Full Title
Modelling suggests blood group incompatibility may substantially reduce SARS-CoV-2 transmission

Short Title
ABO incompatibility and SARS-CoV-2

Authors
Peter J. I. Ellis [1*]

Affiliations

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

[*] corresponding author, email p.j.i.ellis@kent.ac.uk

Abstract

Several independent datasets suggest blood type A is over-represented and type O under-represented among COVID-19 patients. Here, I model a scenario in which ABO transfusion incompatibility reduces the chance of a patient transmitting the virus to an incompatible recipient. Comparison of model outputs to published data on COVID-19 prevalence indicates that if this scenario holds true, ABO incompatibility may reduce virus transmissibility by 60% or more. Paradoxically, however, targeted vaccination of either high-susceptibility type A 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 amongst the remaining susceptible individuals. I stress that these results illustrate a theoretical model of ABO blood group interaction with virus transmission and require confirmation by observation.

Introduction

Several 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 [1-5]. Puzzlingly, however, there is no difference in the severity of disease, with the case fatality ratio (CFR) and the probability of progressing to intensive care appearing independent of blood type. This discrepancy between incidence and severity data has led some authors to challenge the aforementioned findings [6]. Although not remarked on to date, in the majority of these studies type AB appears even more susceptible than type A. Thus, the relative risk of infection is AB > A > B > O, with type A and type B alleles functioning codominantly to increase risk. This is immediately reminiscent of the rules governing blood transfusion compatibility.

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 [7-8], and was also previously proposed for the 2003 epidemic of SARS [9-10], for measles [11], and indeed for enveloped viruses in general [12-13]. 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 [14-15]. The glycosylation status of the M and E proteins has not yet been characterised, nor the glycosylation status of the membrane lipids. Experimental work using pseudotyped virus suggests that transmission of the closely related SARS virus can be blocked by anti-A antibodies when virus particles are grown in an A-expressing cell line [10].

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.

Results

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 v, representing the rate constants for infection and recovery respectively. The duration of the infectious period d is given by $1 / v$. 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 $\beta . v$ 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 [16]. To this model I add a further parameter ρ where $0 \leq \rho < 100\%$, representing the relative probability of cross-type infection (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 of R(t). This will be observed when the population is fully susceptible and transmission is unimpeded, i.e. when ρ = 100% or all currently infected individuals are type O.

Type O individuals are critical for epidemic progression when ρ is low or zero

Figure 1 A-E illustrates the effect of complete ABO incompatibility on transmission dynamics in a typical Western European background with a population blood type distribution of 38% A / 14% B / 4% AB / 44% O, R_{max} = 4 and ρ = 0. If the index case is type A, an epidemic occurs within the A population, and also propagates to AB (since A→AB is a permissible vector of transmission). If the index case is type B or AB, the epidemic never becomes established since there are too few compatible susceptible individuals to sustain transmission. If the index case is type O, epidemics occur within all four populations. In all cases the progress of the epidemic is profoundly suppressed relative to an epidemic with no ABO-interference, however the epidemic stemming from a type O index case is four times the size of that stemming from a type A index case.

Importantly, if $\rho > 0$ and there is *any* degree of spread between incompatible blood types, at least some proportion of the O population will become infected. Thereafter the epidemic will behave similarly to one seeded by type O index cases, as shown in **Figure 1 F-I** for $\rho = 0.1\%$. If the index case is non-O [figure **1F/G/H**], this delays the onset of the epidemic and the date of the peak relative to a type O index case [figure **1I**], but affects neither the size of the peak nor how it progresses between and within different population compartments. For $\rho = 0.1\%$, the epidemic among B individuals is not sustainable from A→B cross-transmission alone but is dependent on the epidemic among O individuals. If these are removed from the susceptible population after 80 days (e.g. by vaccination), this immediately halts the spread among B individuals [**Figure 1J**]. The epidemics in A and AB are also affected, but to a more modest degree since the epidemic is self-sustaining amongst A individuals and these can then transmit to AB.

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

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 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}. **Figure 2** illustrates the nature of this steady state, which depends on ρ (compare rows **A,B,C**), but is independent of R_{max} (compare rows **C,D,E**) and the number and blood type of the index cases (not shown). R_{steady} also depends on the background population blood type distribution (compare rows **C,F,G**). Importantly, during this steady state, R_{steady} is a consistent multiple of R_{max}, indicating that ABO-interference suppressed the epidemic with equal efficiency regardless of the underlying infectiousness of the pathogen. Similarly, the relative risk to each blood type is also dependent only on ρ and the background population blood type distribution. Thus, regardless of whether the epidemic in any given region is progressing quickly or slowly, the relative risk to each blood type will be predictable from the population blood group frequencies and the specific value of ρ.

Estimating ρ for SARS-CoV-2

Since the worldwide SARS-CoV-2 pandemic appears to be still in its infancy, with low seroprevalence in all countries studied, the pandemic appears to fall within the steady-state region of the curve. **Table 1** shows data compiled from all published studies that have reported ABO frequencies amongst infected SARS-CoV-2 patients and controls [1,2,4]. These collectively cover epidemics in two regions of China, three regions of Spain, and one region each of Italy and the United States. A UK Biobank preliminary study has also reported an increased risk for type A individuals [3], but did not infer type AB frequency and was thus excluded from my analysis. 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 calculated the root mean square difference between the predictions and the observed case frequencies. Minimising this difference indicates the most likely value of ρ. Notably, for all data sets this lies between 25% and 40% despite the different underlying blood type frequencies and relative risk ratios in each country.

For the Italian and Spanish study, data was also provided for area-matched control data from blood donors [4]: using these instead of the internal control data does not affect the result.

Three caveats apply to this analysis. Firstly, while overall prevalence in each country remains low, the local prevalence in "hotspots" is much higher [19], and infected individuals will by definition originate preferentially from these hotspots. In these 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. Secondly, 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. Thirdly: 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 hospital patients. This will exaggerate the effects of ABO-interference if uninfected hospitalised patients are used as controls.

Overall, the third of these appears not to be an important factor since for both the Italian and Spanish populations similar results are obtained using blood donor controls. Since the first two effects both bias the estimate of ρ upwards, the central estimate of 35-40% thus represents a best-guess upper bound for ρ. Intriguingly, previous data from direct contact tracing [10] 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.

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 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. **Figure 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 ρ = 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 group frequencies yields **Figure 3B**. 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. A similar shape heat map is obtained for values of ρ = 20% and ρ = 90% (not shown). Vaccinating type O individuals moves the population upwards in **Figure 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, i.e. 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 [20]. 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 ρ = 30% then R_{steady} = 2.32, 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 [**Figure 4**].

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

At present is it 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, i.e. there is steady sustained transmission with overall infection levels neither growing nor shrinking [**Figure 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 extended ABO-interference model, the relative risk to non-O blood groups is less pronounced during this final steady state compared to the initial quasi-steady state during the early phase of pandemic spread, but remains non-zero. In the final steady 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 initial quasi-steady state is independent of R_{max} (see **Figure 2** above), the final steady state for an epidemic with waning immunity does depend on R_{max}. When R_{max} is high, the final relative risk for all blood groups is close to unity. When R_{max} is low, then the final relative risk for all blood groups is similar to that seen in the initial quasi-steady state. 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.

Discussion

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 [21]. "Non-secretor" individuals are homozygous for null mutations in the FUT2 gene. In these 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 are found in virtually all cell types. 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 group. 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 [22], and the estimated value of ρ is ~35-40% (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 [12]. 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 [23]. 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 [10]. Alternatively, IgA and IgM may agglutinate virus particles and trap them in the mucus barrier layer.

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 models 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% [24], 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 [25], 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 [26]

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 groups 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. Evolutionarily, this model provides an interesting case where individual and group advantages differ, since whole 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.

ABO-interference and seasonal coronaviruses with waning immunity

For seasonal 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 groups are indeed more likely to suffer more frequent repeat coronavirus infections. More speculatively, given the emerging evidence that COVID-19 is a multisystemic infection with particular impact on clotting pathways [27], could it be possible that the endemic coronaviruses may also have some effect on the likelihood of thrombosis? This could potentially explain the widely-observed phenomenon that type O individuals are on average slightly less likely to suffer from thrombosis and clotting-related myocardial infarction. Mechanistically, this is ascribed to an effect of ABO blood group on von Willebrand factor (vWF) levels, but the causal link by which ABO antigens regulate vWF levels remains obscure [28].

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, and whether A- or B-specific antisera can neutralise virus 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.

(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.

(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 [29, 30]

Materials and Methods

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 / β and 1 / ν represent the average time required for one infectious individual to transmit to one susceptible individual, and the average duration of the infectious period. The product $\beta \cdot \nu$ 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 / ω 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 subcompartments representing the four ABO blood group phenotypes. These are then linked by a set of twelve equations:

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

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

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

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

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

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

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

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

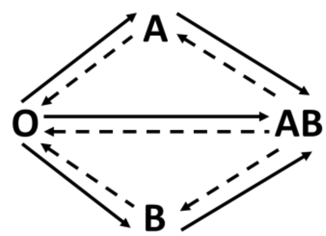
$$\frac{dR_A}{dt} = \nu \cdot I_A - \omega \cdot R_A$$

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

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

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

This can be graphically represented by the diagram on the right, with full arrows representing unimpeded transmission (infection rate = β) and dashed arrows representing impeded transmission (infection rate = $\rho \cdot \beta$).



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 product $\beta \cdot \nu$, 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 ρ = 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). Except where otherwise specified in the main text, all index cases were assumed to be of blood group O. For all analyses except that presented in Figure 5, immunity was assumed to be permanent and thus ω = 0.

References and Notes

1. Zhao J, Yang Y, Huang H, et al. 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.
2. Zietz M, Tatonetti NP. 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.
3. Kolin D, Kulm S, Elemento O. Clinical and Genetic Characteristics of Covid-19 Patients from UK Biobank. (https://www.medrxiv.org/content/10.1101/2020.05.05.20075507v1). Preprint.
4. Ellinghaus D, Degenhardt F, Bujanda L, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N Engl J Med. 2020 Jun 17:NEJMoa2020283.
5. Gérard C, Maggipinto G, Minon JM. COVID-19 and ABO blood group: another viewpoint. Br J Haematol. 2020 May 26:10.1111/bjh.16884.
6. Dzik S, Eliason K, Morris EB, Kaufman RM, North CM. COVID-19 and ABO blood groups. Transfusion. 2020 Jun 19:10.1111/trf.15946.
7. Neil SJ, McKnight A, Gustafsson K, Weiss RA. HIV-1 incorporates ABO histo-blood group antigens that sensitize virions to complement-mediated inactivation. Blood. 2005 Jun 15;105(12):4693-9.
8. Arendrup M, Hansen JE, Clausen H, Nielsen C, Mathiesen LR, Nielsen JO. 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. 1991 Apr;5(4):441-4.
9. Cheng Y, Cheng G, Chui CH, Lau FY, Chan PK, Ng MH, Sung JJ, Wong RS. ABO blood group and susceptibility to severe acute respiratory syndrome. JAMA. 2005 Mar 23;293(12):1450-1.

10. Guillon P, Clément M, Sébille V, Rivain JG, Chou CF, Ruvoën-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 Dec;18(12):1085-93.

11. Preece AF, Strahan KM, Devitt J, Yamamoto F, Gustafsson K. 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. 2002 Apr 1;99(7):2477-82.

12. Cooling L. Blood Groups in Infection and Host Susceptibility. Clin Microbiol Rev. 2015 Jul;28(3):801-70.

13. Seymour RM, Allan MJ, Pomiankowski A, Gustafsson K. Evolution of the human ABO polymorphism by two complementary selective pressures. Proc Biol Sci. 2004 May 22;271(1543):1065-72.

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

15. Focosi D. Anti-A Isohemagglutinin titers and SARS-CoV2 neutralization: implications for children and convalescent plasma selection. Br J Haematol. 2020 Jun 9:10.1111/bjh.16932.

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

17. Chandra T, Gupta A. Frequency of ABO and rhesus blood groups in blood donors. Asian J Transfus Sci. 2012 Jan;6(1):52-3.

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

19. Percivalle E, Cambiè G, Cassaniti I, et al. 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. 2020 Jun;25(24):2001031.

20. Edinur HA, Dunn PP, Lea RA, Chambers GK. Molecular approaches to transfusion medicine in Polynesians and Maori in New Zealand. Int J Immunogenet. 2013 Dec;40(6):460-70.

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

22. Nordgren J, Sharma S, Kambhampati A, Lopman B, Svensson L. Innate Resistance and Susceptibility to Norovirus Infection. PLoS Pathog. 2016 Apr 26;12(4):e1005385.

23. Cerutti A, Chen K, Chorny A. Immunoglobulin responses at the mucosal interface. Annu Rev Immunol. 2011;29:273-93.

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

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

26. Ferguson N, Laydon, D, Nedjati Gilani G, et al. 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/10044/1/77482). Preprint.

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

28. Franchini M, Lippi G. Relative Risks of Thrombosis and Bleeding in Different ABO Blood Groups. Semin Thromb Hemost. 2016 Mar;42(2):112-7.

29. Liu Y, Steinacker JM, Haeussinger L, Dinse-Lambracht A. Association between epidemic dynamics of Covid-19 infection and ABO blood group types. July 15, 2020 (https://www.medrxiv.org/content/10.1101/2020.07.12.20152074v1). Preprint.

30. Miotto M, Di Rienzo L, Gosti G, Milanetti E, Ruocco G. Does blood type affect the COVID-19 infection pattern? July 13, 2020 (https://arxiv.org/abs/2007.06296). Preprint.

Acknowledgments

I would like to thank Dr Rupert Beale, Dr Jeremy Rossman, Dr Mike Clark and Mr C.F. Helix for helpful discussions. I am funded by the UK Higher Education Funding Council for England (HEFCE), the Leverhulme Trust (RPG-2019-414 194) and the Eastern Academic Research Consortium. I have no competing interests to declare. All data and code used in the analysis are available in the manuscript and supplementary materials.

Figures and Tables

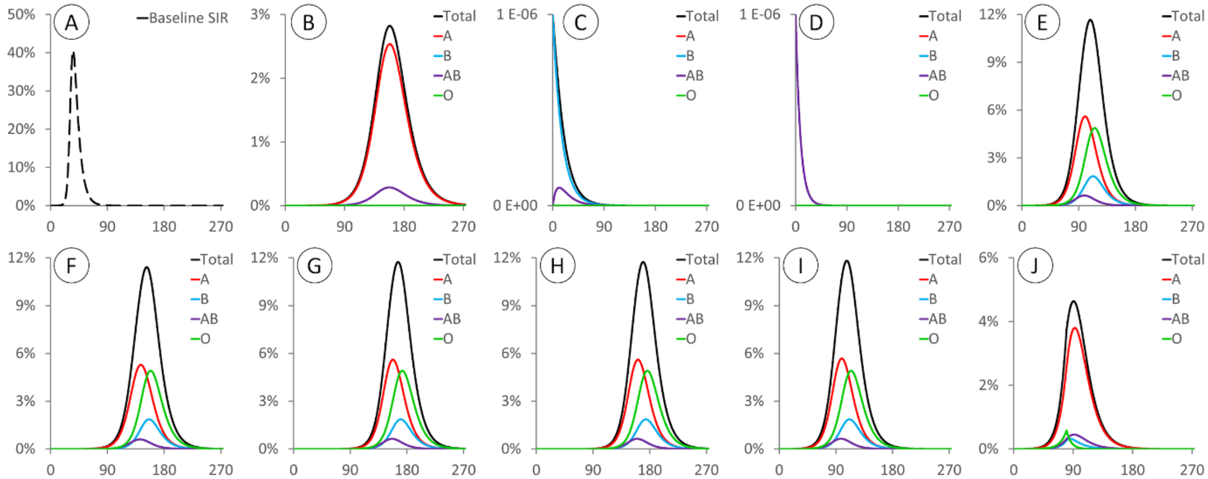


Figure 1

X axis: days. Y axis: proportion of the population infected (different scales used in each panel to show detail). All epidemics are based on a typical "Western" blood type distribution of 38% A / 14% B / 4% / AB / 44% O and initialised at t=0 with 1/1,000,000 of the population infected. **A:** Epidemic with R_{max} = 4 and no ABO-interference with transmission, peaking at 40.5% of the population on day 35. **B/C/D/E:** Epidemics with R_{max} = 4 and $\rho = 0$, with index cases of type A/B/AB/O respectively. **F/G/H/I:** Epidemics with R_{max} = 4 and ρ = 0.1%, with index cases of type A/B/AB/O respectively. **J:** Epidemics with R_{max} = 4 and ρ = 0.1%, with index cases of type B and all susceptible type O individuals assumed to be vaccinated at day 80.

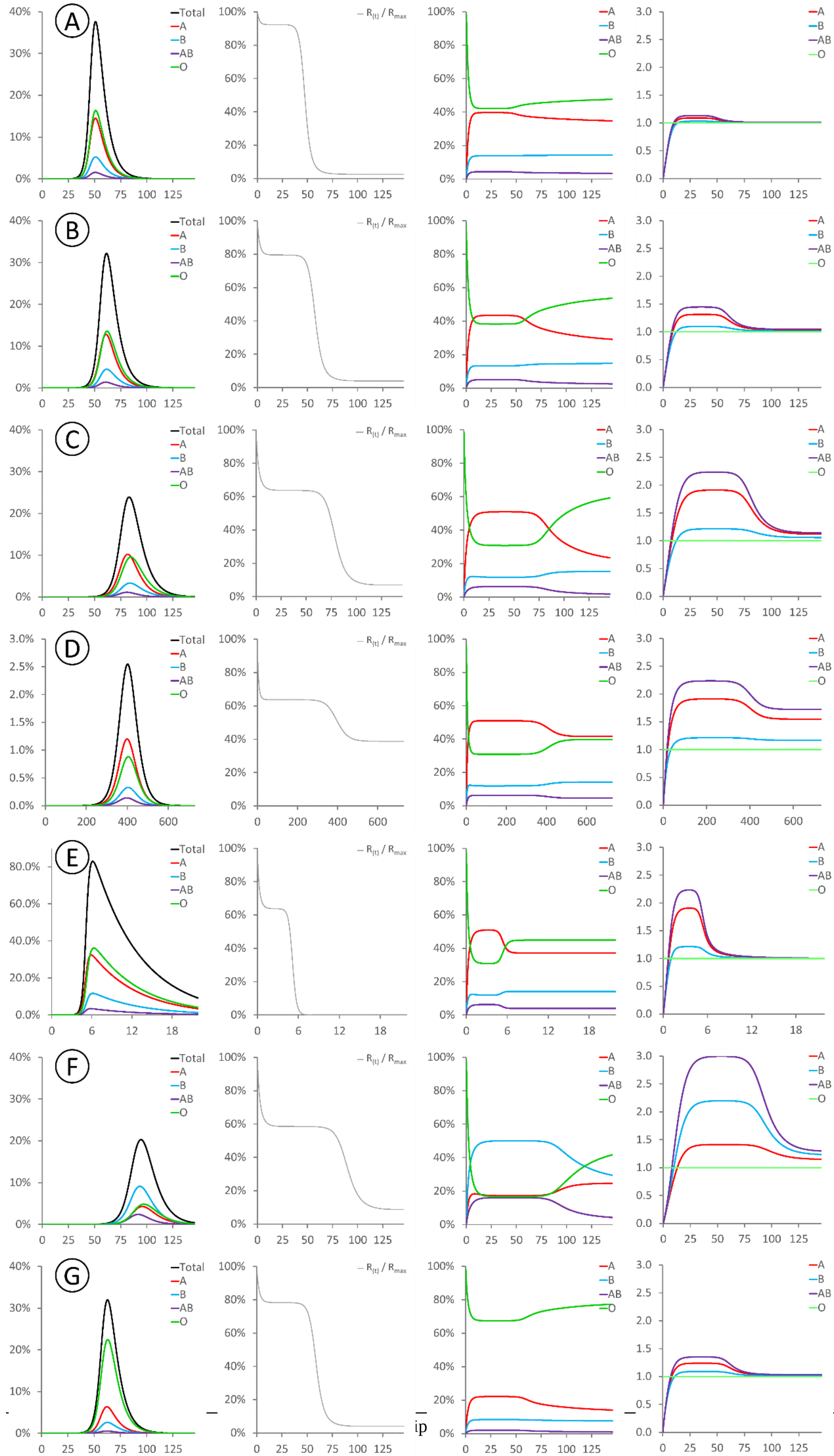

A
B
C
D
E
F
G
Total
A
B
AB
O
$R_{(t)}$ / R_{max}

Figure 2

These panels show the evolution of various modelled parameters for epidemics initialised under varying conditions. All epidemics were seeded at with type O index cases, and thus $R(t) = R_{max}$ at time t=0. 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.

Rows **A-E** are initialised with the same typical Western population blood type distribution used in Figure 1. **A:** $R_{max} = 4$, $\rho = 80\%$. **B:** $R_{max} = 4$, $\rho = 50\%$. **C:** $R_{max} = 4$, $\rho = 20\%$. **D:** $R_{max} = 2$, $\rho = 20\%$. **E:** $R_{max} = 40$, $\rho = 20\%$.

Row **F:** $R_{max} = 4$, $\rho = 20\%$, initialised with an Indian population blood type distribution of 21.4%/39.9%/9.4%/29.3% type A/B/AB/O [17].

Row **G:** $R_{max} = 4$, $\rho = 20\%$, initialised with a Peruvian population blood type distribution of 18.9%/8.1%/1.6%/71.4% type A/B/AB/O [18].

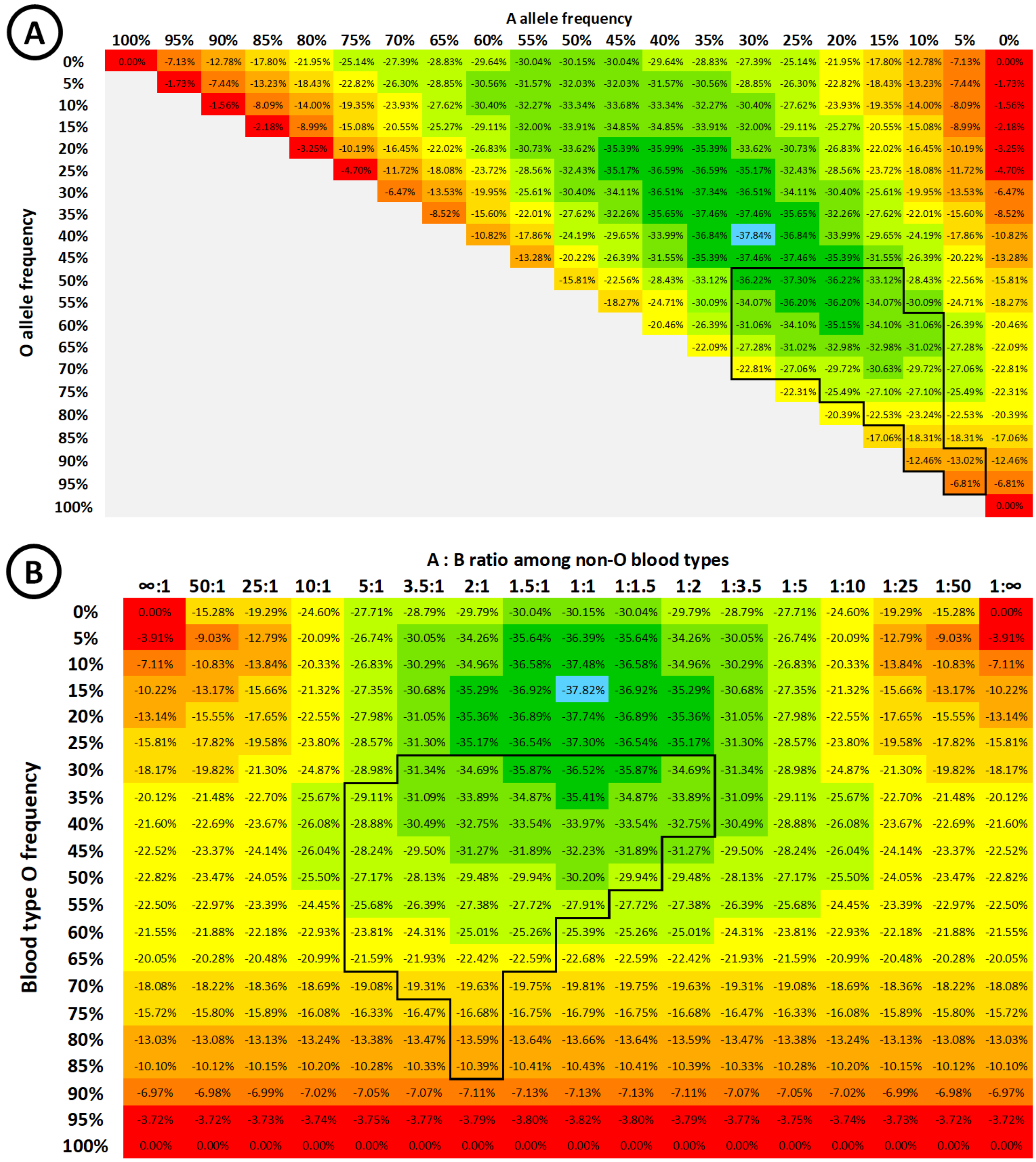


Figure 3

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 the epidemic, 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. 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.

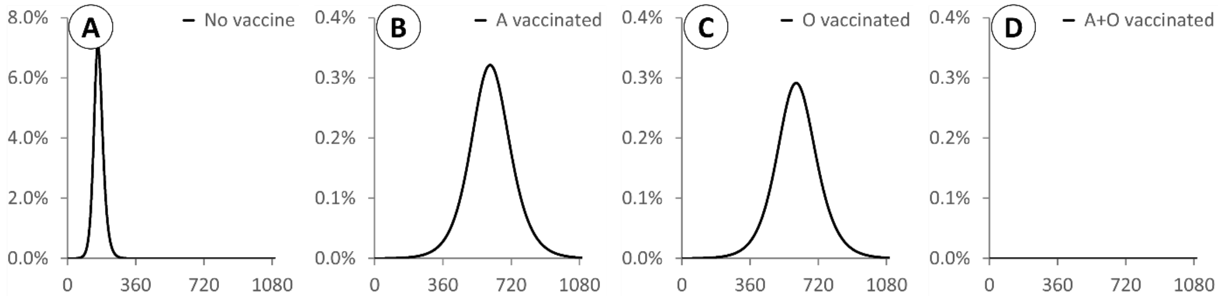


Figure 4

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 ρ = 30% for all epidemics. **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.

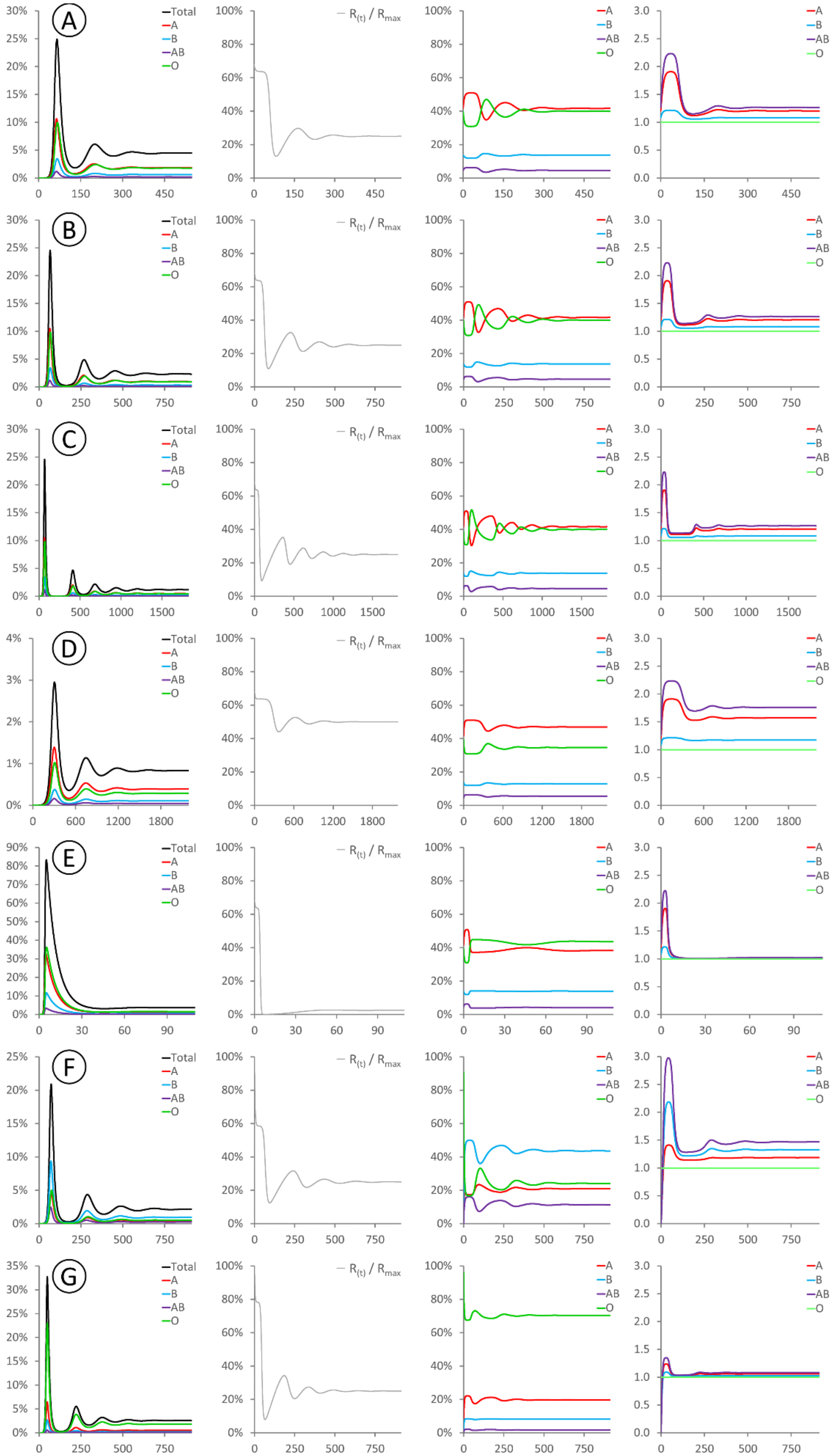

A
B
C
D
E
F
G
Total
A
B
AB
O
R(t) / Rmax

Figure 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). All epidemics were seeded at with type O index cases, and thus $R(t) = R_{max}$ at time t=0. 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.

Rows **A-E** are initialised with the same typical Western population blood type distribution used in Figures 1 and 2. **A:** $R_{max} = 4$, $\rho = 20\%$, $\omega = 1/90$ days^{-1}. **B:** $R_{max} = 4$, $\rho = 20\%$, $\omega = 1/180$ days^{-1}. **C:** $R_{max} = 4$, $\rho = 20\%$, $\omega = 1/360$ days^{-1}. **D:** $R_{max} = 2$, $\rho = 20\%$, $\omega = 1/90$ days^{-1}. **E:** $R_{max} = 40$, $\rho = 20\%$, $\omega = 1/90$ days^{-1}.

Row **F:** $R_{max} = 4$, $\rho = 20\%$, $\omega = 1/90$ days^{-1}, initialised with an Indian population blood type distribution of 21.4%/39.9%/9.4%/29.3% type A/B/AB/O [17].

Row **G:** $R_{max} = 4$, $\rho = 20\%$, $\omega = 1/90$ days^{-1}, initialised with a Peruvian population blood type distribution of 18.9%/8.1%/1.6%/71.4% type A/B/AB/O [18].

The relative risks to different blood groups during the final steady state are independent of ω (compare **A**,**B**,**C**), but depend on R_{max} (compare **B**,**D**,**E**) and on the background blood group distribution (compare **B**,**F**,**G**). The final case burden depends on all of R_{max}, ω, ρ and the background blood group distribution.

		Observed		Predicted case distribution for selected ρ value																			
		Controls	Cases	100%	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%	45%	40%	35%	30%	25%	20%	15%	10%	5%
Italy (Milan area)	A	35.94%	46.47%	35.94%	36.29%	36.67%	37.07%	37.50%	37.96%	38.46%	39.00%	39.58%	40.22%	40.93%	41.71%	42.58%	43.57%	44.71%	46.05%	47.66%	49.66%	52.29%	56.14%
(internal controls)	B	12.99%	10.90%	12.99%	12.96%	12.93%	12.90%	12.86%	12.82%	12.77%	12.72%	12.65%	12.58%	12.49%	12.39%	12.26%	12.12%	11.94%	11.71%	11.42%	11.04%	10.49%	9.61%
	AB	3.98%	5.15%	3.98%	4.06%	4.14%	4.22%	4.31%	4.40%	4.50%	4.60%	4.72%	4.84%	4.97%	5.12%	5.28%	5.45%	5.65%	5.88%	6.15%	6.47%	6.87%	7.42%
	O	47.09%	37.49%	47.09%	46.69%	46.26%	45.81%	45.33%	44.82%	44.27%	43.68%	43.05%	42.36%	41.61%	40.79%	39.88%	38.86%	37.70%	36.36%	34.77%	32.84%	30.35%	26.82%
	rms error vs observed cases			7.23%	6.96%	6.67%	6.37%	6.05%	5.71%	5.34%	4.94%	4.52%	4.05%	3.55%	2.99%	2.38%	1.72%	1.05%	0.81%	1.58%	2.90%	4.69%	7.31%
Spain (mixed areas)	A	41.89%	48.65%	41.89%	42.38%	42.89%	43.44%	44.02%	44.64%	45.31%	46.03%	46.81%	47.66%	48.59%	49.62%	50.76%	52.05%	53.52%	55.24%	57.29%	59.81%	63.13%	67.97%
(internal controls)	B	6.84%	9.16%	6.84%	6.80%	6.75%	6.70%	6.64%	6.58%	6.51%	6.44%	6.36%	6.27%	6.17%	6.06%	5.93%	5.78%	5.61%	5.41%	5.15%	4.84%	4.41%	3.76%
	AB	2.63%	4.65%	2.63%	2.67%	2.72%	2.77%	2.82%	2.87%	2.93%	2.99%	3.06%	3.13%	3.20%	3.29%	3.38%	3.48%	3.59%	3.72%	3.88%	4.06%	4.30%	4.64%
	O	48.63%	37.55%	48.63%	48.15%	47.64%	47.09%	46.52%	45.90%	45.24%	44.54%	43.77%	42.94%	42.04%	41.04%	39.93%	38.69%	37.27%	35.63%	33.68%	31.28%	28.17%	23.64%
	rms error vs observed cases			6.67%	6.35%	6.01%	5.65%	5.28%	4.89%	4.48%	4.06%	3.62%	3.19%	2.79%	2.48%	2.36%	2.53%	3.07%	3.94%	5.15%	6.76%	8.95%	12.21%
China (Wuhan)	A	32.16%	37.87%	32.16%	32.32%	32.49%	32.67%	32.86%	33.06%	33.29%	33.53%	33.79%	34.07%	34.39%	34.75%	35.15%	35.61%	36.15%	36.80%	37.61%	38.68%	40.25%	43.07%
(internal controls)	B	24.91%	26.17%	24.91%	24.93%	24.96%	24.99%	25.02%	25.04%	25.07%	25.09%	25.10%	25.11%	25.12%	25.11%	25.09%	25.06%	24.99%	24.87%	24.69%	24.36%	23.75%	22.31%
	AB	9.10%	10.22%	9.10%	9.30%	9.52%	9.75%	9.99%	10.26%	10.54%	10.85%	11.18%	11.54%	11.94%	12.38%	12.87%	13.43%	14.07%	14.82%	15.71%	16.83%	18.32%	20.55%
	O	33.84%	25.74%	33.84%	33.45%	33.03%	32.60%	32.13%	31.64%	31.11%	30.54%	29.93%	29.27%	28.55%	27.76%	26.89%	25.91%	24.80%	23.51%	21.99%	20.12%	17.68%	14.07%
	rms error vs observed cases			5.03%	4.81%	4.59%	4.35%	4.10%	3.84%	3.58%	3.30%	3.01%	2.73%	2.45%	2.21%	2.06%	2.04%	2.24%	2.69%	3.41%	4.45%	5.96%	8.44%
China (Shenzhen)	A	28.77%	28.77%	28.77%	28.91%	29.06%	29.22%	29.39%	29.58%	29.77%	29.98%	30.21%	30.45%	30.72%	31.02%	31.36%	31.73%	32.16%	32.67%	33.28%	34.04%	35.08%	36.69%
(internal controls)	B	25.14%	29.12%	25.14%	25.22%	25.30%	25.39%	25.47%	25.57%	25.66%	25.76%	25.87%	25.98%	26.10%	26.22%	26.35%	26.49%	26.63%	26.78%	26.94%	27.08%	27.20%	27.20%
	AB	7.32%	13.68%	7.32%	7.48%	7.65%	7.83%	8.02%	8.22%	8.44%	8.67%	8.93%	9.21%	9.51%	9.85%	10.22%	10.64%	11.12%	11.68%	12.34%	13.16%	14.23%	15.77%
	O	38.77%	28.42%	38.77%	38.39%	37.99%	37.56%	37.12%	36.64%	36.13%	35.58%	34.99%	34.36%	33.67%	32.91%	32.07%	31.14%	30.08%	28.87%	27.44%	25.71%	23.49%	20.35%
	rms error vs observed cases			6.39%	6.19%	5.97%	5.75%	5.51%	5.26%	5.00%	4.72%	4.43%	4.12%	3.81%	3.48%	3.15%	2.84%	2.60%	2.49%	2.64%	3.15%	4.13%	5.83%
US	A	27.94%	34.16%	27.94%	28.15%	28.38%	28.63%	28.88%	29.16%	29.45%	29.77%	30.11%	30.48%	30.88%	31.33%	31.82%	32.37%	32.99%	33.70%	34.53%	35.52%	36.76%	38.38%
(internal controls)	B	15.51%	17.01%	15.51%	15.53%	15.55%	15.57%	15.59%	15.61%	15.62%	15.64%	15.66%	15.67%	15.68%	15.68%	15.68%	15.68%	15.66%	15.63%	15.58%	15.51%	15.38%	15.17%
	AB	5.36%	3.08%	5.36%	5.46%	5.56%	5.67%	5.79%	5.91%	6.04%	6.18%	6.33%	6.49%	6.67%	6.86%	7.07%	7.30%	7.56%	7.84%	8.18%	8.57%	9.04%	9.63%
	O	51.20%	45.75%	51.20%	50.86%	50.51%	50.13%	49.74%	49.32%	48.88%	48.41%	47.90%	47.36%	46.77%	46.13%	45.43%	44.66%	43.80%	42.83%	41.71%	40.41%	38.83%	36.82%
	rms error vs observed cases			4.36%	4.19%	4.01%	3.83%	3.65%	3.46%	3.27%	3.08%	2.89%	2.72%	2.57%	2.46%	2.41%	2.45%	2.60%	2.89%	3.33%	3.96%	4.81%	5.99%
Italy (Milan area)	A	38.40%	46.47%	38.40%	38.78%	39.19%	39.62%	40.09%	40.59%	41.13%	41.71%	42.35%	43.05%	43.82%	44.68%	45.65%	46.75%	48.03%	49.54%	51.37%	53.70%	56.83%	61.66%
(local area donor controls	B	13.25%	10.90%	13.25%	13.21%	13.16%	13.11%	13.06%	12.99%	12.92%	12.84%	12.75%	12.64%	12.52%	12.37%	12.20%	12.00%	11.75%	11.44%	11.04%	10.50%	9.72%	8.43%
	AB	4.59%	5.15%	4.59%	4.68%	4.77%	4.87%	4.97%	5.08%	5.20%	5.33%	5.46%	5.61%	5.77%	5.94%	6.13%	6.35%	6.59%	6.86%	7.19%	7.59%	8.09%	8.82%
	O	43.76%	37.49%	43.76%	43.33%	42.88%	42.39%	41.88%	41.34%	40.75%	40.12%	39.44%	38.71%	37.90%	37.01%	36.02%	34.91%	33.64%	32.16%	30.40%	28.22%	25.36%	21.09%
	rms error vs observed cases			5.25%	4.97%	4.67%	4.36%	4.02%	3.67%	3.29%	2.89%	2.47%	2.03%	1.59%	1.25%	1.17%	1.53%	2.24%	3.20%	4.43%	6.00%	8.13%	11.39%
Spain (San Sebastien)	A	40.66%	49.17%	40.66%	41.13%	41.64%	42.18%	42.75%	43.35%	44.00%	44.70%	45.46%	46.28%	47.17%	48.15%	49.25%	50.47%	51.86%	53.46%	55.35%	57.66%	60.63%	64.80%
(Basque donor controls)	B	5.16%	6.31%	5.16%	5.12%	5.09%	5.05%	5.01%	4.96%	4.91%	4.86%	4.80%	4.74%	4.67%	4.59%	4.50%	4.40%	4.28%	4.14%	3.97%	3.76%	3.49%	3.10%
	AB	1.99%	3.65%	1.99%	2.03%	2.06%	2.09%	2.13%	2.17%	2.21%	2.25%	2.30%	2.35%	2.40%	2.46%	2.52%	2.59%	2.67%	2.77%	2.87%	3.00%	3.16%	3.39%
	O	52.19%	40.86%	52.19%	51.72%	51.21%	50.68%	50.12%	49.52%	48.88%	48.19%	47.44%	46.64%	45.76%	44.80%	43.73%	42.54%	41.19%	39.64%	37.80%	35.57%	32.72%	28.71%
	rms error vs observed cases			7.16%	6.83%	6.48%	6.11%	5.72%	5.31%	4.87%	4.41%	3.91%	3.39%	2.84%	2.29%	1.79%	1.52%	1.76%	2.52%	3.66%	5.17%	7.17%	10.03%
Spain (Barcelona)	A	42.03%	49.82%	42.03%	42.50%	42.99%	43.51%	44.07%	44.67%	45.31%	46.01%	46.77%	47.60%	48.52%	49.53%	50.67%	51.96%	53.45%	55.19%	57.30%	59.94%	63.47%	68.79%
(local area donor controls	B	9.65%	12.82%	9.65%	9.59%	9.53%	9.47%	9.39%	9.32%	9.23%	9.13%	9.02%	8.90%	8.77%	8.61%	8.43%	8.22%	7.97%	7.66%	7.29%	6.79%	6.11%	5.02%
	AB	3.76%	2.56%	3.76%	3.83%	3.90%	3.98%	4.06%	4.14%	4.23%	4.32%	4.43%	4.54%	4.66%	4.79%	4.93%	5.09%	5.28%	5.48%	5.73%	6.02%	6.40%	6.95%
	O	44.55%	34.80%	44.55%	44.08%	43.58%	43.05%	42.48%	41.88%	41.23%	40.53%	39.78%	38.96%	38.06%	37.07%	35.97%	34.73%	33.31%	31.66%	29.69%	27.24%	24.02%	19.24%
	rms error vs observed cases			6.47%	6.16%	5.84%	5.50%	5.15%	4.78%	4.40%	4.00%	3.60%	3.22%	2.88%	2.64%	2.60%	2.84%	3.40%	4.30%	5.54%	7.21%	9.52%	13.06%
Spain (Madrid)	A	41.18%	46.27%	41.18%	41.63%	42.11%	42.62%	43.17%	43.76%	44.39%	45.07%	45.82%	46.63%	47.52%	48.51%	49.62%	50.88%	52.32%	54.02%	56.06%	58.61%	62.00%	67.08%
(local area donor controls	B	9.55%	8.46%	9.55%	9.49%	9.44%	9.38%	9.31%	9.24%	9.16%	9.07%	8.97%	8.86%	8.73%	8.59%	8.42%	8.23%	7.99%	7.71%	7.37%	6.91%	6.28%	5.29%
	AB	3.55%	8.96%	3.55%	3.61%	3.68%	3.75%	3.82%	3.90%	3.98%	4.07%	4.17%	4.27%	4.39%	4.51%	4.64%	4.79%	4.96%	5.16%	5.38%	5.66%	6.01%	6.51%
	O	45.73%	36.32%	45.73%	45.26%	44.77%	44.25%	43.69%	43.10%	42.47%	41.78%	41.04%	40.24%	39.36%	38.39%	37.32%	36.10%	34.72%	33.11%	31.19%	28.82%	25.71%	21.11%
	rms error vs observed cases			6.02%	5.73%	5.42%	5.10%	4.77%	4.43%	4.08%	3.72%	3.38%	3.07%	2.82%	2.70%	2.78%	3.11%	3.72%	4.62%	5.83%	7.45%	9.66%	13.04%

Table 1

Observed case and control numbers are taken from references [1,2,4]. The Italian and Spanish data sets provided both internal and external controls. For each area, the control blood group frequencies were used to project the steady-state blood group frequencies among infected patient for varying values of ρ. Calculating the root mean square (rms) difference between the predicted frequencies and the actual case frequencies allows region-specific estimation of ρ (yellow highlight).

Supplementary Materials

Data file S1: Excel spreadsheet containing the implementation of the model described in the Methods.