Modelling suggests ABO histo-incompatibility may substantially reduce SARS-CoV-2 transmission

Peter J.I. Ellis

University of Kent School of Biosciences, Stacey Building, Canterbury, KENT, CT2 7NZ, UK

A B S T R A C T

Several independent datasets suggest blood type A is over-represented and type O under-represented among COVID-19 patients. However, blood group antigens appear not to be conventional susceptibility factors in that they do not affect disease severity, and the relative risk to non-O individuals is attenuated when population prevalence is high. Here, I model a scenario in which ABO transfusion incompatibility reduces the chance of a patient transmitting the virus to an incompatible recipient – thus in Western populations type A and AB individuals are "super-recipients" while type O individuals are "super-spreaders". This results in an offset in the timing of the epidemic among individuals of different blood types, and an increased relative risk to type A/AB patients that is most pronounced during early stages of the epidemic. However, once the majority of any given population is infected, the relative risk to each blood type approaches unity. Published data on COVID-19 prevalence from regions in the early stages of the SARS-CoV-2 epidemic suggests that if this model holds true, ABO incompatibility reduces virus transmissibility by at least 60 %. Exploring the implications of this model for vaccination strategies shows that paradoxically, targeted vaccination of either high-susceptibility type A/AB or "super-spreader" type O individuals is less effective than random vaccination at blocking community spread of the virus. Instead, the key is to maintain blood type diversity among the remaining susceptible individuals. Given the good agreement between this model and observational data on disease prevalence, the underlying biochemistry urgently requires experimental investigation.

1. Introduction

Multiple recent published studies and preprints have suggested that the prevalence of COVID-19 disease varies by blood type, with type A being relatively susceptible and type O being less susceptible. Initial reports from China were followed by confirmation in the US and Europe (Zhao et al., 2020; Zietz and Tatonetti, 2020; Kolin et al., 2021; Ellinghaus et al., 2020; Gérard et al., 2020), and have now been substantiated across the globe (reviewed in (le Pendu et al., 2021)). Although some studies report an association with disease severity, this is not consistently observed, with the case fatality ratio (CFR) and the probability of progressing to intensive care appearing independent of blood type in other studies (e.g. (Dzik et al., 2020)). Consistent with this, the most recent data (release r5) from the Covid-19 Host Genetics Initiative (The COVID-19 Host Genetics Initiative, 2020) indicates that the ABO gene is the most significant region genome-wide for SARS-CoV-2 infection rate, yet has no detectable association with disease severity when comparing hospitalized and non-hospitalized COVID-19 patients. Similarly, an association with disease incidence,

but not severity, is seen in data from home genetics tests carried out by 23andMe and AncestryDNA (Anon., 2021a; Roberts et al., 2021). However, despite the strong worldwide association between ABO blood group and infection risk, a recent study of 1769 crewmembers from the French Navy nuclear aircraft carrier *Charles de Gaulle* conclusively demonstrates that in this scenario, with an overall attack rate of 75.8 %, there was no correlation between blood type and incidence of infection (Boudin et al., 2020). These conflicting observations require explanation.

Here, I investigate the behaviour of "ABO-interference": a model of epidemic spread in which the transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, the causal agent of COVID-19) is dependent on the ABO blood type compatibility between an infected individual and the susceptible people they encounter. Mechanistically, this models a scenario in which infectious virions acquire the glycosylation pattern and hence the ABO antigen status of their current host. This in turn allows shed virions to be "rejected" by incompatible recipients, blocking the initial infective step. This immediately explains the lack of correlation with disease severity, since once an infection is

established, virions produced within the new host are necessarily self-compatible and able to spread freely between cells.

The plausibility of this hypothesis has already been established by work on HIV (Neil et al., 2005; Arendrup et al., 1991), and was also previously proposed for the 2003 epidemic of SARS (Cheng et al., 2005; Guillon et al., 2008), for measles (Preece et al., 2002), and indeed for enveloped viruses in general (Seymour et al., 2004; Cooling, 2015). Modelling of the proposed evolutionary interaction between virus transmission and host ABO antigens provides a convincing explanation for the worldwide distribution of A, B and O alleles at this locus (Seymour et al., 2004) (see also Discussion). SARS-CoV-2 has an outer lipid membrane containing spike, membrane and envelope (S, M and E) proteins, all of which are exposed to immune recognition and any or all of which may be glycosylated. Structural studies show that S is heavily glycosylated, including fucosylated glycans that may potentially bear ABO determinants (Cooling, 2015; Watanabe et al., 2020), and *in vitro* studies show that spike protein does indeed acquire ABO determinants when produced in cells expressing the relevant glycosyltransferases (Deleers et al., 2020). The glycosylation status of the M and E proteins has not yet been characterised, nor the glycosylation status of the membrane lipids. Cell-cell adhesion experiments using engineered CHO cells expressing S and an ACE2-expressing cell line show that anti-A antibodies can block spike-mediated adhesion, but only when the S-expressing cells also express the A glycosyltransferase (Guillon et al., 2008). This study also showed via modelling that ABO-interference can reduce the progress of an epidemic dependent on the magnitude of the block to transmission and the local population structure. However, despite the mechanistic plausibility of this hypothesis and the preliminary data from the SARS epidemic, there has as yet been no detailed modelling exploring the implications of ABO-interference for the relative susceptibility of individuals with different blood types at different stages of the epidemic, or for vaccination strategies.

In this analysis, I develop an extended SIR (**S**usceptible, **I**nfected, **R**ecovered) epidemiological model which allows for a partial or total block to virus transmission from an infected patient to an incompatible recipient, and explore the implications of this model for epidemic progression and for vaccination strategies. Strikingly, I find that this model is able to reconcile all the observations to date. ABO-interference does indeed lead to significant differences in disease incidence between different blood types, but only when the overall population prevalence is low.

2. Results

2.1. Modelling ABO-interference with virus transmission

The simplest of all epidemic models assumes a homogeneously mixing population divided into susceptible, infectious and recovered groups or "compartments". This yields three variables S, I and R, which respectively represent the proportion of the population with each status. Such a model is completely described by two parameters β and ν, representing the rate constants for infection and recovery respectively. The duration of the infectious period d is given by $1 / \nu$. All simulations presented here are based on $d = 7$ days, and none of the results presented are sensitive to this parameter. R0 is given by the product $\beta.d$ (or equivalently $\frac{\beta}{\nu}$) and represents the basic reproduction number, i.e. the mean number of individuals infected by a typical infectious individual during their illness, in the context of a completely susceptible population (Jones, 2007).

To this model I add a further parameter ρ where $0 \leq \rho < 100$ %, representing the relative probability of cross-type infection (Fig. 1, see Methods for full details). In this extended model, R0 is not a well-defined quantity since even in a fully susceptible population the current effective R value, denoted R(t), depends on ρ and on the blood type distributions in both S and I. Below, I use R_{max} to indicate the maximal possible value
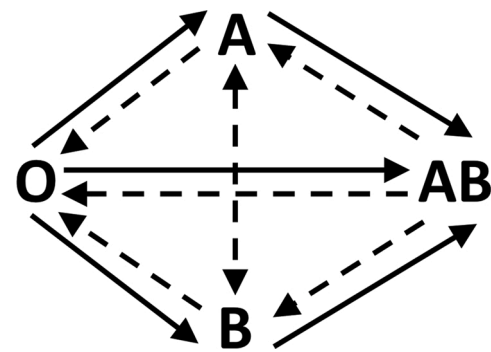

Fig. 1. Illustration of the transmission model. Full arrows represent unimpeded transmission (infection rate $= \beta$) and dashed arrows represent impeded transmission (infection rate $= \rho \cdot \beta$).

of R(t). This will be observed when the population is fully susceptible and transmission is unimpeded, i.e. when $\rho = 100$ % or all currently infected individuals are type O.

2.2. A quasi steady state is obtained when all blood types are infected

If $\rho = 0$ and the index case is non-O, then it is impossible for the epidemic to enter the O sub-population. Since this is not the case in the real world, this scenario is not explored further here. Fig. 2 illustrates the progress of simulated epidemics occurring under varying values of ρ and different population blood group distributions. In row **2A**, I analyse an epidemic with $R_{max} = 4$ and $\rho = 20$ %, occurring within a typical Western population with a high frequency of each of types O and A, and lower frequencies of types B and AB. Since $\rho = 20$ % represents a substantial block to transmission, the transmission rate R(t) during initial stages of the epidemic is depressed by around 1/3 compared to an unmitigated epidemic. In the early stages of the epidemic, the relative risk is highest for types A and AB, and since AB is rare at the population level, the majority of infections are type A individuals. Conversely, type O is relatively protected during early stages of the epidemic. By the later stages of the epidemic the pool of susceptible A, B and AB individuals has become depleted and thus type O cases predominate. In terms of cumulative cases, type AB (as the universal recipient) is at highest relative risk, and type O (as the most restrictive recipient) at lowest relative risk, however the magnitude of the difference is starkly different depending on whether the epidemic is in the late or early phase.

In row **2B,** I analyse an epidemic with $R_{max} = 4$ and $\rho = 80$ %, occurring within the same Western population. The same general pattern is observed, however the level of suppression of R(t) is less pronounced, and the overall difference in risk between the blood groups is greatly attenuated, due to the higher level of cross-type transmission. Rows **2C** and **2D** illustrate epidemics with $R_{max} = 4$ and $\rho = 20$ %, occurring within populations with blood type distributions typical of India or Peru, respectively. In the "Indian" model epidemic, the degree of R(t) suppression is similar to that seen in a Western population, but type B individuals are at higher risk than type A. In the "Peruvian" model epidemic, the high frequency of type O (universal transmitter) in the population means that R(t) is barely suppressed even though cross-type transmission is greatly restricted.

Importantly, whenever infection is present among individuals with all four blood types (i.e. there is at least one type O index case, or $\rho > 0$), the blood type distribution among infected people rapidly converges upon a new equilibrium that is significantly skewed relative to the initial population distribution. This equilibration process causes R(t) to also settle on an equilibrium value during the early stages of the epidemic, hereafter called R_{steady}. This is a rapid process, with R(t) reaching R_{steady} within a few serial intervals: in these simulations this equates to ~0.1 % of the population being infected. Therefore, in real-world epidemics
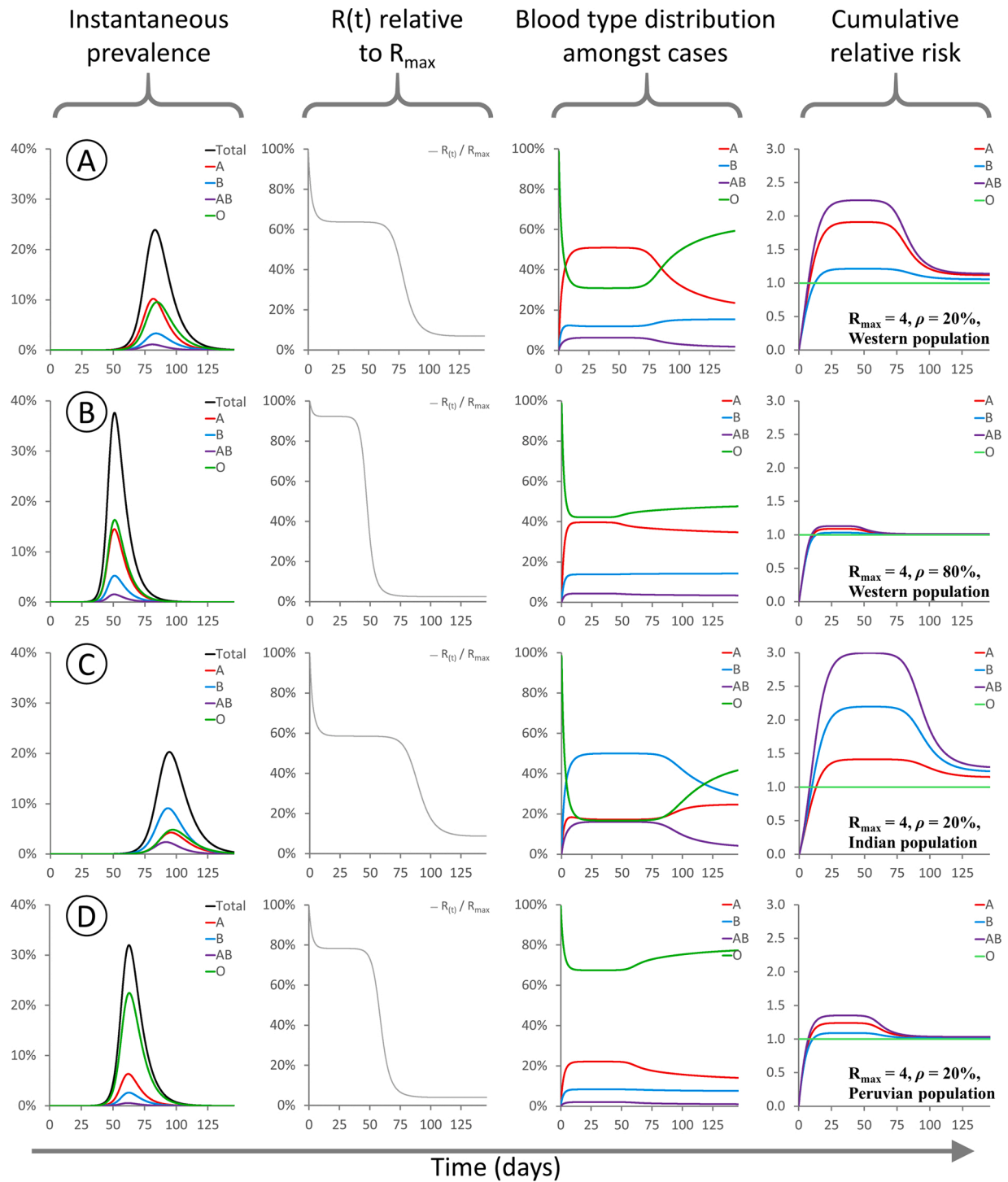
Fig. 2. These panels show the evolution of various modelled parameters for epidemics initialised under different conditions. For all epidemics $R_{max} = 4$ and the model is initialized with type O index cases comprising 1/1,000,000 of the population. Since type O transmits freely to all recipients, $R(t) = R_{max}$ at time $t = 0$.
A: $R_{max} = 4$, $\rho = 20$ %, "Western" population blood type distribution, 38 % A / 14 % B / 4 % AB / 44 % O.
B: $R_{max} = 4$, $\rho = 80$ %, "Western" population blood type distribution as above.
C: $R_{max} = 4$, $\rho = 20$ %, "Indian" population blood type distribution, 21.4 % A / 39.9 % B / 9.4 % AB / 29.3 % O (Cooling, 2015).
D: $R_{max} = 4$, $\rho = 20$ %, "Peruvian" population blood type distribution, 18.9 % A / 8.1 % B / 1.6 % AB / 71.4 % O (Watanabe et al., 2020).
For all sub-panels, X axis denotes days. Y axes: (first column) proportion of the population infected; (second column) R(t) as a fraction of R_{max}; (third column) distribution of blood types among currently infected individuals; (fourth column) cumulative risk of infection for each blood type relative to type O.

subject to ABO-interference, estimates of R0 based on population statistics are likely to actually measure R_{steady} and thus underestimate the true value of R_{max}.

2.3. Estimating ρ for SARS-CoV-2

During the early steady state described above, R_{steady} is a consistent multiple of R_{max}, indicating that ABO-interference suppresses the

epidemic with equal efficiency regardless of both the underlying infectiousness of the pathogen and any effect of non-pharmaceutical interventions on R(t). Similarly, the relative risk to each blood type is also dependent only on ρ and the background population blood type distribution, and independent of R_{max} (Supplementary Fig. 1). Thus, regardless of whether the epidemic in any given region is progressing quickly or slowly, the relative risk to each blood type is predictable from the population blood type frequencies and the value of ρ. Conversely, so long as total seroprevalence is low and the epidemic lies within the steady-state region of the curve, measurement of the relative risk for different blood types allows estimation of ρ.

Table 1 shows data compiled from a range of published studies that have reported ABO frequencies amongst infected SARS-CoV-2 patients and controls (Zhao et al., 2020; Zietz and Tatonetti, 2020; Ellinghaus et al., 2020) during the early stages of the first wave of the worldwide pandemic. These collectively cover epidemics in two regions of China, three regions of Spain, and one region each of Italy and the United States. For each data set, I used the control blood type distribution to predict the expected steady state distribution among infected patients for different values of ρ, and thus the expected case frequencies for each study, dependent upon sample size. A chi-square goodness-of-fit test was then used to compare the observed case frequencies to the model-based expectations. A low p value indicates that the observed case frequencies are incompatible with the model prediction, and thus the corresponding value of ρ is unlikely. Conversely, a higher p value indicates that the corresponding value of ρ is compatible with the observed data. Notably, for all data sets the chi-square peak p value lies between $\rho = 20\%$ and $\rho = 55\%$ despite the different underlying blood type frequencies and relative risk ratios in each country, indicating that in all regions of the world, the most plausible estimate of ρ lies in this range. For the Italian and Spanish study, data was also provided for area-matched control data from blood donors (Ellinghaus et al., 2020): using these instead of the internal control data does not affect the result.

Several caveats apply to this analysis. Firstly, poor demographic matching between cases and controls may distort the results. This is apparent in the Spanish sample from (Ellinghaus et al., 2020), where the distribution of cases and controls was uneven between contributing hospitals. Thus, for the aggregate sample, all values of ρ have a low chi-square p value indicating a poor fit between model and data. In this case, using locally matched blood donor samples results in a better fit without altering the overall findings (Table 1 and Supplementary Fig. 2).

Secondly, the "steady state" only holds during the early stages of the epidemic, and the relative risk approaches unity once the majority of the population is infected as in the case of the *Charles de Gaulle* study (Boudin et al., 2020)). Even in the countries analysed here, in which overall prevalence was low at the time of study, the local prevalence in "hotspots" is much higher (Percivalle et al., 2020). Infected individuals will by definition originate preferentially from hotspot regions. In such areas, the epidemic may have progressed beyond the steady state region of the epidemic curve, reducing the degree of blood type skewing among the infected population and partially masking the effects of ABO-interference on virus transmission. In this case, in order to generate the observed changes in blood type frequency, then the degree of ABO-interference must be even more pronounced. i.e. the estimate of ~40 % is a maximum bound, and the true frequency of cross-type transmission may be even lower.

Thirdly, the SIR model assumes that population mixing and opportunities for transmission are independent of blood type. This is unlikely to be the case since blood relatives living together are both more likely to infect each other and more likely to share a blood type. This effect will also tend to mask the effect of ABO-interference. This effect will be most pronounced in areas with relatively segregated ethnic communities, as for example in the New York sample where black and white Americans have differing blood type distributions and social mixing is also likely to be uneven.

Fourthly, when there is nosocomial spread within a hospital, an increase in frequency of one blood type among infected patients will lead to a decrease in frequency of that blood type in the remaining uninfected

Table 1

Observed case and control numbers are taken from references Zhao et al. (2020), Zietz and Tatonetti (2020), Ellinghaus et al. (2020). The Italian and Spanish data sets provided both internal and external controls. For each area, the control blood group frequencies were used to calculate expected case numbers for each blood group in each study, for varying values of ρ. Calculating the chi-square goodness of fit between the predicted case numbers and observed numbers allows region-specific estimation of ρ. The orange highlight indicates the most plausible value of ρ for each study, while the yellow highlight indicates all values with p > 0.05 after Bonferroni multiple testing correction.

	Controls n	Controls %	Cases n	Cases %	100%		90%		80%		70%		60%		50%		40%		30%		20%		10%	
Italy (Milan area) (internal controls)																								
A	451	35.94%	388	46.47%	300.1	303.0	306.2	309.6	313.1	317.0	321.1	325.6	330.5	335.9	341.7	348.3	355.6	363.8	373.4	384.5	397.7	414.7	436.7	468.8
B	163	12.99%	91	10.90%	108.5	108.0	108.0	107.7	107.4	107.0	106.6	106.2	105.6	105.0	104.3	103.4	102.4	101.2	99.7	97.8	95.4	92.2	87.6	80.3
AB	50	3.98%	43	5.15%	33.3	33.9	34.5	35.2	36.0	36.7	37.5	38.4	39.4	40.4	41.5	42.7	44.1	45.5	47.2	49.1	51.3	54.0	57.4	62.0
O	591	47.09%	313	37.49%	393.2	389.8	386.3	382.5	378.5	374.2	369.7	364.8	359.5	353.7	347.5	340.6	333.0	324.4	314.8	303.6	290.3	274.2	253.4	224.0
χ2 p value for observed vs expected cases					0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0020	0.0096	0.0408	0.1413	0.3657	0.6342	0.6695	0.3121	0.0238	0.0000	0.0000	0.0000
Spain (mixed areas) (internal controls)																								
A	398	41.89%	377	48.65%	324.7	328.4	332.4	336.7	341.2	346.0	351.2	356.7	362.8	369.4	376.6	384.5	393.4	403.4	414.8	428.1	444.0	463.6	489.2	526.7
B	65	6.84%	71	9.16%	53.0	52.7	52.3	51.9	51.5	51.0	50.5	49.9	49.3	48.6	47.8	47.0	46.0	44.8	43.5	41.9	40.0	37.5	34.2	29.1
AB	25	2.63%	36	4.65%	20.4	20.7	21.1	21.5	21.8	22.3	22.7	23.2	23.7	24.2	24.8	25.5	26.2	27.0	27.8	28.9	30.1	31.5	33.3	35.9
O	462	48.63%	291	37.55%	376.9	373.1	369.2	365.0	360.5	355.7	350.6	345.2	339.2	332.8	325.8	318.1	309.5	299.8	288.8	276.1	261.0	242.5	218.3	183.2
χ2 p value for observed vs expected cases					0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
China (Wuhan) (internal controls)																								
A	1188	32.16%	715	37.87%	607.2	610.2	613.4	616.8	620.4	624.3	628.4	633.0	637.9	643.3	649.3	656.0	663.6	672.3	682.5	694.7	710.1	730.4	760.0	813.1
B	920	24.91%	494	26.17%	470.2	470.8	471.3	471.8	472.3	472.8	473.2	473.6	473.9	474.2	474.2	474.1	473.8	473.1	471.8	469.6	466.1	459.9	448.3	421.3
AB	336	9.10%	193	10.22%	171.7	175.6	179.7	184.0	188.7	193.6	199.0	204.8	211.0	217.9	225.4	233.7	243.0	253.5	265.6	279.7	296.7	317.8	345.9	387.9
O	1250	33.84%	486	25.74%	638.9	631.5	623.7	615.4	606.6	597.3	587.3	576.7	565.1	552.7	539.1	524.2	507.6	489.1	468.2	443.9	415.2	379.9	333.8	265.7
χ2 p value for observed vs expected cases					0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0006	0.0011	0.0011	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000
China (Shenzhen) (internal controls)																								
A	6728	28.77%	82	28.77%	82.0	82.4	82.8	83.3	83.8	84.3	84.8	85.4	86.1	86.8	86.8	87.6	88.4	89.4	90.4	91.7	93.1	94.8	97.0	100.0
B	5880	25.14%	83	29.12%	71.7	71.9	72.1	72.4	72.6	72.9	73.1	73.4	73.7	74.0	74.4	74.7	75.1	75.5	75.9	76.3	76.8	77.2	77.5	77.5
AB	1712	7.32%	39	13.68%	20.9	21.3	21.8	22.3	22.8	23.4	24.0	24.7	25.4	26.2	27.1	28.1	29.1	30.3	31.7	33.3	35.2	37.5	40.6	44.9
O	9066	38.77%	81	28.42%	110.5	109.4	108.3	107.1	105.8	104.4	103.0	101.4	99.7	97.9	96.0	93.8	91.4	88.7	85.7	82.3	78.2	73.3	66.9	58.0
χ2 p value for observed vs expected cases					0.0000	0.0000	0.0001	0.0001	0.0003	0.0007	0.0015	0.0033	0.0071	0.0149	0.0306	0.0604	0.1131	0.1959	0.3046	0.4061	0.4305	0.3034	0.0847	0.0017
US (internal controls)																								
A	245	27.94%	233	34.16%	190.5	192.0	193.6	195.2	197.0	198.9	200.9	203.0	205.4	207.9	210.6	213.7	217.0	220.7	225.0	229.8	235.5	242.3	250.7	261.7
B	136	15.51%	116	17.01%	105.8	105.9	106.0	106.2	106.3	106.4	106.6	106.7	106.8	106.9	106.9	107.0	107.0	106.9	106.8	106.6	106.3	105.8	104.9	103.5
AB	47	5.36%	21	3.08%	36.5	37.2	37.9	38.7	39.5	40.3	41.2	42.2	43.2	44.3	45.5	46.8	48.2	49.8	51.5	53.5	55.8	58.4	61.6	65.7
O	449	51.20%	312	45.75%	349.2	346.9	344.5	341.9	339.2	336.4	333.4	330.1	326.7	323.0	319.0	314.6	309.8	304.6	298.7	292.1	284.5	275.6	264.8	251.1
χ2 p value for observed vs expected cases					0.0001	0.0001	0.0002	0.0003	0.0004	0.0005	0.0006	0.0008	0.0009	0.0009	0.0009	0.0008	0.0006	0.0004	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
Italy (Milan area) (local area donor controls)																								
A	5629	38.40%	388	46.47%	320.7	323.8	327.2	330.9	334.7	338.9	343.4	348.3	353.6	359.4	365.9	373.1	381.2	390.4	401.0	413.7	429.0	448.4	474.5	514.9
B	1942	13.25%	91	10.90%	110.6	110.3	109.9	109.5	109.0	108.5	107.9	107.2	106.4	105.5	104.5	103.3	101.9	100.2	98.1	95.5	92.2	87.7	81.2	70.4
AB	673	4.59%	43	5.15%	38.3	39.1	39.8	40.7	41.5	42.4	43.4	44.5	45.6	46.8	48.1	49.6	51.2	53.0	55.0	57.3	60.0	63.3	67.6	73.6
O	6414	43.76%	313	37.49%	365.4	361.8	358.0	354.0	349.7	345.2	340.3	335.0	329.4	323.2	316.5	309.0	300.8	291.5	280.9	268.5	253.8	235.6	211.7	176.1
χ2 p value for observed vs expected cases					0.0000	0.0000	0.0000	0.0001	0.0005	0.0048	0.0140	0.0881	0.1790	0.2995	0.3930	0.3770	0.2281	0.0647	0.0052	0.0000	0.0000	0.0000	0.0000	0.0000
Spain (San Sebastien) (Basque donor controls)																								
A	7235	40.66%	148	49.17%	122.4	123.8	125.3	126.9	128.7	130.5	132.5	134.6	136.8	139.3	142.0	144.9	148.2	151.9	156.1	160.9	166.6	173.6	182.5	195.1
B	918	5.16%	19	6.31%	15.5	15.4	15.3	15.2	15.1	14.9	14.8	14.6	14.5	14.3	14.0	13.8	13.5	13.2	12.9	12.5	12.0	11.3	10.5	9.3
AB	355	1.99%	11	3.65%	6.0	6.1	6.2	6.3	6.4	6.5	6.6	6.8	6.9	7.1	7.2	7.4	7.6	7.8	8.1	8.3	8.6	9.0	9.5	10.2
O	9288	52.19%	123	40.86%	157.1	155.7	154.2	152.6	150.9	149.0	147.1	145.0	142.8	140.4	137.7	134.8	131.6	128.1	124.0	119.3	113.8	107.1	98.5	86.4
χ2 p value for observed vs expected cases					0.0005	0.0010	0.0018	0.0034	0.0062	0.0113	0.0200	0.0345	0.0574	0.0909	0.1355	0.1864	0.2311	0.2491	0.2198	0.1423	0.0548	0.0083	0.0002	0.0000
Spain (Barcelona) (local area donor controls)																								
A	14617	42.03%	136	49.82%	114.8	116.0	117.4	118.8	120.3	121.9	123.7	125.6	127.7	130.0	132.4	135.2	138.3	141.9	145.9	150.7	156.4	163.6	173.3	187.8
B	3355	9.65%	35	12.82%	26.3	26.2	26.0	25.8	25.6	25.4	25.2	24.9	24.6	24.3	23.9	23.5	23.0	22.4	21.8	20.9	19.9	18.5	16.7	13.8
AB	1309	3.76%	7	2.56%	10.3	10.5	10.7	10.9	11.1	11.3	11.5	11.8	12.1	12.4	12.7	13.1	13.5	13.9	14.4	15.0	15.6	16.4	17.5	19.0
O	15493	44.55%	95	34.80%	121.6	120.3	119.0	117.5	116.0	114.3	112.6	110.7	108.6	106.4	103.9	101.2	98.2	94.8	90.9	86.4	81.0	74.4	65.6	52.5
χ2 p value for observed vs expected cases					0.0034	0.0049	0.0069	0.0096	0.0132	0.0175	0.0226	0.0280	0.0329	0.0360	0.0359	0.0316	0.0234	0.0134	0.0053	0.0011	0.0001	0.0000	0.0000	0.0000
Spain (Madrid) (local area donor controls)																								
A	15145	41.18%	93	46.27%	82.8	83.7	84.6	85.7	86.8	88.0	89.2	90.6	92.1	93.7	95.5	97.5	99.7	102.3	105.2	108.6	112.7	117.8	124.6	134.8
B	3511	9.55%	17	8.46%	19.2	19.1	19.0	18.9	18.7	18.6	18.4	18.2	18.0	17.8	17.6	17.3	16.9	16.5	16.1	15.5	14.8	13.9	12.6	10.6
AB	1306	3.55%	18	8.96%	7.1	7.3	7.4	7.5	7.7	7.8	8.0	8.2	8.4	8.6	8.8	9.1	9.3	9.6	10.0	10.4	10.8	11.4	12.1	13.1
O	16819	45.73%	73	36.32%	91.9	91.0	90.0	88.9	87.8	86.6	85.4	84.0	82.5	80.9	79.1	77.2	75.0	72.6	69.8	66.6	62.7	57.9	51.7	42.4
χ2 p value for observed vs expected cases					0.0001	0.0001	0.0002	0.0004	0.0007	0.0013	0.0023	0.0040	0.0067	0.0111	0.0175	0.0261	0.0357	0.0436	0.0446	0.0346	0.0168	0.0033	0.0001	0.0000

hospital patients. This will exaggerate the effects of ABO-interference if uninfected hospitalised patients are used as controls.

Since the majority of these effects will tend to oppose the effects of ABO-interference and thus bias the estimate of ρ upwards, the central estimate of ~40 % represents a best-guess upper bound for ρ. Intriguingly, previous data from direct contact tracing (Guillon et al., 2008) imply ρ = 47.7 % for the 2003 hospital outbreak of SARS in Hong Kong, suggesting that values around this range may be a feature of coronavirus infections in general.

2.4. Is targeting by blood type a useful vaccination strategy?

Initial preprints noting the increased risk to type A individuals have proposed that these may require additional surveillance and priority for protection. However, ABO-interference with virus transmission presents a unique and striking scenario that has not previously been modelled in
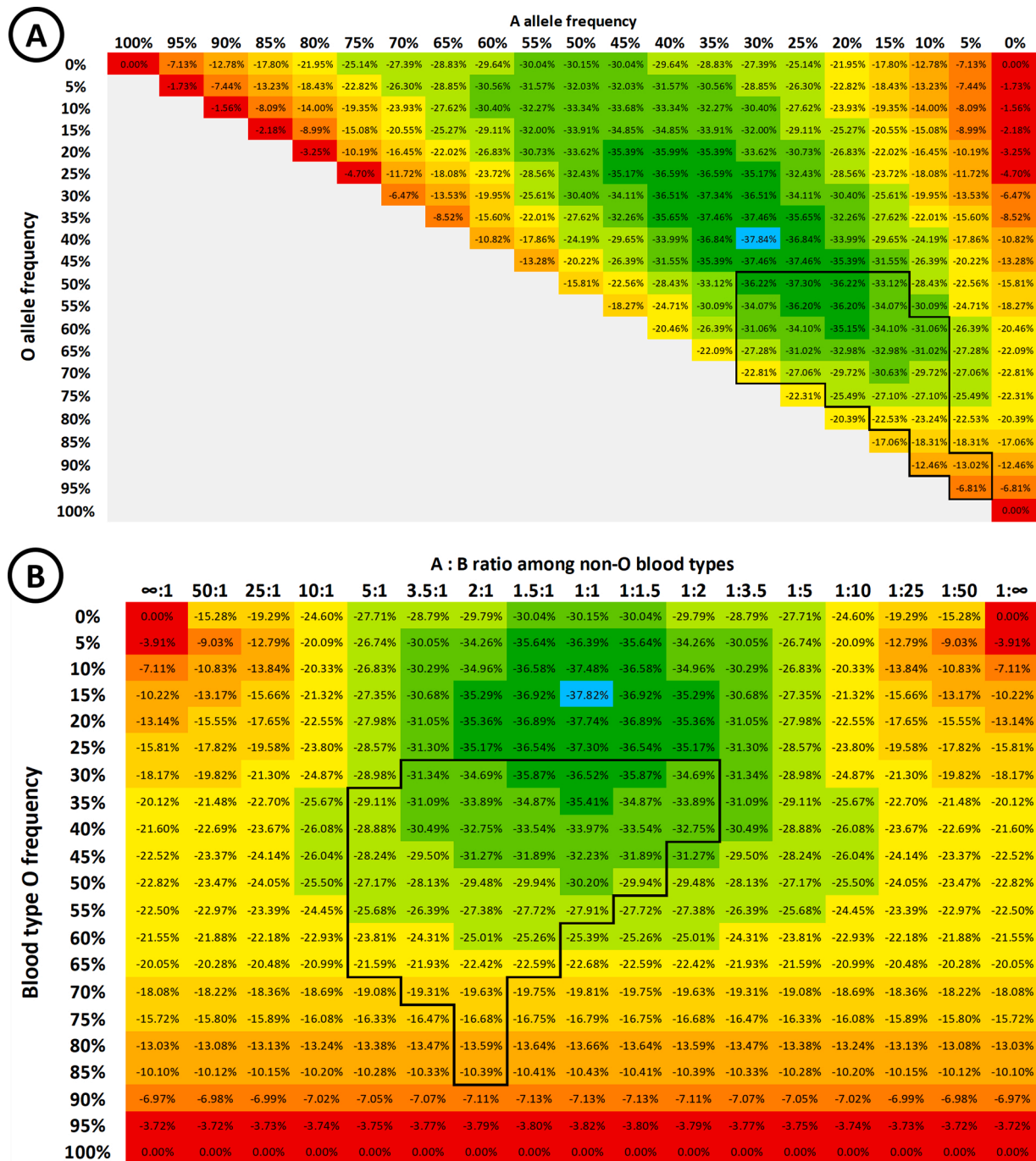
A — A allele frequency

O allele freq	100%	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%	45%	40%	35%	30%	25%	20%	15%	10%	5%	0%
0%	0.00%	-7.13%	-12.78%	-17.80%	-21.95%	-25.14%	-27.39%	-28.83%	-29.64%	-30.04%	-30.15%	-30.04%	-29.64%	-28.83%	-27.39%	-25.14%	-21.95%	-17.80%	-12.78%	-7.13%	0.00%
5%		-1.73%	-7.44%	-13.23%	-18.43%	-22.82%	-26.30%	-28.85%	-30.56%	-31.57%	-32.03%	-32.03%	-31.57%	-30.56%	-28.85%	-26.30%	-22.82%	-18.43%	-13.23%	-7.44%	-1.73%
10%			-1.56%	-8.09%	-14.00%	-19.35%	-23.93%	-27.62%	-30.40%	-32.27%	-33.34%	-33.68%	-33.34%	-32.27%	-30.40%	-27.62%	-23.93%	-19.35%	-14.00%	-8.09%	-1.56%
15%				-2.18%	-8.99%	-15.08%	-20.55%	-25.27%	-29.11%	-32.00%	-33.91%	-34.85%	-34.85%	-33.91%	-32.00%	-29.11%	-25.27%	-20.55%	-15.08%	-8.99%	-2.18%
20%					-3.25%	-10.19%	-16.45%	-22.02%	-26.83%	-30.73%	-33.62%	-35.39%	-35.99%	-35.39%	-33.62%	-30.73%	-26.83%	-22.02%	-16.45%	-10.19%	-3.25%
25%						-4.70%	-11.72%	-18.08%	-23.72%	-28.56%	-32.43%	-35.17%	-36.59%	-36.59%	-35.17%	-32.43%	-28.56%	-23.72%	-18.08%	-11.72%	-4.70%
30%							-6.47%	-13.53%	-19.95%	-25.61%	-30.40%	-34.11%	-36.51%	-37.34%	-36.51%	-34.11%	-30.40%	-25.61%	-19.95%	-13.53%	-6.47%
35%								-8.52%	-15.60%	-22.01%	-27.62%	-32.26%	-35.65%	-37.46%	-37.46%	-35.65%	-32.26%	-27.62%	-22.01%	-15.60%	-8.52%
40%									-10.82%	-17.86%	-24.19%	-29.65%	-33.99%	-36.84%	-37.84%	-36.84%	-33.99%	-29.65%	-24.19%	-17.86%	-10.82%
45%										-13.28%	-20.22%	-26.39%	-31.55%	-35.39%	-37.46%	-37.46%	-35.39%	-31.55%	-26.39%	-20.22%	-13.28%
50%											-15.81%	-22.56%	-28.43%	-33.12%	-36.22%	-37.30%	-36.22%	-33.12%	-28.43%	-22.56%	-15.81%
55%												-18.27%	-24.71%	-30.09%	-34.07%	-36.20%	-36.20%	-34.07%	-30.09%	-24.71%	-18.27%
60%													-20.46%	-26.39%	-31.06%	-34.10%	-35.15%	-34.10%	-31.06%	-26.39%	-20.46%
65%														-22.09%	-27.28%	-31.02%	-32.98%	-32.98%	-31.02%	-27.28%	-22.09%
70%															-22.81%	-27.06%	-29.72%	-30.63%	-29.72%	-27.06%	-22.81%
75%																-22.31%	-25.49%	-27.10%	-27.10%	-25.49%	-22.31%
80%																	-20.39%	-22.53%	-23.24%	-22.53%	-20.39%
85%																		-17.06%	-18.31%	-18.31%	-17.06%
90%																			-12.46%	-13.02%	-12.46%
95%																				-6.81%	-6.81%
100%																					0.00%

B — A : B ratio among non-O blood types

Blood type O frequency	∞:1	50:1	25:1	10:1	5:1	3.5:1	2:1	1.5:1	1:1	1:1.5	1:2	1:3.5	1:5	1:10	1:25	1:50	1:∞
0%	0.00%	-15.28%	-19.29%	-24.60%	-27.71%	-28.79%	-29.79%	-30.04%	-30.15%	-30.04%	-29.79%	-28.79%	-27.71%	-24.60%	-19.29%	-15.28%	0.00%
5%	-3.91%	-9.03%	-12.79%	-20.09%	-26.74%	-30.05%	-34.26%	-35.64%	-36.39%	-35.64%	-34.26%	-30.05%	-26.74%	-20.09%	-12.79%	-9.03%	-3.91%
10%	-7.11%	-10.83%	-13.84%	-20.33%	-26.83%	-30.29%	-34.96%	-36.58%	-37.48%	-36.58%	-34.96%	-30.29%	-26.83%	-20.33%	-13.84%	-10.83%	-7.11%
15%	-10.22%	-13.17%	-15.66%	-21.32%	-27.35%	-30.68%	-35.29%	-36.92%	-37.82%	-36.92%	-35.29%	-30.68%	-27.35%	-21.32%	-15.66%	-13.17%	-10.22%
20%	-13.14%	-15.55%	-17.65%	-22.55%	-27.98%	-31.05%	-35.36%	-36.89%	-37.74%	-36.89%	-35.36%	-31.05%	-27.98%	-22.55%	-17.65%	-15.55%	-13.14%
25%	-15.81%	-17.82%	-19.58%	-23.80%	-28.57%	-31.30%	-35.17%	-36.54%	-37.30%	-36.54%	-35.17%	-31.30%	-28.57%	-23.80%	-19.58%	-17.82%	-15.81%
30%	-18.17%	-19.82%	-21.30%	-24.87%	-28.98%	-31.34%	-34.69%	-35.87%	-36.52%	-35.87%	-34.69%	-31.34%	-28.98%	-24.87%	-21.30%	-19.82%	-18.17%
35%	-20.12%	-21.48%	-22.70%	-25.67%	-29.11%	-31.09%	-33.89%	-34.87%	-35.41%	-34.87%	-33.89%	-31.09%	-29.11%	-25.67%	-22.70%	-21.48%	-20.12%
40%	-21.60%	-22.69%	-23.67%	-26.08%	-28.88%	-30.49%	-32.75%	-33.54%	-33.97%	-33.54%	-32.75%	-30.49%	-28.88%	-26.08%	-23.67%	-22.69%	-21.60%
45%	-22.52%	-23.37%	-24.14%	-26.04%	-28.24%	-29.50%	-31.27%	-31.89%	-32.23%	-31.89%	-31.27%	-29.50%	-28.24%	-26.04%	-24.14%	-23.37%	-22.52%
50%	-22.82%	-23.47%	-24.05%	-25.50%	-27.17%	-28.13%	-29.48%	-29.94%	-30.20%	-29.94%	-29.48%	-28.13%	-27.17%	-25.50%	-24.05%	-23.47%	-22.82%
55%	-22.50%	-22.97%	-23.39%	-24.45%	-25.68%	-26.39%	-27.38%	-27.72%	-27.91%	-27.72%	-27.38%	-26.39%	-25.68%	-24.45%	-23.39%	-22.97%	-22.50%
60%	-21.55%	-21.88%	-22.18%	-22.93%	-23.81%	-24.31%	-25.01%	-25.26%	-25.39%	-25.26%	-25.01%	-24.31%	-23.81%	-22.93%	-22.18%	-21.88%	-21.55%
65%	-20.05%	-20.28%	-20.48%	-20.99%	-21.59%	-21.93%	-22.42%	-22.59%	-22.68%	-22.59%	-22.42%	-21.93%	-21.59%	-20.99%	-20.48%	-20.28%	-20.05%
70%	-18.08%	-18.22%	-18.36%	-18.69%	-19.08%	-19.31%	-19.63%	-19.75%	-19.81%	-19.75%	-19.63%	-19.31%	-19.08%	-18.69%	-18.36%	-18.22%	-18.08%
75%	-15.72%	-15.80%	-15.89%	-16.08%	-16.33%	-16.47%	-16.68%	-16.75%	-16.79%	-16.75%	-16.68%	-16.47%	-16.33%	-16.08%	-15.89%	-15.80%	-15.72%
80%	-13.03%	-13.08%	-13.13%	-13.24%	-13.38%	-13.47%	-13.59%	-13.64%	-13.66%	-13.64%	-13.59%	-13.47%	-13.38%	-13.24%	-13.13%	-13.08%	-13.03%
85%	-10.10%	-10.12%	-10.15%	-10.20%	-10.28%	-10.33%	-10.39%	-10.41%	-10.43%	-10.41%	-10.39%	-10.33%	-10.28%	-10.20%	-10.15%	-10.12%	-10.10%
90%	-6.97%	-6.98%	-6.99%	-7.02%	-7.05%	-7.07%	-7.11%	-7.13%	-7.13%	-7.13%	-7.11%	-7.07%	-7.05%	-7.02%	-6.99%	-6.98%	-6.97%
95%	-3.72%	-3.72%	-3.73%	-3.74%	-3.75%	-3.77%	-3.79%	-3.80%	-3.82%	-3.80%	-3.79%	-3.77%	-3.75%	-3.74%	-3.73%	-3.72%	-3.72%
100%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Fig. 3. A: Heat map showing $\left(1 - \frac{R_{steady}}{R_{max}}\right)$, i.e. the degree of suppression of R0 during the early "steady state" portion of an epidemic with ρ = 30 %, according to the population ABO allele frequencies. B allele frequency = (100 % - O% - A%). **B:** Heat map showing the degree of suppression of R0 during the early stages of the epidemic, according to the background population blood type frequencies. In both heat maps, ABO allele frequencies are assumed to be in Hardy-Weinberg equilibrium – i.e. allele frequencies are not changing from generation to generation, there is no net selection or gene flow, and that mating is random with respect to genotype. The colour gradient from red to green indicates the degree of suppression in 5 % increments, with the blue highlight indicating maximal suppression of virus transmission. The boxed areas of the plots indicate allele / blood type frequencies that are typically observed in human populations (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

detail, in which those most prone to infection are those least likely to pass it on, and vice versa. This raises the question as to whether it is more important to vaccinate the most susceptible individuals, or the most infectious individuals. Fig. 3A shows $(1 - \frac{R_{steady}}{R_{max}})$, i.e. the degree to which R0 is suppressed by ABO-interference, across the full spectrum of potential ABO allele frequencies for $\rho = 30$ %. ABO-interference suppresses transmission most efficiently when the allele frequency ratio is approximately 40 % O / 30 % A / 30 % B alleles. Translating allele frequencies to blood type frequencies yields Fig. 3B. Thus, ABO-interference suppresses transmission most efficiently when type O individuals make up 15 % of the population and type A / type B individuals are present in equal proportions. This is however not the case for any existing human populations, since O is the most common blood type worldwide (see Discussion). A similar shape heat map is obtained for values of $\rho = 20$ % and $\rho = 90$ % (not shown). Amongst actual populations worldwide, the effect of ABO-interference in lowering R(t) is likely to be most pronounced within East and Southeast Asian countries, where type O is only moderate frequency, and A and B are more evenly balanced.

From the standpoint of vaccination strategy, vaccinating type O individuals effectively moves the population upwards in Fig. 3B, while vaccinating type A or B moves the population right or left respectively. In principle, an optimal vaccination strategy will cause the distribution among susceptible individuals to move "down" the gradient towards more effective suppression of the epidemic. Conversely, the vaccination strategy must also be careful *not* to disrupt the intrinsic protection afforded by ABO-interference. To illustrate this, consider a population with 50 % type A and 50 % type O individuals, similar to Māori and some Polynesian populations where the type B frequency is very low (Edinur et al., 2013). In general, the predicted herd immunity threshold is $\left(\frac{R0-1}{R0}\right)$, so in the absence of ABO-interference the threshold for an epidemic with an R_{max} of 3 is 66.7 %. In such a population, if $\rho = 30$ % then $R_{steady} = 2.32$, ald the risk for type A individuals is 1.82 times higher than type O individuals, and type O individuals are 1.54 times as infectious as type A individuals.

A well-intentioned strategy to reduce infection might prioritise vaccinating type O "super-spreaders" before type A. However, once all type O individuals have been immunised, the protective effect of ABO-interference is abolished since the remaining susceptible population is now exclusively type A, and an infected type A individual can freely transmit to any remaining susceptible individual. Herd immunity will therefore only be attained when the full 66.7 % of the population is vaccinated. The same applies in reverse if the more vulnerable type A individuals are instead prioritised for vaccination. However, vaccinating both blood types equally produces herd immunity when 56.9 % of the population is vaccinated, consistent with R_{steady} in this population. This effect is further magnified if transmission is brought down by other means, for example non-pharmaceutical interventions including social distancing. For the same population and same value of $\rho = 30$ %, if

$R_{max} = 2$ then $R_{steady} = 1.55$. In this case the herd immunity threshold is 50 % of the population if preferentially vaccinating either type O or type A, but only 35.4 % if vaccinating individuals at random [Fig. 4].

2.5. The effect of waning immunity for a virus subject to ABO-interference

At present it is unknown whether SARS-CoV-2 leads to long-lasting immunity. The four other endemic human coronaviruses do not confer lasting immunity, and repeat infection is common. In an SIR model, when immunity is allowed to wane over time, then population spread of the virus can resume once the population level of immunity falls below the herd immunity threshold. Over time this leads to damped oscillatory behaviour, with recurrent pulses of infection converging on a final steady state. In this final steady state, by definition R(t) = 1 and there is steady sustained transmission with overall infection levels neither growing nor shrinking [Fig. 5]. The final population disease burden depends on the duration of immunity, with a shorter immune duration leading to a higher population-wide average prevalence. In the endemic state, with waning immunity and recurrent infection, the relative risk to each blood type should be interpreted as a difference in the frequency of infection: e.g. if type A has a relative risk of 1.1 compared to type O, it means type A individuals are 10 % more likely to suffer an infection within any given time period.

Intriguingly, while the relative risk to different blood groups during disease emergence is independent of R_{max} (see Supplementary Fig. 1), the final steady state for an endemic disease with waning immunity is independent of ω but does depend on R_{max} (compare **5A, B, C**). When R_{max} is high, the relative risk at equilibrium for all blood types is close to unity. When R_{max} is low, then the relative risk at equilibrium for all blood types is similar to that seen during disease emergence. Intuitively, this follows from the fact that for a highly contagious disease with a high R_{max}, everyone will become infected as soon as their immunity wears off, irrespective of blood type. For a less contagious disease with a low R_{max}, then there is more scope for differential susceptibility to play a part in how frequently individuals become infected.

3. Discussion

3.1. The biological interpretation of the parameter ρ

In this model, ρ represents the relative probability of virus transmission between an infected individual and an ABO-incompatible target individual. Mechanistically, this will encompass at least three sources of variability: (a) the extent to which the infected individual deposits ABO antigen on the surface of the virions produced, (b) the extent to which the target individual carries anti-A or anti-B antibodies, and (c) the ability of these antibodies to access the incoming virions and prevent infection.

Of these, (a) will depend not only on the host's ABO genotype, but also on their Secretor (Se) status (Watkins, 1980). "Non-secretor" individuals are homozygous for null mutations in the FUT2 gene. In these
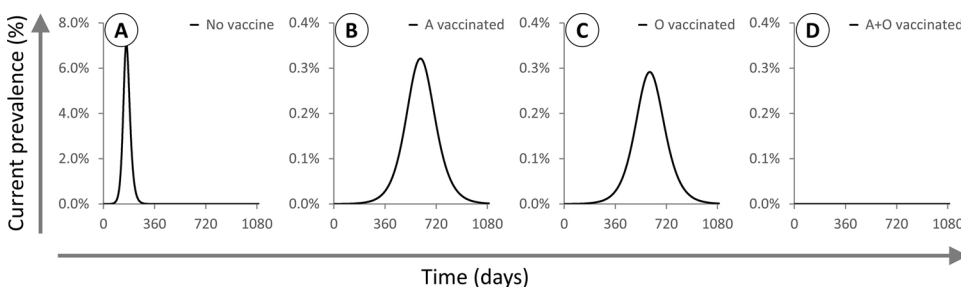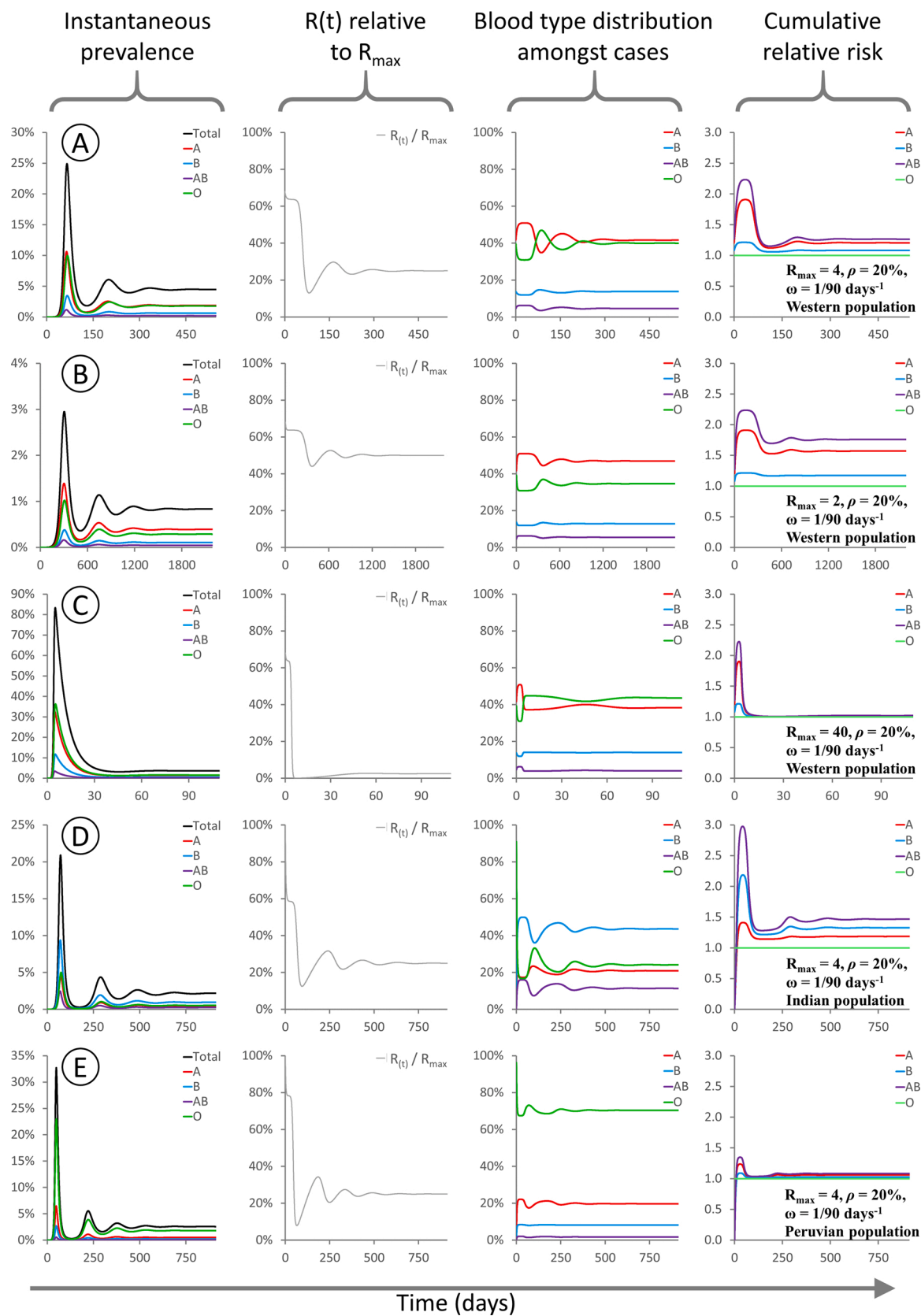

Fig. 4. X axis: days. Y axis: proportion of the population infected (note different scale used for panel A). All epidemics are based on a population with 50 % A / 50 % O individuals and initialised at t = 0 with type O index cases and 1/1,000,000 of the population infected. $R_{max} = 2$ and $\rho = 30$ % for all epidemics modelled. **A:** No vaccination. **B:** 35.4 % of the population vaccinated at t = 0, all type A. **C:** 35.4 % of the population vaccinated at t = 0, all type O. **D:** 35.4 % of the population vaccinated at t = 0, half type A and half type O. Vaccinating O is slightly more effective than vaccinating A, but vaccinating both types leads to

herd immunity at a lower threshold.

Instantaneous prevalence · R(t) relative to R_{max} · Blood type distribution amongst cases · Cumulative relative risk

A: $R_{max} = 4$, $\rho = 20\%$, $\omega = 1/90$ days^{-1}, Western population

B: $R_{max} = 2$, $\rho = 20\%$, $\omega = 1/90$ days^{-1}, Western population

C: $R_{max} = 40$, $\rho = 20\%$, $\omega = 1/90$ days^{-1}, Western population

D: $R_{max} = 4$, $\rho = 20\%$, $\omega = 1/90$ days^{-1}, Indian population

E: $R_{max} = 4$, $\rho = 20\%$, $\omega = 1/90$ days^{-1}, Peruvian population

Time (days)

(caption on next page)

Fig. 5. These panels show the evolution of various modelled parameters for epidemics involving waning immunity. Waning immunity is described by a parameter ω representing the rate of loss of immunity (see Methods for details). For all epidemics $\rho = 20$ %, and the model is initialized with type O index cases comprising 1/1,000,000 of the population. Since type O transmits freely to all recipients, $R(t) = R_{max}$ at time $t = 0$.
A: $R_{max} = 4$, $\rho = 20$ %, $\omega = 1/90$ days^{-1}, "Western" population blood type distribution, 38 % A / 14 % B / 4 % AB / 44 % O.
B: $R_{max} = 2$, $\rho = 20$ %, $\omega = 1/90$ days^{-1}, "Western" population blood type distribution as above.
C: $R_{max} = 40$, $\rho = 20$ %, $\omega = 1/90$ days^{-1}, "Western" population blood type distribution as above. **D:** $R_{max} = 4$, $\rho = 20$ %, $\omega = 1/90$ days^{-1}, "Indian" population blood type distribution, 21.4 % A / 39.9 % B / 9.4 % AB / 29.3 % O (Chandra and Gupta, 2012).
E: $R_{max} = 4$, $\rho = 20$ %, $\omega = 1/90$ days^{-1}, "Peruvian" population blood type distribution, 18.9 % A / 8.1 % B / 1.6 % AB / 71.4 % O (Anon., 2021b).
For all sub-panels, X axis denotes days. Y axes: (first column) proportion of the population infected; (second column) $R(t)$ as a fraction of R_{max}; (third column) distribution of blood types among currently infected individuals; (fourth column) cumulative risk of infection for each blood type relative to type O. The relative risks to different blood types during the final steady state depend on R_{max} (compare **A,B,C**) and on the background blood type distribution (compare **A,D,E**). The final case burden depends on all of R_{max}, ω, ρ and the background blood type distribution.

individuals, ABO antigen is expressed exclusively in red blood cells and the vascular endothelium and is not present in other cell types. Conversely, in "secretor" individuals with at least one functional FUT2 allele, ABO antigens will be present on many different types of epithelial cells including the lining of the respiratory and digestive tracts. In the context of the ABO-interference model described here, non-secretor individuals will still form anti-A and anti-B antibodies, and thus exhibit disease susceptibility according to their ABO blood type. However, they are unlikely to deposit A or B determinants on virions produced in lung cells, and so will transmit the virus freely, as if they were type O. Since around 20 % of individuals in Western countries are non-secretors (Nordgren et al., 2016), and the estimated value of ρ is ~40 % (see above), then non-secretor individuals likely account for around half the observed cross-type transmission.

Factors (b) and (c) are related and I shall consider them together. While titres vary, virtually all individuals carry at least some antibodies to non-self ABO glycans. These are generally acquired in the first few years of life following exposure to microbial surface glycans similar to A or B determinants (Cooling, 2015). Anti-A and anti-B antibodies are typically IgM and IgA subtypes, however IgG can also be seen in patients following heterologous transfusion and systemic exposure to antigen. This has implications for factor (c) in that secretory IgA is the predominant antibody type found in mucosal secretions, along with smaller amounts of secreted IgM (Cerutti et al., 2011). While the level of anti-ABO antibody in respiratory secretions has not been well studied, it is plausible that some amounts of both IgA and IgM may be present. These will not however be able to trigger complement-mediated inactivation of virus particles as the full complement cascade requires serum and is not present in the case of mucosal immunity. Anti-A/B antibodies may however block virus entry directly if the AB determinants are borne on the virus spike glycoprotein, as previously shown (Guillon et al., 2008; Preece et al., 2002; Deleers et al., 2020). Alternatively, IgA and IgM may agglutinate virus particles and trap them in the mucus barrier layer.

3.2. Implications of the estimated level of ABO-interference for public health strategy

If ABO-interference is the cause of the widely observed bias in SARS-CoV-2 infection rates among different blood types, then this model allows us to conclude that ABO incompatibility reduces SARS-CoV-2 transmission by at least 60 % and potentially more. This implies that the apparent R0 for most of the largest epidemics around the world has already been suppressed by at least ~25 %, and that R_{max} is likely to be substantially higher than the actually-measured R_{steady}. However, it is key to appreciate that no blood type is necessarily high- or low-risk: rather the nature of any protection is entirely context dependent. The presence of a diverse mix of blood types within any given community (i.e. within the pool of individuals that freely mix and may transmit the virus to each other) confers significant protection. Conversely, communities with limited blood type diversity have little or no inherent protection and will suffer disproportionately from SARS-CoV-2 infection. Herd immunisation threshold estimates derived from current data

may therefore substantially underestimate the level of vaccination required to protect vulnerable communities such as Native populations in both North and South America. In these, the type O frequency approaches 100 % (Ottensooser and Pasqualin, 1949), thus the true infectious potential of SARS-CoV-2 will be unmasked and the local R0 will tend towards R_{max}.

This heterogeneity in transmissibility means that in general, the risk to non-O and in particular type AB individuals in most countries will be higher than risk to type O individuals, while type O individuals are more infectious than non-O individuals. This may contribute to the marked overdispersion in transmission frequency for SARS-CoV-2 (Endo et al., 2020), and help explain why a small subset of patients are responsible for the majority of transmission events. If other polymorphic surface glycans (e.g. Lewis and P antigens) behave similarly, this will further magnify the differences between "super-spreaders" and "super-recipients".

Paradoxically, however, although in this model both disease vulnerability and infectiousness vary substantially between different blood types, it is important *not* to simplistically target vaccination on this basis. Rather, once a vaccine is available, care should be taken not to inadvertently destroy the existing blood type frequency structure that provides population-wise disease resistance, by ensuring good vaccine uptake among all communities. There is a danger that the growing public perception that "type O = low risk" will lead type O individuals to neglect or even refuse vaccination. If this tendency is not monitored and compensated for, it may disproportionately reduce the efficacy of public vaccination programs. Other types of blood-type-aware non-pharmaceutical interventions are not modelled here. If protective equipment is in limited supply, it may for example be appropriate for hospitals and care facilities to emphasise source control measures for type O "super-spreaders", and recipient protection measures for type A and AB "super-recipients". More sophisticated agent-based approaches will be needed to model this possibility (Ferguson et al., 2020)

3.3. Implications of this model for evolution of the ABO polymorphism

Irrespective of the detailed epidemiology of any disease, at the individual level type O alleles are always selectively favoured under this model, while type A and B alleles are subject to frequency-dependent selection. This predicts that - as seen across the globe - type O will have the highest allele frequency, while A and B alleles will be at lower frequency and more nearly similar to each other. Extending the SIR model to cover the case of waning immunity shows that the elevated risk to non-O blood types remains present even for endemic rather than epidemic disease, and thus long-term population morbidity and mortality from diseases subject to ABO-interference may be one factor affecting ABO allele population frequency, as previously modelled by Seymour et al. (Seymour et al., 2004).

Evolutionarily, this model provides an interesting case where individual and group advantages differ, since while O is always individually favoured, population-level disease resistance is optimised at relatively low type O frequencies. This tension between individual and group optima leads to a "tragedy of the commons" in which selection drives O

alleles to a higher frequency than the group optimum and leaves the resulting population more vulnerable to disease. Seymour et al. show that there must necessarily also be some countervailing selection pressure to prevent fixation of type O – they attribute this to a concomitant "arms race" with bacterial pathogens that use ABO antigens as adhesins and thus provide a constant balancing selection in favour of rare blood groups (Seymour et al., 2004). The net result is thus a contest between the bacterial interactions (which generally favour equal frequencies of all blood groups) and virus interactions (which always favour type O).

However, under this model it is not clear why type A and type B are not globally more equal in frequency. A recent preprint (Souilmi et al., 2021) may shed light on this debate, providing genetic evidence for past coronavirus-driven selective sweeps occurring specifically within East Asian populations. As well as indicating that coronavirus emergence in this region has been a recurrent threat over thousands of years, this regionally-contained selective pressure provides a plausible basis for the more balanced A:B frequencies found in these regions. More generally, it is likely that ABO frequency in the population will be locally determined by their interactions with the pathogens most prevalent in each region over evolutionary time scales.

3.4. ABO-interference and seasonal coronaviruses with waning immunity

For seasonal endemic coronaviruses, if these are subject to ABO-interference, this model predicts that that (in Western populations) type A and AB individuals should be more susceptible to infection, and type O less susceptible, but to a less pronounced degree than for SARS-CoV-2. It will therefore be interesting to determine whether non-O blood types are indeed more likely to suffer more frequent repeat endemic coronavirus infections. More speculatively, given the emerging evidence that COVID-19 is a multisystemic infection with particular impact on clotting pathways (Levi et al., 2020), could it be possible that the endemic coronaviruses may also have some effect on the likelihood of thrombosis? In addition to the known effect of ABO blood type on von Willebrand factor (vWF) levels (Franchini and Lippi, 2016), this could be an additional pro-thrombotic risk factor for non-O blood types.

3.5. Key observations and new experiments needed to test this model

In testing this model, the key experiment will be to directly determine whether A or B antigens are present on the virus envelope in *in vivo* infections, and whether A- or B-specific antisera can neutralise virus harvested from patients with the appropriate blood types. However, there are at least five other testable predictions from this model that may be addressable using existing epidemiological data:

(i) in countries with high type B frequency such as India, type B individuals should be at higher risk than type A individuals. This prediction has already been substantiated by two studies published during review of this modelling work (Rahim et al., 2020; Padhi et al., 2020)

(ii) in studies of super-spreading events, the index cases should be disproportionately type O individuals and/or non-secretor individuals.

(iii) Direct contact-tracing data should in general follow the blood transfusion rules. For in-family tracing, transmission between blood relatives - being more likely to share a blood type - may be more common than transmission between spouses, though this will be substantially confounded by age and behavioural effects.

(iv) In hotspot areas, once the overall case frequency exceeds approximately 20 %, the relative risks will begin to decline, as the epidemic proceeds through the type O population in a delayed manner. Longitudinal studies will be needed to address this, however this factor may explain the lack of any blood type correlation with susceptibility in high prevalence scenarios such as the *Charles de Gaulle* study (Boudin et al., 2020).

(v) In communities with high type O frequency, R_{steady} will be higher, the doubling time shorter and the relative risk to non-O blood types reduced when compared to otherwise similar communities with lower type O frequency. Similarly, communities with a highly skewed A:B ratio will have a higher R_{steady} and a shorter doubling time than communities with more nearly equal numbers of A and B individuals. This latter prediction has already been supported by two recent preprints (Liu et al., 2020; Miotto et al., 2020)

4. Materials and methods

4.1. An SIR model of ABO-interference with virus transmission

A standard SIR model of infection divides the population into three compartments, **S**usceptible, **I**nfectious, **R**ecovered, representing the proportion of individuals in the population with each status, and thus $S + I + R = 1$. These compartments are linked by three differential equations:

$$\frac{dS}{dt} = -\beta \cdot S \cdot I + \omega \cdot R$$

$$\frac{dI}{dt} = \beta \cdot S \cdot I - \nu \cdot I$$

$$\frac{dR}{dt} = \nu \cdot I - \omega \cdot R$$

The parameters β and ν represent the rate constants for infection and recovery, and their reciprocals $1 / \beta$ and $1 / \nu$ represent the average time required for one infectious individual to transmit to one susceptible individual, and the average duration of the infectious period. The ratio β / ν represents the basic reproductive number R0. The parameter ω represents the rate constant for loss of immunity, and thus transfer from the R compartment back to S. Its reciprocal $1 / \omega$ represents the average duration of immunity following recovery. In the analyses presented here, I extend this model by splitting each of S, I and R into four sub-compartments representing the four ABO blood types. These are then linked by a set of twelve equations:

$$\frac{dS_A}{dt} = -\beta \cdot S_A \cdot (I_A + \rho I_B + \rho I_{AB} + I_O) + \omega \cdot R_A$$

$$\frac{dS_B}{dt} = -\beta \cdot S_B \cdot (\rho I_A + I_B + \rho I_{AB} + I_O) + \omega \cdot R_B$$

$$\frac{dS_{AB}}{dt} = -\beta \cdot S_{AB} \cdot (I_A + I_B + I_{AB} + I_O) + \omega \cdot R_{AB}$$

$$\frac{dS_O}{dt} = -\beta \cdot S_O \cdot (\rho I_A + \rho I_B + \rho I_{AB} + I_O) + \omega \cdot R_O$$

$$\frac{dI_A}{dt} = \beta \cdot S_A \cdot (I_A + \rho I_B + \rho I_{AB} + I_O) - \nu \cdot I_A$$

$$\frac{dI_B}{dt} = \beta \cdot S_B \cdot (\rho I_A + I_B + \rho I_{AB} + I_O) - \nu \cdot I_B$$

$$\frac{dI_{AB}}{dt} = \beta \cdot S_{AB} \cdot (I_A + I_B + I_{AB} + I_O) - \nu \cdot I_{AB}$$

$$\frac{dI_O}{dt} = \beta \cdot S_O \cdot (\rho I_A + \rho I_B + \rho I_{AB} + I_O) - \nu \cdot I_O$$

$$\frac{dR_A}{dt} = \nu \cdot I_A - \omega \cdot R_A$$

$$\frac{dR_B}{dt} = \nu \cdot I_B - \omega \cdot R_B$$

$$\frac{dR_{AB}}{dt} = \nu \cdot I_{AB} - \omega \cdot R_{AB}$$

$$\frac{dR_O}{dt} = \nu \cdot I_O - \omega \cdot R_O$$

These equations were implemented as a Microsoft Excel spreadsheet (Supplementary Data File 1). In this model R0 is not a well-defined quantity since even in a fully susceptible population the current effective R value, denoted R(t), depends on ρ and on the blood type distributions in both S and I. In this paper, I use R_{max} to indicate the ratio β / ν, i.e. the R0 that would be observed in the absence of any ABO-interference. This is the maximal possible value of R(t), and would be observed if $\rho = 100$ %, or if all currently infected individuals are type O. For the work presented here, all epidemics were initiated by transferring 1/1,000,000 of the population from S to I at time t = 0. This for example represents an initial importation of ~9 infected index cases into a city the size of London. Varying these boundary conditions from 1/10,000 to 1/100,000,000 has no effect other than accelerating or retarding the initial progress of the epidemic (not shown). All index cases were assumed to be of blood type O. For all analyses except that presented in Fig. 5, immunity was assumed to be permanent and thus $\omega = 0$.

CRediT authorship contribution statement

Peter J.I. Ellis: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Project administration, Supervision, Resources.

Declaration of Competing Interest

The authors report no declarations of interest.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.epidem.2021.100446.

References

https://blog.23andme.com/23andme-research/blood-type-and-covid-19/.

Retrieved July 3, 2020, from https://rhnegativoperu.com/donaciones-de-sangre/porce ntajes-grupo-sanguineo/.

Arendrup, M., Hansen, J.E., Clausen, H., Nielsen, C., Mathiesen, L.R., Nielsen, J.O., 1991. Antibody to histo-blood group A antigen neutralizes HIV produced by lymphocytes from blood group A donors but not from blood group B or O donors. AIDS 5 (4), 441–444. Apr.

Boudin, L., Janvier, F., Bylicki, O., Dutasta, F., 2020. ABO blood groups are not associated with risk of acquiring the SARS-CoV-2 infection in young adults. Haematologica. Jul 23:haematol.2020.265066.

Cerutti, A., Chen, K., Chorny, A., 2011. Immunoglobulin responses at the mucosal interface. Annu. Rev. Immunol. 29, 273–293.

Chandra, T., Gupta, A., 2012. Frequency of ABO and rhesus blood groups in blood donors. Asian J. Transfus. Sci. 6 (1), 52–53. Jan.

Cheng, Y., Cheng, G., Chui, C.H., Lau, F.Y., Chan, P.K., Ng, M.H., Sung, J.J., Wong, R.S., 2005. ABO blood group and susceptibility to severe acute respiratory syndrome. JAMA 293 (12), 1450–1451. Mar 23.

Cooling, L., 2015. Blood groups in infection and host susceptibility. Clin. Microbiol. Rev. 28 (3), 801–870. Jul.

Deleers, M., Breiman, A., Daubie, V., Maggetto, C., Barreau, I., Besse, T., Clémenceau, B., Ruvoën-Clouet, N., Fils, J.F., Maillart, E., Doyen, V., Mahadeb, B., Jani, J.C., Van der

Linden, P., Cannie, M.M., Hayef, N., Corazza, F., Le Pendu, J., El Kenz, H., 2020. Covid-19 and blood groups: ABO antibody levels may also matter. Int. J. Infect. Dis. 104, 242–249. Dec 14.

Dzik, S., Eliason, K., Morris, E.B., Kaufman, R.M., North, C.M., 2020. COVID-19 and ABO blood groups. Transfusion. https://doi.org/10.1111/trf.15946. Jun 19.

Edinur, H.A., Dunn, P.P., Lea, R.A., Chambers, G.K., 2013. Molecular approaches to transfusion medicine in Polynesians and Maori in New Zealand. Int. J. Immunogenet. 40 (6), 460–470. Dec.

Ellinghaus, D., Degenhardt, F., Bujanda, L., et al., 2020. Genomewide association study of severe Covid-19 with respiratory failure. N. Engl. J. Med. Jun 17: NEJMoa2020283.

Endo, A., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott, S., et al., 2020. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Res. 5, 67.

Ferguson, N., Laydon, D., Nedjati Gilani, G., et al., 2020. Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and Healthcare Demand. March 27, 2020 (https://spiral.imperial.ac.uk:8443/handle/10044/1/774 82). Preprint.

Franchini, M., Lippi, G., 2016. Relative risks of thrombosis and bleeding in different ABO blood groups. Semin. Thromb. Hemost. 42 (2), 112–117. Mar.

Gérard, C., Maggipinto, G., Minon, J.M., 2020. COVID-19 and ABO blood group: another viewpoint. Br. J. Haematol. https://doi.org/10.1111/bjh.16884. May 26.

Guillon, P., Clément, M., Sébille, V., Rivain, J.G., Chou, C.F., Ruvoën-Clouet, N., Le Pendu, J., 2008. Inhibition of the interaction between the SARS-CoV spike protein and its cellular receptor by anti-histo-blood group antibodies. Glycobiology 18 (12), 1085–1093. Dec.

Jones, J.H., 2007. Notes on R0. Retrieved July 3, 2020, from. Stanford University, Stanford. https://web.stanford.edu/jhj1/teachingdocs/Jones-on-R0.pdf.

Kolin D., Kulm S., Elemento O. Clinical and Genetic Characteristics of Covid-19 Patients from UK Biobank. (https://www.medrxiv.org/content/10.1101/2020.05.05.200 75507v1). Preprint.

le Pendu, J., Breiman, A., Rocher, J., Dion, M., Ruvoën-Clouet, N., 2021. ABO blood types and COVID-19: spurious, anecdotal, or truly important relationships? A reasoned review of available data. Viruses 13 (2), 160. https://doi.org/10.3390/ v13020160. Jan 22.

Levi, M., Thachil, J., Iba, T., Levy, J.H., 2020. Coagulation abnormalities and thrombosis in patients with COVID-19. Lancet Haematol. 7 (6), e438–e440. Jun.

Liu, Y., Steinacker, J.M., Haeussinger, L., Dinse-Lambracht, A., 2020. Association Between Epidemic Dynamics of Covid-19 Infection and ABO Blood Group Types. July 15, 2020 (https://www.medrxiv.org/content/10.1101/2020.07.12.201 52074v1). Preprint.

Miotto, M., Di Rienzo, L., Gosti, G., Milanetti, E., Ruocco, G., 2020. Does Blood Type Affect the COVID-19 Infection Pattern? July 13, 2020 (https://arxiv.org/abs/2 007.06296). Preprint.

Neil, S.J., McKnight, A., Gustafsson, K., Weiss, R.A., 2005. HIV-1 incorporates ABO histo-blood group antigens that sensitize virions to complement-mediated inactivation. Blood 105 (12), 4693–4699. Jun 15.

Nordgren, J., Sharma, S., Kambhampati, A., Lopman, B., Svensson, L., 2016. Innate resistance and susceptibility to norovirus infection. PLoS Pathog. 12 (4), e1005385. Apr 26.

Ottensooser, F., Pasqualin, R., 1949. Blood types of Brazilian Indians (Matto grosso). Am. J. Hum. Genet. 1 (2), 141–155. Dec.

Padhi, S., Suvankar, S., Dash, D., Panda, V.K., Pati, A., Panigrahi, J., Panda, A.K., 2020. ABO blood group system is associated with COVID-19 mortality: an epidemiological investigation in the Indian population. Transfus. Clin. Biol. 27 (4), 253–258. Nov.

Percivalle, E., Cambiè, G., Cassaniti, I., et al., 2020. Prevalence of SARS-CoV-2 specific neutralising antibodies in blood donors from the Lodi Red Zone in Lombardy, Italy, as at 06 April 2020. Euro Surveill. 25 (24), 2001031. Jun.

Preece, A.F., Strahan, K.M., Devitt, J., Yamamoto, F., Gustafsson, K., 2002. Expression of ABO or related antigenic carbohydrates on viral envelopes leads to neutralization in the presence of serum containing specific natural antibodies and complement. Blood 99 (7), 2477–2482. Apr 1.

Rahim, F., Amin, S., Bahadur, S., Noor, M., Mahmood, A., Gul, H., 2020. ABO / Rh-D blood types and susceptibility to Corona Virus Disease-19 in Peshawar, Pakistan. Pak. J. Med. Sci. 37 (1).

Genevieve H.L. Roberts, Danny S. Park, Marie V. Coignet, Shannon R. McCurdy, Spencer C. Knight, Raghavendran Partha, Brooke Rhead, Miao Zhang, Nathan Berkowitz, AncestryDNA Science Team, Asher K. Haug Baltzell, Harendra Guturu, Ahna R. Girshick, Kristin A. Rand, Eurie L. Hong, Catherine A. Ball. AncestryDNA COVID-19 Host Genetic Study Identifies Three Novel Loci. (https://www.medrxiv.org/content/ 10.1101/2020.10.06.20205864v1). Preprint.

Seymour, R.M., Allan, M.J., Pomiankowski, A., Gustafsson, K., 2004. Evolution of the human ABO polymorphism by two complementary selective pressures. Proc. Biol. Sci. 271 (1543), 1065–1072. May 22.

Yassine Souilmi, M. Elise Lauterbur, Ray Tobler, Christian D. Huber, Angad S. Johar, David Enard. An ancient viral epidemic involving host coronavirus interacting genes more than 20,000 years ago in East Asia. (https://www.biorxiv.org/content/10. 1101/2020.11.16.385401v2). Preprint.

The COVID-19 Host Genetics Initiative, 2020. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur. J. Hum. Genet. 28, 715–718.

Watanabe, Y., Allen, J.D., Wrapp, D., McLellan, J.S., Crispin, M., 2020. Site-specific analysis of the SARS-CoV-2 glycan shield. Science, eabb9983. May 4.

Watkins, W.M., 1980. Biochemistry and Genetics of the ABO, Lewis, and P blood group systems. Adv. Hum. Genet. 10, 1-136, 379-85.

Zhao, J., Yang, Y., Huang, H., et al., 2020. Relationship Between the ABO Blood Group and the COVID-19 Susceptibility. March 27, 2020 (https://www.medrxiv.org/content/10.1101/2020.03.11.20031096v2). Preprint.

Zietz, M., Tatonetti, N.P., 2020. Testing the Association between Blood Type and COVID-19 Infection, Intubation, and Death. April 11, 2020 (https://www.medrxiv.org/content/10.1101/2020.04.08.20058073v1). Preprint.