Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Transfusion clinique et biologique 30 (2023) 116–122

Available online at ScienceDirect www.sciencedirect.com

Elsevier Masson France EM|consulte www.em-consulte.com/en

Original article

Blood-Type-A is a COVID-19 infection and hospitalization risk in a Turkish cohort

Meryem Cetin [a], Sirin Cetin [b], Ayse Ulgen [c,d,*], Wentian Li [e,*]

[a] Department of Medical Microbiology, Faculty of Medicine, Amasya University, Amasya, Turkey
[b] Department of Biostatistics, Amasya University, Amasya, Turkey
[c] Department of Biostatistics, Faculty of Medicine, Girne American University, 99320 Karmi, Cyprus
[d] Department of Mathematics, School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NF, UK
[e] The Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, USA

ARTICLE INFO

Article history:
Available online 13 October 2022

Keywords:
ABO blood group
Anti-A antibody
COVID-19
Logistic regression

ABSTRACT

We have shown in an ethnically homogenous Turkey cohort with more than six thousand cases and 25 thousand controls that ABO blood types that contain anti-A antibody (O and B) are protective against COVID-19 infection and hospitalization, whereas those without the anti-A antibody (A and AB) are risks. The A + AB frequency increases from 54.7 % in uninfected controls to 57.6 % in COVID-19 outpatients, and to 62.5 % in COVID-19 inpatients. The odds-ratio (OR) for lacking of anti-A antibody risk for infection is 1.16 (95 % confidence interval (CI) 1.1–1.22, and Fisher test p-value 1.8×10^{-7}). The OR for hospitalization is 1.23 (95 %CI 1.06–1.42, Fisher test p-value 0.005). A linear regression treating controls, outpatients, inpatients as three numerical levels over anti-A antibody leads to a p-value of 5.9×10^{-9}. All these associations remain to be statistically significant after conditioning over age, even though age itself is a risk for both infection and hospitalization. We also attempted to correct the potential effect from vaccination, even though vaccination information is not available, by using the date of the data collection as a surrogate to vaccination status. Although no significant association between infection/hospitalization with Rhesus blood system was found, forest plots are used to illustrate possible trends.

Introduction

The ABO blood types have been shown to be associated with COVID-19 infection susceptibility, severity, and mortality [1–6]. Further genetic studies show a direct link between COVID-19 severity and genetic variants at the ABO locus on chromosome 9 [7]. The same locus is also identified in an expression quantitative trait loci (eQTL) analysis in multiple tissues [8]. Although ABO blood type is mostly used as a biomarker, or to be matched in blood transfusion, biomolecules associated with the ABO type do have biological functions. In particular, ABO blood type is linked to hemostasis and thrombosis [9–11], with non-O blood type associated with thromboembolism risk [12]. Also, if the ABO blood type specific antibody (anti-histo-blood group antibodies, e.g. anti-A, anti-B) [13] protects the binding between coronavirus spike protein and cell receptor [14], there could be differences of immune responses to the SARS-Cov-2 virus between people with different blood types [15–17].

Because the ABO type frequencies differ in different population and ethnic groups, association results from one study in one region need to be validated in other ethnic groups. The ABO frequency pattern doesn't have a simple correlation with continent [18]. More than half of the countries in the world have O% > A% > B% > AB% (including, e.g., UK, Spain); about 30 % countries have A % > O% B% > AB% (e.g., Norway, Sweden); and more than 10 % of countries have O% > B% > A% > AB% (e.g. Thailand, Vietnam). Turkey belongs to the A% > O% > B% > AB% group, thus it has a higher A% than other blood types. Multiple neighboring countries in central Asia, Middle East, and Eastern Europe share a comparable ABO blood type distributions.

Previous studies show blood type-A is associated with increased risk of infection, and type-O with decreased risk [19–23], type-A and AB is associated with increased risk of requiring ventilation

Abbreviations: ICU, intensive care unit; COVID-19, coronavirus disease 2019.

* Corresponding author at: Department of Biostatistics, Faculty of Medicine, Girne American University, 99320 Karmi, Cyprus.
E-mail addresses: ayshe.ulgen@global.t-bird.edu (A. Ulgen), wtli2012@gmail.com (W. Li).

https://doi.org/10.1016/j.tracli.2022.10.003
1246-7820/

[24]), but a New York study showed a slight decrease of intubation risk for type-A COVID-19 patients [25],lower risk for death for type-O [21] and lower risk for Rh(D)-negative type [21]. On the other hand, there are reports on lack of evidence for association between ABO blood types with COVID-19 infection, severity, and outcomes [26,27].

For Turkish cohorts, it is reported in [28] that type-A is more frequent in infected COVID-19 patients (A% = 57 %, sample size n = 186) than healthy non-patients (A% = 38 %, n = 1881), and type-O is less frequent in COVID-19 patients (O% = 24.8 %) than general population (37.2 %). In another study [29], type-A is linked to a higher ICU admission rate (p-value = 0.027), while type-O is not significant (total number of patients is n = 39850). In the publication [30], however, type-A is more common in asymptomatic patients (n = 56) than symptomatic ones (n = 76). In a nationwide collection of convalescent plasma donors (who were infected with COVID-19) and blood donation samples (which are treated as population level baseline) from Turkish Red Crescent, it is observed that type-A, type-AB, and Rh(D)+ were over-represented in infected COVID-19 patients than general population [31].

All previous studies collectively present a partial and tentative picture of the impact of ABO blood type on COVID-19 infection, severity and outcome. However, different studies are still not completely consistent. Studies with small sample sizes are unlikely to produce a significant evidence, whereas larger sample size collected from a heterogenous population may mix the ethnic groups with different ABO blood type distributions, leading to spurious results [32]. Using control samples from the same ethnic group and same geographic region as the comparison group is of crucial important in a study design [33,34]). It is therefore our intention to add yet another cohort with large enough samples (several thousands COVID-19 patients), from the homogenous population (city of Amasya, Turkey), with well defined infection status and disease severity (inpatients being more severe, and outpatients being less severe), to help to clarify the association between ABO blood type and COVID-19 disease.

Knowing that COVID-19 disease is not impacted by ABO blood type alone and the severity could be influenced by many other factors [35–39], we analyzed our data a step further than most other studies, by conditional on age and gender-- known risk factors for COVID-19 disease severity, whenever possible, in the examination of ABO-severity or ABO-infection association. Our results confirm that type-A increases the infection and inpatient risk, whereas type-O and type-B decrease it. Lastly, the issue of vaccination was addressed even when we do not have the vaccination status data, by logistic regression conditional on the month a sample's blood was collected. Our conclusion concerning infection risk is not changed by this conditioning, but the conclusion on hospitalization would require more samples (more than the 9 hundred inpatients in our collection) to confirm.

Material and methods

We have surveyed 6553 COVID-19 patients from Amasya State Hospital, Amasya, Turkey, with 5628 outpatients and 925 inpatients. Most patients had mild symptoms in the first week, such as fever, muscle or joint pains, cough, and sore throat without distress (respiration rate per minute less than 24, oxygen saturation (SpO2) level larger than 95 %). In this period, patients were monitored either at home or at hospital, a decision made by attending physician on a case-by-case basis.

In the second week, the chance to developing severe illness is higher. Severe symptoms include shortness of breath, continuation of fever, low blood oxygen level, etc. All monitored patients with an onset of severe symptoms were admitted to hospital as inpatients. The admission date was from April 2020 to August 2021.

The ABO and Rh blood system type is determined for all samples. Other information, such as gender and age, are also collected. Although inpatients who decides to be monitored in a hospital setting may not have severe symptoms, this consists a very small proportion of all inpatients.

The general population is represented by a group of controls with medical records in Amasya State Hospital, Amasya, Turkey, collected from January 2020 to November 2021. There are 25,163 persons in this collection. Most of the controls are non-COVID-19 patients in the hospital, while roughly 15 % were babies newly born at the hospital.

All analysis and tests are carried out by R statistical packages (https://www.r-project.org). In particular, *glm* function (generalized linear model) is used for carrying out logistic regression, where the family is chosen to be "binomial". The R notation *glm* ($y \sim x$, *data* = . . ., *family*="*binomial*") (y is the binary outcome variable, x is an independent variable) implies the following logistic regression (for more applications of logistic regression, see, e.g., [40]):

$$\mathrm{Pro}b(y=1)=\frac{1}{1+e^{-c-a\cdot x}}$$

The p-value for variable x is that for testing coefficient a to be zero. If $p < p_0$ where p_0 is a small value (e.g. $p_0 = 0.01$), variable x is significantly associated with y at level-p_0 (the practice of claiming statistical significance without mentioning the level p_0 is strongly discouraged [41,42]. Similarly, logistic regression can be carried out conditional on another co-variate z:

$$\mathrm{Pro}b(y=1)=\frac{1}{1+e^{-c-a\cdot x-b\cdot z}}$$

The number of covariates can be more than one.

Besides *glm*, *forestplot* from the forestplot package is to show the forest plot [43], *lm* is used for linear regression, *fisher.test* function is used to carry out Fisher's test on 2-by-2 count tables. If a count table has more than 2 rows or 2 columns, *chisq.test* function is used to carry out a χ^2 test. For 2-by-2 count tables, χ^2 test and Fisher's test lead to similar p-values.

Results

Type-A is a risk for COVID-19 infection and type-O is a protection: Table 1 compares the ABO blood type distribution in the COVID-19 patients (both outpatients and inpatients, n = 6553) and other hospital patients who are not infected with SARS-Cov-2 varus (n = 25163). The type-A is overrepresented in COVID-19 patients (49.5 %) than non-COVID patients (46.8 %), whereas type-O is underrepresented in COVID-19 patients (26.9 %) vs non– COVID patients (30.4 %). The type-B frequency is only slightly higher in non-COVID patients, and type-AB is slightly higher in COVID-19 patients. The ABO distribution between infected and uninfected groups is significantly different (p-value = 1.8×10^{-7}, from 4-by-2 count table).

For each individual ABO type, a 2-by-2 table is used for counts belong to, and not belong to, that type. The Fisher's test can be carried out. The p-values for type-A, O, B, AB from the Fisher's test are 9.5×10^{-5}, 3.3×10^{-8}, 0.86, and 0.022. In other words, the statistical evidence is stronger for type-A and type-O. The odds-ratios (OR) for type-A, O, B, AB are 1.115 (risk), 0.843 (protection), 0.992 (not significant), 1.121 (risk). These odds-ratios in the range of 1.1–1.2 indicate that the signal strength is not particularly strong.

By the hypothesis that anti-A antibody may play a role in preventing virial enter of cell, blood type-O and B both have anti-A

Table 1
Number of persons infected with COVID-19 (n(COVID) column) and without COVID-19 (n(non)) stratified by the ABO blood type (rows), as well as Rh blood system (last two blocks). Blood types can also be grouped as those with anti-A antibody (O and B) and those without (A and AB), shown in the last two rows. The distribution (frequencies) of these blood types are shown in parenthesis. The Fisher's test *p*-value refers to that of the 2-by-2 count table with a particular blood type and without, in COVID-19 patients and in uninfected controls. The odds-ratio (OR) refer to the particular blood type in favor (risk) for infection.

type	n(COVID), n(non)	Fisher pv	OR	Rh(D)+: n(COVID), n(non)	Rh(D)-: n(COVID), n(non)
A	3244 (49.5 %), 11775 (46.8 %)	9.5E-5	1.115	2889 (49.5 %), 10486 (46.8 %)	362 (49.7 %), 1289 (46.8 %)
O	1766 (26.9 %), 7658 (30.4 %)	3.3E-8	0.843	1587 (27.2 %), 6836 (30.5 %)	181 (24.9 %), 822 (29.9 %)
B	968 (14.77 %), 3742 (14.87 %)	0.86	0.992	865 (14.8 %), 3318 (14.8 %)	105 (14.4 %), 424 (15.4 %)
AB	575 (8.77 %),1988 (7.9 %)	0.022	1.121	495 (8.5 %), 1771 (7.9 %)	80 (10.99 %), 217 (7.88 %)
A + AB	3819 (58.3 %), 13763 (54.7 %)	1.9E-7	1.157	3377 (57.97 %), 12257 (54.69 %)	442 (60.7 %), 1506 (54.7 %)
O + B	2734 (41.7 %), 11400 (45.3 %)	same	0.864	2448 (42.0 %), 10154 (45.3 %)	286 (39.3 %), 1246 (45.3 %)

antibody where type-A and AB do not. Therefore, type-O and B are combined in one group and A and AB in another, forming a 2-by-2 count table (Table 1). The Fisher's test *p*-value is 1.9×10^{-7}, and odds-ratio 1.157 (95 %CI: (1.095–1.222)). All results in Table 1 are also graphically illustrated by a forest plot, where the ORs and 95 % confidence intervals of OR in different situations are shown.

The Rhesus blood system does not seem to affect the result: Table 1 shows the Rh(D)+ and Rh(D)- specific counts and distribution of ABO blood types and the same pattern can be seen. Because most results concerning Rhesus blood system are not significant, these results are only presented in a forest plot in Fig. 1, at least to show possible trends. In Fig. 1, all 95 % confidence intervals of OR related to Rhesus system bracket the OR = 1 vertical line (the last 7 rows). The trend from our data (though not statistically significant) is that Rh(D)+ is protective for all samples and for A + AB subgroup, but is a risk for O + B subgroup.

Type-A is a risk for COVID-19 hospitalization and type-B is a protection: Table 2 shows the ABO blood type distribution in COVID-19 inpatients (n = 925) and outpatients (n = 5628). The type-A is overrepresented in inpatients (54.4 %) compared to outpatients (48.7 %). The type-B is overrepresented in outpatients (15.2 %) versus inpatients (12.2 %). There are also trends for type-O and AB to be slightly more frequent in outpatients than inpatients. The χ^2 test *p*-value for the 4-by-2 count table (4 blood types, inpatient and outpatient) is 0.009. This *p*-value is not as small as that in Table 1 mainly due to a smaller sample size in Table 2 than in Table 1.

For individual ABO type, the same comparison was carried out between count of a type and the rest, in 2-by-2 count tables. The Fisher's *p*-values for type-A, O, B, AB are 0.0014, 0.23, 0.019, 0.49, and odds-ratio are 1.255 (risk), 0.905 (not significant), 0.777 (protection), 0.905 (not significant). For type-A and type-B inpatient-outpatient comparisons, the test result is statistically significant

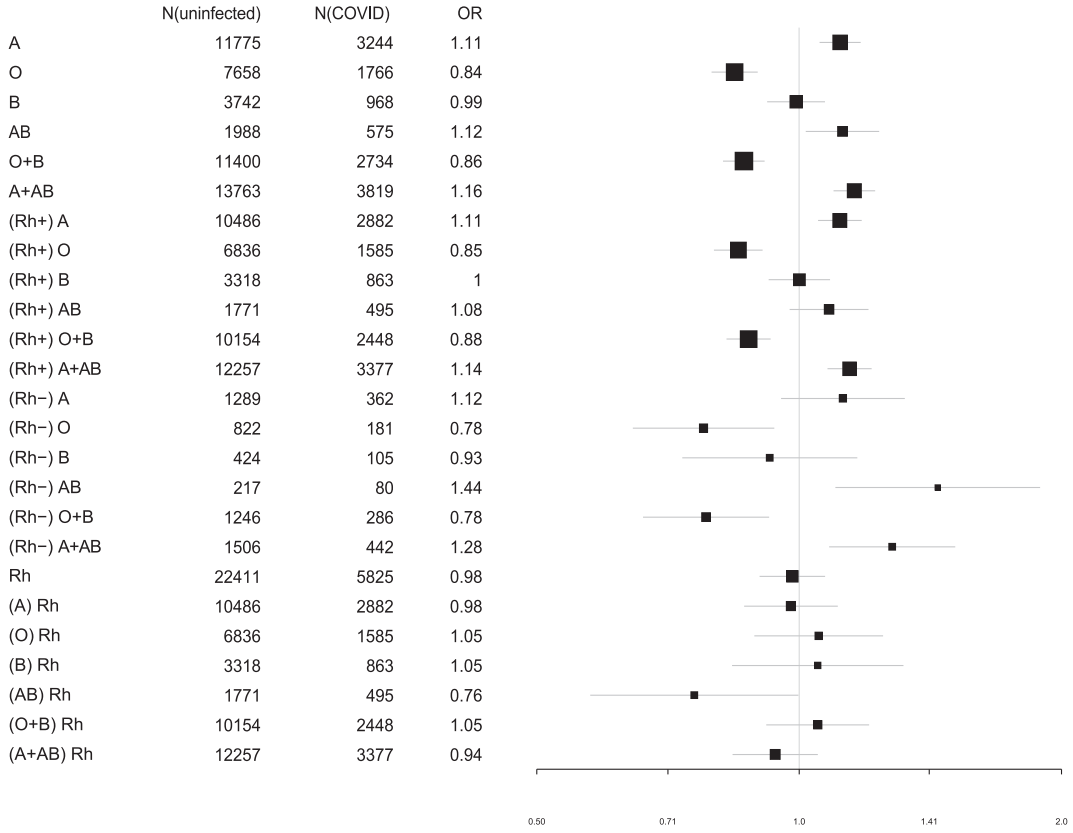

Fig. 1. Forest plot for infection in various risk groups. The plot shows the odds-ratio (in log scale) with a condition (e.g., A type in the first row), and its 95 % confidence interval. The size of the square indicates the sample size. The four columns are: the stratified group and the risk variable, number of uninfected samples in the group, number of COVID-19 infected samples in the group, and odds-ratio (larger than 1 if the positive risk value increases the infection rate). Examples: (Rh(D)+)A means for Rh(D)+ samples only and consider A as the risk variable; (A) Rh means for type A samples to consider Rh as the risk variable; etc.

Table 2
Similar to Table 1, the number of COVID-19 patients who are hospitalized (n(in)) and those who are not (n(out)) are shown, stratified by the ABO blood type, as well as Rh blood system.

type	n(in), n(out)	Fisher pv	OR	Rh(D)+: n(in), n(out)	Rh(D)-: n(in), n(out)
A	503(54.4%), 2741(48.7%)	0.0014	1.255	442(54.2%), 2440(48.7%)	61(55.5%), 301 (48.7%)
O	234 (25.3%), 1532(27.2%)	0.23	0.905	213(26.1%), 1372(27.4%)	21 (19.1%), 160 (25.9%)
B	113(12.2%), 855(15.2%)	0.019	0.777	99(12.1%), 764(15.2%)	14(12.7%), 91 (14.7%)
AB	75(8.1%), 500(8.9%)	0.49	0.905	61 (7.5%), 434(8.7%)	14(12.7%), 66 (10.7%)
A + AB	578 (62.5%), 3241 (57.6%)	0.005	1.227	503 (61.7%), 2874 (57.4%)	75(68.2%), 367(59.4%)
O + B	347 (37.5%), 2387 (42.4%)	same	0.816	312(38.3%), 2136(42.6%)	35(31.8%), 251(40.6%)

(at 0.02 level). Again, by combining type-A and AB, type-O and B, the 4-by-2 count table collapses to a 2-by-2 count table. The Fisher's *p*-value for this table is 0.005 and OR = 1.227 favoring A+AB in inpatient group (95 %CI: (1.063–1.416)). All results in Table 2 are graphically shown in a forest plot in Fig. 2.

Similar to Table 1 and Fig. 1, there are not enough samples to conclude on the impact from Rhesus system on hospitalization. However, similar to Fig. 1, the Rh(D) + is trending a risk (though not significant) in O + B subgroup for hospitalization, and trending a protection, again not significant, in A + AB subgroup.

Linear regression covering both infection and hospitalization: Interestingly, if Table 1 and Table 2 are compared, there is a trend of increasing type-A+AB frequency from non-COVID patient (54.7 %) to COVID-outpatient (57.6 %) to COVID-inpatient (62.5 %). Similarly, the O + B frequency increases from COVID-inpatient (37.5 %) to COVID- outpatient (42.4 %) to non-COVID patient (45.3 %). This indicates that the risk (protection) conveyed by lacking (having) anti-A antibody is reflected both in COVID-19 infection and severity.

This led to an idea of running a regression covering three levels: uninfected controls, infected outpatients, and infected inpatients, as compared to the two levels in logistic regression. This might be carried out by the proportional odds model, an example of the ordinal regression. However, the assumption made in a proportional odds model, that odds are the same from level-1 to level-2 as that from level-2 to level-3, will make it equally reasonable to use the linear regression where the three levels are represented by the numerical values 0,1, and 2.

The linear regression of the three-level indicator variable over the anti-A antibody status leads to a p-value of 5.9×10^{-9}. The association between anti-A antibody and the uninfected-unhospita lized-hospitalized status is also significant (p-value = 2.9×10^{-8}). by $\chi^2(df = 2)$ test. The corresponding p-value from the χ^2 test is 2.9×10^{-8}.

Infection risk/protection conditional on age: The next question to address is whether the association between type-A, type-O and infection still holds true when other infection risk is also considered. Since only age information is available, the following logistic regression is used to consider both blood type and age together (using the R notation, where *glm* for generalized linear model, and *family = binomial* indicates the use of logistic regression):

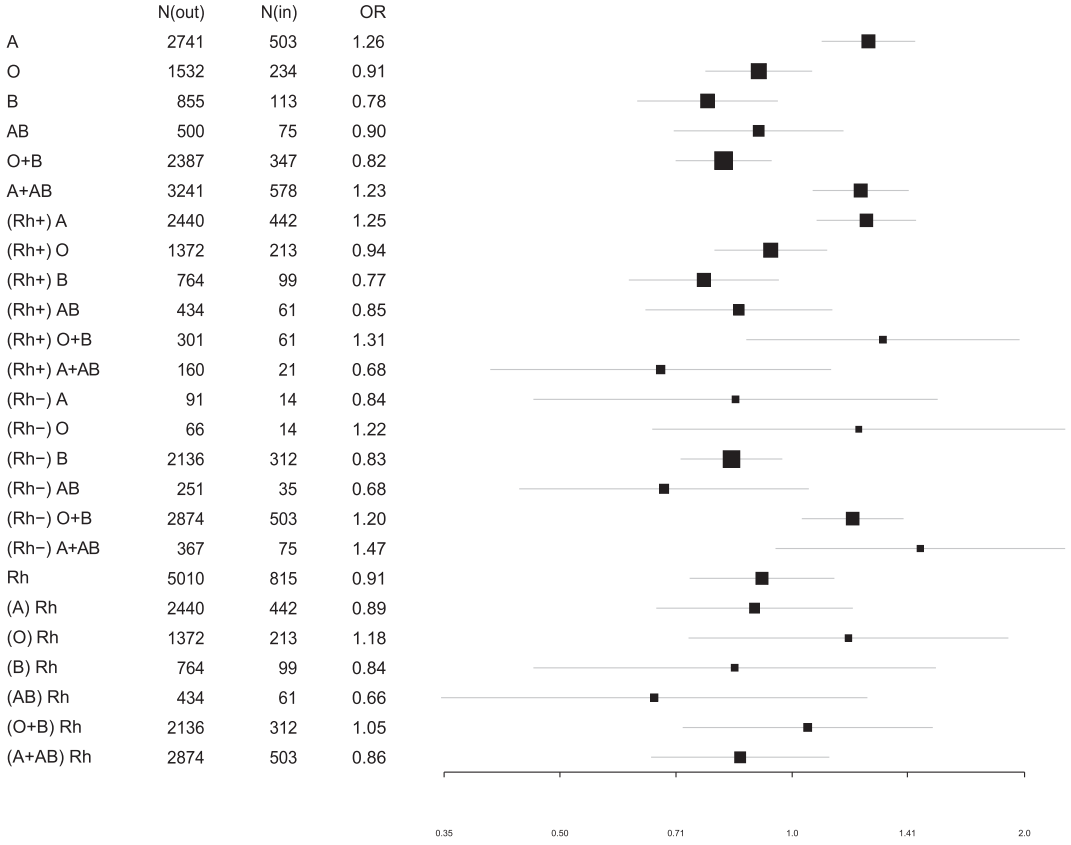

Fig. 2. Similar to Fig. 1, but this forest plot is for hospitalization in various risk groups.

$$glm(COVIDinfection \sim antiAantibody + age, family = "biomial") \quad (1)$$

where variable anti-A antibody takes the value of 1 for type-O and type-B, and 0 for type-A and type-AB.

Both anti-A antibody (p-value = 1.9×10^{-6}) and age (p-value = 2.1×10^{-78}) are significantly associated with infection when conditioning on each other. coefficient) for the anti-A antibody variable is 1.9×10^{-6}, and that for age is 2.1×10^{-78}. The p-value for anti-A antibody variable is comparable to that from Fisher's test in Table 1.

The highly significant result on age is perhaps due to the fact that 15 % of the non-COVID control samples are newborns at the hospital. If only samples with age 25 or up are used, the significance of association with anti-A antibody variable is not changed (p-value = 8.4×10^{-6}), whereas the p-value for age becomes 4.3×10^{-19}. The Rh system variable is not significant in the logistic regression when it is another covariate (result not shown).

Inpatient risk/protection conditional on age: The association between ABO blood type and hospitalization risk can be checked conditional on age and gender by the following logistic regression:

$$glm(hospitalization \sim antiAantibody + age + gender, family = "biomial") \quad (2)$$

The p-value for anti-A antibody without conditioning on any other covariate is 0.043. Compared to the Fisher's test p-value in Table 2, the p-value increases by 10-fold after conditioning on age, but still significant at 0.05 level. We may alter Eq.(2) by conditional on different choices of covariates, such as on Rh only, on age only, etc. The p-value for anti-A antibody term changes only slightly with different these conditionings. These show that anti-A antibody's association with hospitalization is robust. Note that the p-values here are not as small as those for infection test because the sample size is much smaller.

It is straightforward to extend our previous linear regression acrossing three levels of phenotype over anti-A antibody status to a regression with age as a co-variate. Age remains to be significant (p-value = 4.2×10^{-153}). On the other hand, the p-value for anti-A antibody variable is 1.9×10^{-7}, indicating that blood type's contribution to the three level of phenotype is independent from that from the age.

Infection and hospitalization risk/protection conditional on date of measurement as a surrogate for vaccine status: There is an important covariate that contribute to both infection and hospitalization risk: the vaccination status. Unfortunately, information on the vaccination status is not available. However, the proportion of vaccinated individuals is expected to be low because the first vaccination in Turkey only started in early 2021 and only for senior people.

In order to deal with the issue of vaccination status without its information being available, the following assumption is made: the month of the year from the beginning of 2021 could provide partial information on vaccination status. Obviously nobody were vaccinated in 2020, and there are more people in the general population, starting from seniors, being vaccinated with time forward since early 2021. We therefore plot the ABO frequency as a function of the month in which the sample is typed as a function of time (Fig. 3). The black lines represent control samples; pink points are for outpatients and red points for inpatients (when the number of samples in a month is more than 20). We only plot the A and O frequencies for COVID-19 patients because the sample sizes per month for AB and B are low. It can seen from Fig. 3 that A-type frequency in COVID-19 patients were higher, and O-type were lower, than the controls, in particular in the year 2020. In the year 2021, however, these frequencies seem to converge back to those of the controls. A more mathematical analysis using month as a covariate is included in the supplement material.

Discussion

Although we used the term "risk" and "protection" in the Result section, a statistical association is not necessarily a causality until

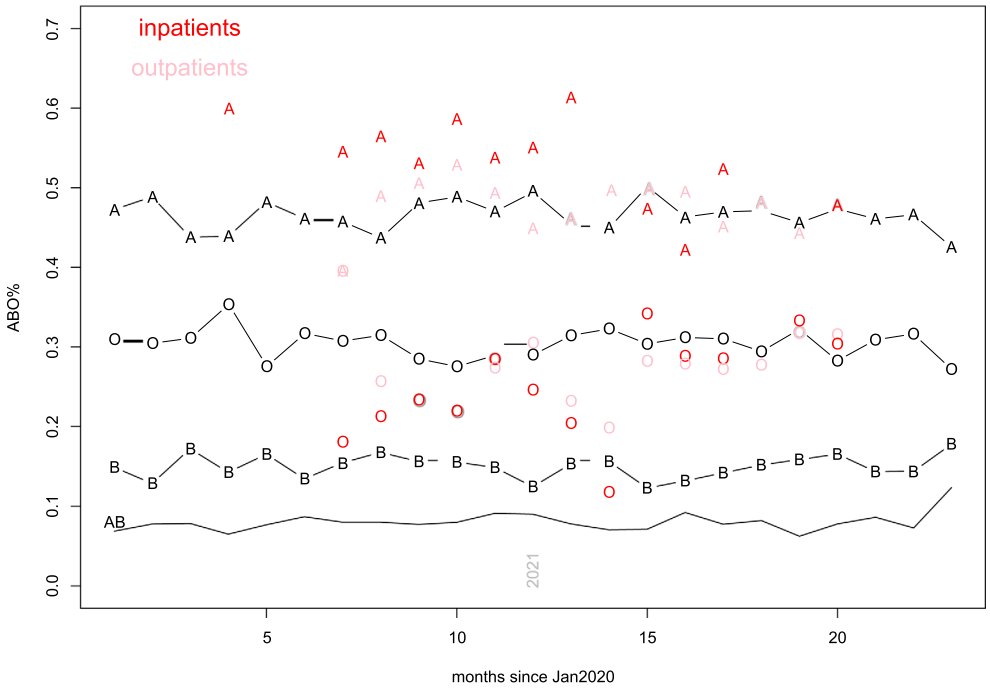

Fig. 3. Frequency of A,B,O,AB blood type for samples registered in a specific month (x-axis the month num- ber since January 2020) (black: uninfected controls, pink: COVID-19 outpatients, red: COVID-19 inpatients). If the number of samples per month in a group is smaller than 20,

proven to be true. Therefore, risk/protection should be understood as a variable status whose increased value is associated with the higher/lower disease incidence and/or prognosis.

In this paper, a trend of increasing more anti-A-antibody-lacking blood types (A and AB) with COVID-19 disease status ladder (54.7 % in unaffected, 57.6 % in out- patients, and 62.5 % in inpatients) is established, which is mostly contributed by the A-type (46.8 % in unaffected, 48.7 % in outpatients, and 54.4 % in inpatients). On the other hand, the trend is opposite for O-type (30.4 % in unaffected, 27.2 % in outpatients, and 25.3 % in inpatients), as well as anti-A-antibody-containing blood types (O and B). Our results are consistent with many other publications, including many studies carried out in Turkey [28,44,45].

However, a few other publications present contradictory conclusions, including some from Turkish studies [30,46,47]. In [30], A-type frequency was 66 % in asymptomatic patients (n = 56), 46.94 % in mild- intermediate patients (n = 51), and 48 % in severe patients (n = 25). It is natural to assume that asymptomatic patients have similar blood type distribution as the general population, but the A frequency is estimated to be 44 % in a much larger sample size (n = 86797) [48], different from the 66 % frequency in [30]. The sample size of n = 56 in [30] might be too small. In [47], O frequency is higher in COVID-19 positive group than negative group. Our recalculation by Fisher's test on the data in [47], however, showed an insignificant conclusion (p-value = 0.11). In [46], O frequency was higher in PCR-positive than PCR-negative (but still had symptom). Not only the result was not statistical significant (at 0.05 level), but also there is a question on if the PCR-positive window had passed if the test was carried out in a late stage. Overall, it is possible that the opposite trend results might be caused by artifacts, in particular small sample sizes.

In [49,50] Rh blood system frequency differences were observed between COVID-19 patients and controls. Although we reproduced the p-value of 0.006 from a binomial distribution in count data in [50] assuming Rh(D) + frequency to be 0.89, we could not reproduce a statistical significant result from the count table in [49] (Fisher's test p-value = 0.69 in our recalculation). We did not find such a difference for Rh(D) in our data.

Efforts were made in our analysis to remove the contributions from confounding variables, if the information was available, such as age. We do not have information on comorbidity for this dataset. Among the few previous work, [51] did not find an association between ABO blood type and obesity in a Saudi Arabian dataset; [52] found blood type B (followed by O) to be a risk for both hypertension and obesity in an India dataset; [53] depicted a more complicated picture that a potential association of blood type with obesity may depend on both gender and age; etc. As hypertension and obesity are known risk for hospitalization, even if there were information on these comobidities in our data, it would not help to explain our conclusion through the "guilty by association" artifact.

Another potential confounding variable is vaccination status. Our data was initially collected before vaccine being available to the general public. Towards the end of time window of our collection, vaccine gradually became available, first to seniors, then to younger age groups. Because the vaccination status information is not available on our samples, the issue of vaccination impact on ABO blood type association with infection and hospitalization is addressed indirectly, by using the month in which the blood sample was collected as an imperfect surrogate to vaccination status. Our analysis and visual inspection of the raw data indicates that the association is slightly weaker when month is conditioned, but more samples are needed to have a conclusive result.

A recent analysis stated that blood type is not a significant predictor of prognosis or mortality after conditional on age, gender, number of comorbidities, etc. in a logistic regression [54] (sample size n = 670). On this result, the ethnic heterogeneity could make the number of variables used in a multi-variable logistic regression relative to the sample size, as a potential issue. Although in [54] the African American and Hispanic ethnicity are two co-variates, the sample size per ethnic group would be effectively smaller. The events per variable (EPV) (before variable selection) for multivariable regression has been discussed in the literature, and EPV = 50 was suggested in [55], among other proposed values. This would point to a number of 13 variables before selection, compared to the 11 variables in multiple logistic regression in [54]. Finally, it is not impossible that the signal contained in blood type overlaps with that contained in the number of comorbidities, then conditional on the latter would remove the signal from the former.

In conclusion, using a large dataset (six thousand COVID-19 patients, 25 thousand controls, more than 900 inpatients) collected in an ethnically homogenous region (Amasya, Turkey), we confirm the over-representation of anti-A-antibody-lacking blood types (A + AB) in COVID-19 patients than controls, and in inpatients than outpatients, and under-representation of anti-A- antibody-containing blood types (O + B). The effect size is relatively small (odds-ratio around $\sim$1.2). The inconsistent results from some of previous studies might be caused by other artifacts that may blur the relatively weak signal from ABO blood type, by genetic heterogeneity, and by small sample sizes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

WL acknowledges the support from Robert S Boas Center for Ge- nomics and Human Genetics. There is no external funding for the data analysis work.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Ethical approval was obtained from the Ministry of Health and Amasya University, Turkey and the protocol number E-63781-2022/38.

Informed consent

As the study was a retrospective analysis of medical records, informed consent was waived.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tracli.2022.10.003.

References

[1] Bhattacharjee S, Banerjee M, Pal R. ABO blood groups and severe outcomes in COVID-19: a meta- analysis. Postgraduate Med J 2022;98:e136–7.
[2] El-Shitany NA, El-Hamamsy M, Alahmadi AA, Eid BG, Neamatallah T, Almukadi HS, et al. The impact of ABO blood grouping on COVID-19 vulnerability and seriousness: a retrospective cross-sectional controlled study among the Arab community. Int J Environ Res Public Health 2021;18:276.

[3] Franchini M, Cruciani M, Mengoli C, Marano G, Candura F, Lopez N, et al. ABO blood group and COVID-19: an updated systematic literature review and meta-analysis. Blood Transfus 2021;19:317–26.
[4] Golinelli D, Boetto E, Maietti E, Fantini MP. The association between ABO blood group and SARS- CoV-2 infection: a meta-analysis. PLoS One 2020;15: e0239508.
[5] Gutíerrez-Valencia M, Leache L, Librero J, Jerićo C, Gerḿan ME, Garćıa-Erce JA. ABO blood group and risk of COVID-19 infection and complications: a systematic review and meta-analysis. *Transfusion* 2022;62:493–505.
[6] Wu BB, Gu DZ, Yu JN, Yang J, Shen WQ. Association between ABO blood groups and COVID-19 infection, severity and demise: a systematic review and meta-analysis. Infect Genet Evol 2020;84:104485.
[7] The Severe Covid-19 GWAS Group. Genomewide association study of severe covid-19 with respiratory failure. New Eng J Med 2020;383:1522–34.
[8] D'Antonio M, Nguyen JP, Arthur TD, Matsui H. The COVID-Host Genetics Initiative, A DAntonio- Chronowska, KA Frazer, SARS-CoV-2 susceptibility and COVID-19 disease severity are associated with genetic variants affecting gene expression in a variety of tissues. Cell Rep 2021;37:110020.
[9] Jenkins PV, O'Donnell JS. ABO blood group determines plasma von Willebrand factor levels: a biologic function after all? Transfusion 2006;46:1836–44.
[10] O'Donnell J, Laffan MA. The relationship between ABO histo-blood group, factor VIII and von Willebrand factor. Transf Med 2001;11:343–51.
[11] Zhou S, Welsby I. Is ABO blood group truly a risk factor for thrombosis and adverse outcomes? World J Cardiol 2014;6:985–92.
[12] Dentali F, Sironi AP, Ageno W, Turato S, Bonfanti C, Frattini F, et al. Non-O blood type is the commonest genetic risk factor for VTE: results from a meta-analysis of the literature. Seminars Thromb Hemost 2012;38:535–48.
[13] Moreau R, Dausset J, Bernard JMJ. Acquired hemolytic anemia with polyagglutinability of erythrocytes by a new factor present in normal blood. Bull Mem Soc Med Hop Paris 1957;73:569587.
[14] Guillon P, Clement M, Sébille V, Rivain JG, Chou CF, Ruvöen-Clouet N, Le Pendu J. Inhibition of the interaction between the SARS-CoV Spike protein and its cellular receptor by anti-histo-blood group antibodies. *Glycobiology* 2008;18:1085–93.
[15] Arend P. Why blood group A individuals are at risk whereas blood group O individuals are protected from SARS-CoV-2 (COVID-19) infection: a hypothesis regarding how the virus invades the human body via ABO(H) blood group-determining carbohydrates. Immunobiology 2021;226:152027.
[16] Deleers M, Breiman A, Daubie V, Maggetto C, Barreau I, Besse T, et al. Covid-19 and blood groups: ABO antibody levels may also matter. Int J Infec Dis 2021;104:242–9.
[17] Le J, Pendu, Breiman A, Rocher J, Dion M, Ruvöen-Clouet, N. ABO blood types and COVID-19: spurious, anecdotal, or truly important relationships? a reasoned review of available data. Virues 2021;13:60.
[18] Cavalli-Sforza LL, Menozzi P, Piazza A. The history and geography of human genes. Princeton University Press); 1994.
[19] Al-Youha SA, Alduaij W, Al-Serri A, Almazeedi SM, Al-Haddad M, Jamal MH, et al. The impact of ABO blood groups on clinical outcomes and susceptibility to COVID-19: a retrospective study in an unselected population. Transfusion 2021;61:1631–41.
[20] Kim Y, Latz CA, DeCarlo CS, Lee S, Png CYM, Kibrik P, et al. Relationship between blood type and outcomes following COVID-19 infection. Seminar Vas Surgery 2021;34:125–31.
[21] Ray JG, Schull MJ, Vermeulen MJ, Park AL. Association between ABO and Rh blood groups and SARS-CoV-2 infection or severe COVID-19 illness. Annals Intern Med 2021;174:308–15.
[22] Szymanski J, Mohrmann L, Carter J, Nelson R, Chekuri S, Assa A, et al. ABO blood type association with SARS-CoV-2 infection mortality: a single-center population in New York City. Transfusion 2021;61:1064–70.
[23] Zhao J, Yang Y, Huang H, Li D, Gu D, Lu X, et al. Relationship between the ABO blood group and the coronavirus disease 2019 (COVID-19) susceptibility. Clin Infec Dis 2021;73:328–31.
[24] Hoiland RL, Fergusson NA, Mitra AR, Griesdale DEG, Devine DV, Stukas S, et al. The association of ABO blood group with indices of disease severity and multiorgan dysfunction in COVID-19. Blood Adv 2020;4:4981–9.
[25] Zietz M, Zucker J, Tatonetti NP. Associations between blood type and COVID-19 infection, intubation, and death. Nat Commun 2020;11:5761.
[26] Almadhi MA, Abdulrahman A, Alawadhi A, Rabaan AA, Atkin S, AlQahtani M. The effect of ABO blood group and antibody class on the risk of COVID19 infection and severity of clinical outcomes. Sci Rep 2021;11:5745.
[27] Latz CA, DeCarlo C, Boitano L, Png CYM, Patell R, Conrad MF, et al. Blood type and outcomes in patients with COVID-19. Annals Hematol 2020;99:2113–8.
[28] Goker H, Aladag E, Demiroglu H, Ayaz C, Buyukasik Y, Inkaya A, Özcebe I, Ünal S, et al. The effects of blood group types on the risk of COVID-19 infection and its clinical outcome. Turkish J Med Sci 2020;50:679–83.
[29] Dal MS, Ata N, Altuntas F, Basci S, Yigenoglu TN, Korakmaz S, et al. COVID-19 clinical course and blood groups: Turkish population-based study. Turkish J Med Sci 2021;51:1659–64.
[30] Beksac M, Akin HY, GencerOncul EB, Yousefzadeh M, Seval GC, Gulten E, et al. A model integrating Killer Immunoglobulinlike Receptor (KIR) haplotypes for risk prediction of COVID19 clinical disease severity. Immunogenet 2021;73:449–58.
[31] Yilmaz M, Karaca A, Sozmen NN, Guncikan MN, Karadag GG, Hafizoglu N, et al. Association between COVID-19 and ABO blood groups: an analysis on convalescent plasma donors in Turkey. Med Bull Haseki 2021;59(suppl 1):7–14.
[32] Li W. Three lectures on case-control genetic association analysis. Brief Bioinf 2008;9:1–13.
[33] Cetin BS, Arslan D. The relationship between blood groups and COVID-19 patients (letter to the editor). Erciyes Med J 2021;43:624–5.
[34] Dzik S, Eliason K, Morris EB, Kaufman RM, North CN. COVID-19 and ABO blood groups (letter to the editor). Transfusion 2020;60:1883–4.
[35] Ulgen A, Cetin S, Balci PO, Sivgin H, Cetin M, Li W. COVID-19 outpatients and surviving inpatients exhibit comparable blood test results that are distinct from non-surviving inpatients. J Health Sci Med 2021;4:306–13.
[36] Ulgen A, Cetin S, Cetin M, Sivgin H, Li W. A composite ranking of risk factors for COVID-19 time- to-event data from a Turkish cohort. *Comp Biol Chem* 2022.
[37] Cetin M, Balci PO, Sivgin H, Cetin S, Ulgen A, Demir HD, et al. Alpha variant (B.1.1.7) of SARS- Cov-2 increases fatality-rate patients under age of 70 years and hospitalization risk overall. Acta Microbiol Immunol Hungarica 2021;68:153–61.
[38] Cetin S, Ulgen A, Balci PO, Sivgin H, Cetin M, Sivgin S, et al. Survival analyses of COVID-19 patients in a Turkish cohort: comparison between using time to death and time to release. SciMedicine J 2021;3(COVID-19 Special issue):1–9.
[39] Cetin S, Ulgen A, Sivgin H, Li W. A study on factors impacting length of hospital stay of COVID-19 patients. J Contemp Med 2021;11:396–404.
[40] Li W, Yang Y. Logistic regression in genome-wide association analysis. In: Elloumi M, Zomaya AY, editors. Biological Knowledge Discovery Handbook: Preprocessing, Mining and Postprocessing of Biological Data. Wiley; 2013. p. 477–500.
[41] Li W, Shih A, Freudenberg-Hua Y, Fury W, Yaning Y. Beyond standard pipeline and $p < 0.05$ in pathway enrichment analyses. Comp Biol Chem 2021;92:107455.
[42] Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond $p < 0.05$. Am Stat 2019;73(119).
[43] Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. BMJ 2001;322:1479–80.
[44] Cakir M. The relationship between blood groups and COVID-19 patients. Erciyes Med J 2021;43:142–5.
[45] Uyuklu M, Ozudogru O. Relationship between ABO blood group and COVID-19: the case of Siirt. Bezmialem Sci 2020;8(suppl 2):36–40.
[46] Coluk Y, Hizli O, Gunaydin S, Yildirim G, Baysal E, Hergul GO. Association of blood subgroups with PCR test positive and lung involvement in patients with COVID-19. Cureus 2021;13:e14172.
[47] Nalbant A, Aydin A, Yaylaci S, Kaya T, Wermeulen CL, Cinemre H. Association of ABO blood group and age with COVID-19 positive test. Rev Assoc Med Bras 2020;67(suppl 1):46–50.
[48] Torun YA, Kaynar LG, Karakukcu C, Yay M, Kurnaz F, Mutlu H, et al. ABO and Rh blood group distribution in Kayseri Province, Turkey. Turk J Haematol 2012;29:97–8.
[49] Arac E, Solmaz U, Akkoc H, Donmezdil S, Karahan Z, Kaya S, et al. Association between the Rh blood group and the COVID-19 susceptibility. Int J Hematol Oncol 2020;30:81–6.
[50] Gunduz M, Temel H. The relationship between the distribution of ABO and Rh blood groups and the clinical course in children with COVID-19. J Health Sci Adiyaman Univ 2021;7:32–7.
[51] Alwasaidi TA, Alrasheed SK, Alhazmi RA, Alfraidy OB, Jameel MA, Alandijani AA. Relation between ABO blood groups and obesity in a Saudi Arabian population. J Taibah Univ Med Sci 2017;12:407–11.
[52] Chandra T, Gupta A. Association and distribution of hypertension, obesity and ABO blood groups in blood donors. Iran J Ped Hematol Oncol 2012;2:140–5.
[53] Flor CR, Moura ICG, Baldoni AO, Loureiro P, Teixeira CM, Carneiro-Proietti AB, et al. Obesity and ABO blood group: Is there an association? Obesity Med 2020;18:100209.
[54] Thomas KE, Karamanis A, Dauchy E, Chapple AG, Loch MM. Assessing the effect of blood type on death and a novel scoring system to assess clinical course in patients with COVID-19. Am J Med Sci 2022;364:7–15.
[55] Steyerberg EW, Eijkemans MJC, Harrell Jr FE, Habbema JDF. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med 2000;19:1059–79.