/CUIMC, though we cannot rule out other differences between tested and general populations that could explain the difference in associations. Moreover, our meta-analysis using COV+ vs general population found significant associations between A, B, and O blood groups, and the NYP/CUIMC data received 20-30 percent weight for each comparison, indicating a large contribution to the pooled associations. Further work is needed to understand how the population of COVID-tested patients differs from the general population.

We did not identify any significant relationships between blood group and intubation or death due to COVID-19. However, intubation and death due to COVID-19 continue in New York as of

April 5, 2020, and individuals currently alive, not intubated, or COV- may reach these outcomes in the future. Our data is preliminary and represents a snapshot of the pandemic in a New York hospital system. When more patients become tested, intubated, and recovered, we will be better able to assess the relationship between blood group and eventual COVID-19 outcomes that may not have occurred at the time of our analysis.

Our study analyzed EHR data collected during the care of patients, not necessarily with research intent. Our sample sizes were relatively small, making explicit stratifications by age, sex, comorbidities, and other risk factors challenging. As an observational study without rigorous corrections for additional possible confounding, our results should be considered preliminary and should not be taken to inform clinical practice or policy.

Conclusion

In this study we found evidence for association between blood groups and COVID-19. Using data from NYP/CUIMC, we found the odds of COVID-19 positive vs negative test results were increased in blood groups A and decreased in blood groups O, consistent with previous results from Wuhan and Shenzhen. While Rh negative blood types are rare, we find evidence of association only for Rh positive blood groups. Though few AB individuals were included in our cohort, we discovered a new significant odds decrease for AB blood groups. In a meta-analysis of our data with data from Wuhan and Shenzhen reported by Zhao et al., we found a new significant COVID-19 odds increase for B blood groups compared to the general population. We demonstrated that the associations we found were not explained by confounding due to several known risk factors. Our results replicate previously-discovered associations between A and O blood groups and COVID-19, and we show novel associations between B, AB, and Rh blood groups.

References

1. **Coronavirus disease 2019 (COVID-19) Situation Report – 76**
World Health Organization
WHO Coronavirus disease 2019 situation reports (2020-04-05)
https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200405-sitrep-76-covid-19.pdf?sfvrsn=6ecf0977_2

2. **Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study**
Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, … Bin Cao
The Lancet (2020-03) https://doi.org/ggnxb3
DOI: 10.1016/s0140-6736(20)30566-3

3. **Relationship between the ABO Blood Group and the COVID-19 Susceptibility**
Jiao Zhao, Yan Yang, Hanping Huang, Dong Li, Dongfeng Gu, Xiangfeng Lu, Zheng Zhang, Lei Liu, Ting Liu, Yukun Liu, … Peng George Wang

medRxiv (2020-03-27) https://doi.org/ggpn3d
DOI: 10.1101/2020.03.11.20031096

4. **ABO Blood Group and Susceptibility to Severe Acute Respiratory Syndrome**
JAMA
(2005-03-23) https://doi.org/ftkw6v
DOI: 10.1001/jama.293.12.1450-c · PMID: 15784866

5. **ABO Blood Group Phenotypes and Plasmodium falciparum Malaria: Unlocking a Pivotal Mechanism**
María‐Paz Loscertales, Stephen Owens, James O'Donnell, James Bunn, Xavier Bosch‐Capblanch, Bernard J. Brabin
Advances in Parasitology (2007) https://doi.org/db42m2
DOI: 10.1016/s0065-308x(07)65001-5

6. **Attachment of Helicobacter pylori to human gastric epithelium mediated by blood group antigens**
T Boren, P Falk, K. Roth, G Larson, S Normark
Science (1993-12-17) https://doi.org/d3wbh6
DOI: 10.1126/science.8018146 · PMID: 8018146

7. **Human susceptibility and resistance to Norwalk virus infection**
Lisa Lindesmith, Christine Moe, Severine Marionneau, Nathalie Ruvoen, Xi Jiang, Lauren Lindblad, Paul Stewart, Jacques LePendu, Ralph Baric
Nature Medicine (2003-04-14) https://doi.org/c7xvwh
DOI: 10.1038/nm860 · PMID: 12692541

8. **ABO blood group, hepatitis B viral infection and risk of pancreatic cancer**
De-Shen Wang, Dong-Liang Chen, Chao Ren, Zhi-Qiang Wang, Miao-Zhen Qiu, Hui-Yan Luo, Dong-Sheng Zhang, Feng-Hua Wang, Yu-Hong Li, Rui-Hua Xu
International Journal of Cancer (2012-07-15) https://doi.org/cpfxbm
DOI: 10.1002/ijc.26376 · PMID: 21858814

9. **Relation of Infection with Neisseria gonorrhoeae to ABO Blood Groups**
M. T. Foster, A. H. Labrum
Journal of Infectious Diseases (1976-02-01) https://doi.org/dt9x9d
DOI: 10.1093/infdis/133.3.329 · PMID: 1254989

10. **COVID-19: Data - NYC Health**https://www1.nyc.gov/site/doh/covid/covid-19-data.page

11. **Meta-Analysis with R**
Guido Schwarzer, James R. Carpenter, Gerta Rücker
Use R! (2015) https://doi.org/drbb
DOI: 10.1007/978-3-319-21416-0

12. **Open collaborative writing with Manubot**
Daniel S. Himmelstein, Vincent Rubinetti, David R. Slochower, Dongbo Hu, Venkat S. Malladi,

Casey S. Greene, Anthony Gitter
PLOS Computational Biology (2019-06-24) https://doi.org/c7np
DOI: 10.1371/journal.pcbi.1007128 · PMID: 31233491 · PMCID: PMC6611653

13. **Frequencies and ethnic distribution of ABO and RhD blood groups in China: a population-based cross-sectional study**
Jue Liu, Shikun Zhang, Qiaomei Wang, Haiping Shen, Yiping Zhang, Min Liu
BMJ Open (2017-12-03) https://doi.org/gcnrk5
DOI: 10.1136/bmjopen-2017-018476 · PMID: 29203504 · PMCID: PMC5736034

14. **ABO and Rh(D) phenotype frequencies of different racial/ ethnic groups in the United States**
George Garratty, Simone A. Glynn, Robin McEntire, Retrovirus Epidemiology Donor Study
Transfusion (2004-05) https://doi.org/dkshr5
DOI: 10.1111/j.1537-2995.2004.03338.x · PMID: 15104651

Supplemental information

Table 4: Procedures used by name to identify individual blood group

Procedure name
TYPE AND SCREEN
BLOOD TYPE ABO AND RH
TYPE (ABO CONFIRMATION ONLY)
NEWBORN PANEL (ABO/RH PLUS DAT PLUS AB SCREEN)
CORD BLOOD PANEL (ABO/RH PLUS DAT)
NEWBORN BLOOD TYPE

Table 5: Codes used to define phenotypes. For each code, we used the code and all descendants of the code to define the phenotype, and assigned individuals based on the presence or absence of any code belonging to the phenotype assigned them. Concept IDs are based on OMOP CDM concept IDs. N individuals is the number of individuals from our analyzed cohort who had a single code included in the phenotype definition.

Risk factor	Concept ID	N Individuals
Hypertension	19829001	724
Cardiovascular diseases	134057	878
Respiratory diseases	320136	848
Diabetes mellitus	201820	445
Overweight status	437525	188

Table 6: Summary of chi-squared tests for association between blood type and COVID-19 outcomes. 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, while ABO+Rh used a 6x2 table for each test, resulting in 3 and 5 degrees of freedom, respectively.

Comparison groups	Blood group type	Group 1 counts	Group 2 counts	p-value
COV+ vs COV-	ABO	A:233 (34.2%), AB:21 (3.1%), B:116 (17%), O:312 (45.7%)	A:245 (27.9%), AB:47 (5.4%), B:136 (15.5%), O:449 (51.2%)	**0.006**
COV+ vs COV-	ABO+Rh	A-:17 (2.5%), A+:216 (31.7%), AB+:21 (3.1%), B-:7 (1%), B+:109 (16%), O-:21 (3.1%), O+:291 (42.7%)	A-:26 (3%), A+:219 (25.1%), AB+:42 (4.8%), B-:14 (1.6%), B+:122 (14%), O-:26 (3%), O+:423 (48.5%)	**0.031**
COV+ vs general population	ABO	A:233 (34.2%), AB:21 (3.1%), B:116 (17%), O:312 (45.7%)	A:35643 (32.7%), AB:4582 (4.2%), B:16229 (14.9%), O:52406 (48.1%)	0.152
COV+ vs general population	ABO+Rh	A-:17 (2.5%), A+:216 (31.7%), AB+:21 (3.1%), B-:7 (1%), B+:109 (16%), O-:21 (3.1%), O+:291 (42.7%)	A-:3447 (3.2%), A+:32196 (29.7%), AB+:4188 (3.9%), B-:1422 (1.3%), B+:14807 (13.7%), O-:4808 (4.4%), O+:47598 (43.9%)	0.166
COV+/Died vs COV+/Alive	ABO	A:27 (33.8%), AB:5 (6.2%), B:12 (15%), O:36 (45%)	A:206 (34.2%), AB:16 (2.7%), B:104 (17.3%), O:276 (45.8%)	0.363
COV+/Died vs COV+/Alive	ABO+Rh	A-:1 (1.2%), A+:26 (32.5%), AB+:5 (6.2%), B-:0 (0%), B+:12 (15%), O-:0 (0%), O+:36 (45%)	A-:16 (2.7%), A+:190 (31.6%), AB+:16 (2.7%), B-:7 (1.2%), B+:97 (16.1%), O-:21 (3.5%), O+:255 (42.4%)	0.283
COV+/Intubated vs COV+/Not intubated	ABO	A:62 (34.6%), AB:8 (4.5%), B:35 (19.6%), O:74 (41.3%)	A:171 (34%), AB:13 (2.6%), B:81 (16.1%), O:238 (47.3%)	0.322
COV+/Intubated vs COV+/Not intubated	ABO+Rh	A-:2 (1.1%), A+:60 (33.5%), AB+:8 (4.5%), B-:2 (1.1%), B+:33 (18.4%), O-:4 (2.2%), O+:70 (39.1%)	A-:15 (3%), A+:156 (31%), AB+:13 (2.6%), B-:5 (1%), B+:76 (15.1%), O-:17 (3.4%), O+:221 (43.9%)	0.441

Table 7: Summary of all one-vs-rest analyses conducted. Each individual test compared the listed blood group with all other blood groups between the listed comparison groups. Shown are comparisons between each blood type and all three COVID-19 outcomes investigated.

Blood group	Blood group type	Comparison groups	OR	95% CI	p-value
A	ABO	COV+ vs general population	1.066	0.906 - 1.252	0.437
A-negative	ABO+Rh	COV+ vs general population	0.779	0.45 - 1.258	0.379
A-positive	ABO+Rh	COV+ vs general population	1.098	0.93 - 1.294	0.257
AB	ABO	COV+ vs general population	0.723	0.444 - 1.116	0.151
AB-positive	ABO+Rh	COV+ vs general population	0.791	0.486 - 1.221	0.368
B	ABO	COV+ vs general population	1.170	0.949 - 1.432	0.131
B-negative	ABO+Rh	COV+ vs general population	0.781	0.312 - 1.622	0.733
B-positive	ABO+Rh	COV+ vs general population	1.203	0.971 - 1.48	0.083
O	ABO	COV+ vs general population	0.908	0.778 - 1.059	0.219
O-negative	ABO+Rh	COV+ vs general population	0.685	0.421 - 1.057	0.092
O-positive	ABO+Rh	COV+ vs general population	0.952	0.815 - 1.111	0.536
A	ABO	COV+ vs COV-	1.338	1.072 - 1.672	**0.009**
A-negative	ABO+Rh	COV+ vs COV-	0.832	0.42 - 1.608	0.641
A-positive	ABO+Rh	COV+ vs COV-	1.382	1.099 - 1.737	**0.004**
AB	ABO	COV+ vs COV-	0.561	0.315 - 0.969	**0.033**
AB-positive	ABO+Rh	COV+ vs COV-	0.628	0.35 - 1.097	0.093
B	ABO	COV+ vs COV-	1.117	0.843 - 1.477	0.446
B-negative	ABO+Rh	COV+ vs COV-	0.636	0.216 - 1.695	0.381
B-positive	ABO+Rh	COV+ vs COV-	1.169	0.874 - 1.563	0.282
O	ABO	COV+ vs COV-	0.804	0.654 - 0.987	**0.036**
O-negative	ABO+Rh	COV+ vs COV-	1.034	0.548 - 1.93	1.000
O-positive	ABO+Rh	COV+ vs COV-	0.790	0.642 - 0.971	**0.024**
A	ABO	COV+/Intubated vs COV+/Not intubated	1.029	0.705 - 1.493	0.927
A-negative	ABO+Rh	COV+/Intubated vs COV+/Not intubated	0.368	0.04 - 1.608	0.263
A-positive	ABO+Rh	COV+/Intubated vs COV+/Not intubated	1.121	0.765 - 1.635	0.575
AB	ABO	COV+/Intubated vs COV+/Not intubated	1.762	0.622 - 4.678	0.214
AB-positive	ABO+Rh	COV+/Intubated vs COV+/Not intubated	1.762	0.622 - 4.678	0.214

B	ABO	COV+/Intubated vs COV+/Not intubated	1.266	0.79 - 1.999	0.298
B-negative	ABO+Rh	COV+/Intubated vs COV+/Not intubated	1.125	0.106 - 6.948	1.000
B-positive	ABO+Rh	COV+/Intubated vs COV+/Not intubated	1.269	0.783 - 2.027	0.342
O	ABO	COV+/Intubated vs COV+/Not intubated	0.785	0.547 - 1.124	0.190
O-negative	ABO+Rh	COV+/Intubated vs COV+/Not intubated	0.654	0.158 - 2.042	0.616
O-positive	ABO+Rh	COV+/Intubated vs COV+/Not intubated	0.820	0.569 - 1.177	0.291
A	ABO	COV+/Died vs COV+/Alive	0.979	0.574 - 1.639	1.000
A-negative	ABO+Rh	COV+/Died vs COV+/Alive	0.464	0.011 - 3.067	0.708
A-positive	ABO+Rh	COV+/Died vs COV+/Alive	1.044	0.608 - 1.757	0.898
AB	ABO	COV+/Died vs COV+/Alive	2.437	0.679 - 7.226	0.088
AB-positive	ABO+Rh	COV+/Died vs COV+/Alive	2.437	0.679 - 7.226	0.088
B	ABO	COV+/Died vs COV+/Alive	0.845	0.402 - 1.646	0.751
B-negative	ABO+Rh	COV+/Died vs COV+/Alive	0.000	0 - 5.26	1.000
B-positive	ABO+Rh	COV+/Died vs COV+/Alive	0.919	0.436 - 1.794	0.872
O	ABO	COV+/Died vs COV+/Alive	0.966	0.586 - 1.585	0.906
O-negative	ABO+Rh	COV+/Died vs COV+/Alive	0.000	0 - 1.429	0.158
O-positive	ABO+Rh	COV+/Died vs COV+/Alive	1.113	0.675 - 1.827	0.718

Table 8: All associations between risk factors and blood groups where logistic regression coefficient p-values were below 0.1. These results are from logistic regression of blood group ~ risk factors. Full data on all coefficients are available on GitHub.

Blood group	Term	Coefficient	Standard error	p-value
A	age	0.008	0.003	**0.005**
AB	age	0.014	0.006	**0.027**
B	diabetes	-0.434	0.195	**0.026**
O	age	-0.008	0.003	**0.003**
O	diabetes	0.248	0.142	0.080
A_neg	hypertension	-0.895	0.511	0.080
A_neg	diabetes	0.880	0.442	**0.047**
A_neg	CV diseases	0.852	0.507	0.093
A_pos	age	0.009	0.003	**0.001**
AB_pos	age	0.013	0.006	**0.043**

B_neg	age	0.018	0.011	0.096
B_pos	diabetes	-0.390	0.200	0.052
O_neg	hypertension	2.124	1.043	**0.042**
O_neg	overweight	-1.359	0.744	0.068
O_pos	age	-0.007	0.003	**0.006**
O_pos	diabetes	0.279	0.143	0.051
O_pos	overweight	0.355	0.168	**0.034**

Table 9: Analysis of deviance for comparisons between null (intercept only), risk factors (RF), and blood groups (ABO and ABO+Rh) on COVID-19 outcomes. 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	Compared models	df	Resid. df	Deviance	Resid. Deviance	p-value
COV+ vs COV-	Risk factors vs Null	7	1550	121.851	2013.777	0.000
COV+ vs COV-	ABO + Risk factors vs Risk factors	3	1547	10.752	2003.024	0.013
COV+ vs COV-	ABO+Rh + Risk factors vs Risk factors	7	1543	17.165	1996.612	0.016
COV+/Intubated vs COV+/Not intubated	Risk factors vs Null	7	674	12.250	772.893	0.093
COV+/Intubated vs COV+/Not intubated	ABO + Risk factors vs Risk factors	3	671	3.021	769.872	0.388
COV+/Intubated vs COV+/Not intubated	ABO+Rh + Risk factors vs Risk factors	6	668	5.876	767.017	0.437
COV+/Died vs COV+/Alive	Risk factors vs Null	7	674	100.012	393.094	0.000
COV+/Died vs COV+/Alive	ABO + Risk factors vs Risk factors	3	671	1.153	391.941	0.764
COV+/Died vs COV+/Alive	ABO+Rh + Risk factors vs Risk factors	6	668	5.641	387.453	0.465

Table 10: Logistic regression coefficients and coefficient p-values for comparisons between nested (outcome vs blood group) and full (outcome vs blood group + risk factors) models. The outcome here is COV+ vs COV-. The coefficients are either changed marginally or more extreme in the full model than the nested model. Were COV+ status conditionally independent of blood type given risk factors, we would expect full model coefficients to be less extreme than in

the nested model. AB blood groups were not included because they are mutually exclusive with A, B, and O blood groups. Full data, including ABO+Rh results, are available on GitHub.

Blood group	Nested model coefficient	Full model coefficient	Nested model p-value	Full model p-value
A	0.759	0.752	0.006	0.009
B	0.647	0.757	0.026	0.012
O	0.442	0.504	0.105	0.074

Table 11: Weights for sites in random-effects meta-analyses conducted for each ABO blood group. Each blood group was compared against all others using data from NYP/CUIMC, and Zhao et al. (Wuhan Jinyintan, Renmin Hospital in Wuhan, and Shenzhen Third People's Hospital).

Blood group	Site	OR	95% CI	%Weight
A	NYP/CUIMC	1.0660	0.9095 - 1.2494	31.8
A	Wuhan Jinyintan	1.2790	1.1364 - 1.4395	39.3
A	Wuhan Renmin	1.3959	0.9519 - 2.0472	10.3
A	Shenzhen	1.0001	0.7727 - 1.2945	18.6
B	NYP/CUIMC	1.1698	0.9573 - 1.4294	23.7
B	Wuhan Jinyintan	1.0828	0.9516 - 1.2321	57.1
B	Wuhan Renmin	0.8566	0.5460 - 1.3440	4.7
B	Shenzhen	1.2233	0.9458 - 1.5822	14.4
AB	NYP/CUIMC	0.7230	0.4678 - 1.1176	23.5
AB	Wuhan Jinyintan	1.1139	0.9201 - 1.3487	30.2
AB	Wuhan Renmin	1.5297	0.8783 - 2.6643	20.0
AB	Shenzhen	2.0071	1.4266 - 2.8237	26.3
O	NYP/CUIMC	0.9084	0.7810 - 1.0566	31.1
O	Wuhan Jinyintan	0.6799	0.5993 - 0.7715	32.9
O	Wuhan Renmin	0.6441	0.4179 - 0.9925	13.2
O	Shenzhen	0.6272	0.4842 - 0.8124	22.8

Table 12: Heterogeneity across meta-analysis sites. I-squared is a measure of the variation due to heterogeneity, instead of chance. Q refers to Cochran's Q and has a chi-squared distribution with Q df degrees of freedom.

Blood group	I-squared	I-squared 95% CI	Q	Q df
A	47.1%	0.0 - 82.4	5.67	3
B	0.0%	0.0 - 79.4	2.23	3

AB	80.6%	48.9 - 92.6	15.43	3
O	72.1%	21.0 - 90.2	10.76	3