

RESEARCH ARTICLE

“Hot-spotting” to improve vaccine allocation by harnessing digital contact tracing technology: An application of percolation theory

Mark D. Penney^{1,2*}, Yigit Yargic^{1,3}, Lee Smolin¹, Edward W. Thommes^{4,5,6}, Madhur Anand⁷, Chris T. Bauch²

1 Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada, **2** Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada, **3** Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada, **4** Vaccine Epidemiology and Modeling, Sanofi Pasteur, Toronto, Ontario, Canada, **5** Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada, **6** Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada, **7** School of Environmental Sciences, University of Guelph, Guelph, Ontario, Canada

* m2penney@uwaterloo.ca



OPEN ACCESS

Citation: Penney MD, Yargic Y, Smolin L, Thommes EW, Anand M, Bauch CT (2021) “Hot-spotting” to improve vaccine allocation by harnessing digital contact tracing technology: An application of percolation theory. *PLoS ONE* 16(9): e0256889. <https://doi.org/10.1371/journal.pone.0256889>

Editor: Martial L Ndeffo Mbah, Texas A&M University College Station, UNITED STATES

Received: March 19, 2021

Accepted: August 17, 2021

Published: September 22, 2021

Copyright: © 2021 Penney et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying the results presented in the study was collected and published by other researchers and is available at <https://doi.org/10.5281/zenodo.3746312>.

Funding: This work was supported by the NSERC COVID-19 Alliance Grant ALLRP 554380- 20 (MDP, MA, CTB) and the Ontario Ministry of Colleges and Universities COVID19 Rapid Response Grant C-299-137-BAUCH (MA, CTB). The NSERC COVID19 Alliance Grant was held in

Abstract

Vaccinating individuals with more exposure to others can be disproportionately effective, in theory, but identifying these individuals is difficult and has long prevented implementation of such strategies. Here, we propose how the technology underlying digital contact tracing could be harnessed to boost vaccine coverage among these individuals. In order to assess the impact of this “hot-spotting” proposal we model the spread of disease using percolation theory, a collection of analytical techniques from statistical physics. Furthermore, we introduce a novel measure which we call the efficiency, defined as the percentage decrease in the reproduction number per percentage of the population vaccinated. We find that optimal implementations of the proposal can achieve herd immunity with as little as half as many vaccine doses as a non-targeted strategy, and is attractive even for relatively low rates of app usage.

Introduction

Vaccine allocation decisions for pandemic infectious diseases such as COVID-19 must weigh a complex set of competing factors based on transmission characteristics [1], differences in vulnerability across demographic groups [2, 3], and the principle of reciprocity, which states that those who accept the greatest risks to mitigate the effects of the pandemic should be vaccinated first [4]. The initially limited supply of vaccines that is expected during a pandemic highlights the importance of these trade-offs, necessitating careful decisions about their use [5].

One particular trade-off that has been explored in the influenza modelling literature is whether to use the vaccine to protect vulnerable groups, or to reduce community transmission [6]. Recent models of SARS-CoV-2 transmission dynamics suggest that in certain parameter

partnership with Sanofi Pasteur who made only a contribution in-kind in the form of their employee (EWT). The funders provided support in the form of salaries for authors (MDP, EWT), but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the ‘author contributions’ section.

Competing interests: EWT is an employee of Sanofi Pasteur. All authors are inventors of United States Patent Application No. 63/122,587 derived from this work. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

regimes and for COVID-19 vaccines that have some effectiveness against both disease and infection, vaccinating to block transmission may prevent more deaths than targeting vulnerable age groups [1, 7, 8]. The strong potential effects of using vaccines to interrupt transmission, and the initially restricted supply of vaccines that is expected during a pandemic, motivate us to ask the following question: How can one best allocate a limited supply of vaccines in order to achieve the greatest reduction in disease transmission?

Allocation of vaccines is often based on age or other demographic factors, since we know that contact patterns differ between demographic groups. However, heterogeneity within demographic groups is much greater than the differences between them [9]. Empirical evidence from the ongoing COVID-19 pandemic has pointed to significant overdispersion in the number of secondary cases [10] which is not entirely explained by differing viral loads [11, 12]. Because individuals with more contacts (and, thus, exposure) have the potential to cause more transmission than individuals with low exposure to others, targeting these individuals for vaccination has been found to be highly effective in mathematical models [13–15]. But such strategies are difficult to implement in practice since public health authorities only have information on broad demographic or regional factors upon which to act.

It is exactly in this regard that digital contact tracing technologies have opened up a new avenue for public health interventions. The core functionality of Bluetooth-based exposure notification apps is the creation of an encounter log between app users. This encounter log is useful for more than just exposure notification: It also quantifies the user’s exposure to others. Said another way, these apps are also sensors that measure the duration of exposure—an epidemiologically significant quantity.

We propose that Bluetooth exposure notification apps can be harnessed to improve vaccine program effectiveness. These apps could be used to increase vaccine coverage among individuals with very high exposure rates, and thereby provide a highly efficient means to limit transmission in the population. We present mathematical modelling results in support of our proposal based on percolation theory: a collection of analytical techniques from statistical physics [16] that have been applied to material sciences [17], the spread of forest fires [18, 19], infectious diseases [20] and other areas. We assume a public health goal of interrupting transmission as efficiently as possible, and where vaccines can prevent transmission. The tools we developed show that the improved efficiency gained by piggy-backing vaccine allocation strategies on top of exposure notification apps is a robust phenomenon, as it derives its power from the very heterogeneities in contact patterns that shape the spread of infectious diseases.

Bluetooth exposure notification

In March 2020, COVID Watch released a white paper detailing an anonymous Bluetooth-based system that exploits the ubiquity of Android and iOS smartphones to support contact tracing [21]. The idea has seen widespread adoption, with nearly every developed country having incorporated it into their digital contact tracing solutions. Both the Google/Apple and BlueTrace frameworks are also based on this technology, the former of which is available throughout North America and the European Union and the latter in Singapore and Australia. Uptake has varied significantly between countries (Table 1), with most countries having made the usage of the app a personal choice. Singapore is a notable exception to this, having achieved high uptake rates due to mandatory use of the software.

Protecting users’ privacy was a fundamental principle underlying the design process. Every 10–20 minutes, anonymous tokens are exchanged between app users who are in close proximity to one another. Each user’s device stores these tokens in an encounter log. The data stored by the app is able to determine neither the number of contacts nor the duration of any

Table 1. Exposure notification app download rates in selected countries based on official sources.

Country	Downloads per 100 people
Canada [22]	23
Germany [23]	36
Italy [24]	19
New Zealand [25]	60
Singapore [26]	>90

<https://doi.org/10.1371/journal.pone.0256889.t001>

individual contact. Instead, the number of tokens logged over a given time frame is a measure of the user’s *total exposure time* to others: the sum, over each contact, of the duration of that contact.

Theoretical vaccination strategies that vaccinate individuals in accordance to their number of contacts have previously been studied in the context of immunizing scale-free networks [13]. Using exposure notification apps one can implement strategies which preferentially vaccinate individuals in accordance with their total exposure time. In what follows we will develop a model for more general vaccination strategies which depend on both the number and duration of contacts. We present this more general model not just because it is no more difficult, but also because such strategies could, with changes to the exposure notification systems, be implemented without the centralized collection of user data. Of course, any changes to the protocols come with concerns around privacy and data protection, and so for our simulations we restrict ourselves to considering a hot-spotting strategy that can be implemented without introducing any new privacy concerns.

The means by which the decentralized encounter log is used to facilitate contact tracing is by the central collection and communication of those tokens associated with individuals who have tested positive for COVID-19. In the Google/Apple framework public health authorities provide COVID-19 positive individuals with a code which, when voluntarily entered in the app, uploads their tokens to a central repository.

This same system can be used for ring vaccination strategies, which have been used in campaigns against smallpox [27] and Ebola [28]. However, there would have to be an important change in the type of information sent to the central repository from infected users. Instead of sending their own tokens, they would share the tokens they’ve collected from other users. With the same change, neighbor vaccination strategies [29] are also possible. Despite the fact that ring vaccination has already been studied using percolation theory [30, 31], and neighbor vaccination strategies certainly can in principle be, we restricted our attention to exposure-based strategies for the same reasons as in the preceding paragraph.

Weighted percolation theory

Many infectious diseases spread through close contact, and contact patterns in human populations display a high degree of heterogeneity [32, 33]. One successful approach to understanding the impact of heterogeneity is to model the spread of infectious disease as a percolation process on the network of contacts [34–36]. In a series of foundational works [37, 38], Newman developed analytical techniques based on *probability generating functions*. In this formalism, the information needed from the network is the degree distribution $\{p_k\}$, where p_k is the fraction of the individuals in the network which have k contacts or, equivalently, have degree k . Using the probability generating function of the degree distribution, $G_0(x) = \sum_k p_k x^k$, a standard technique for studying discrete probability distributions [39], Newman derived formulas for key epidemiological quantities in terms of G_0 and its derivatives.

More concretely, the infectious process is modelled as an instance of bond percolation. The disease spreads through occupied edges, and in Newman’s work each edge has a uniform probability T of being occupied. In the language of percolation theory, the basic reproduction number R_0 is the expected number of occupied edges attached to an infected vertex, with one edge removed from each infected vertex to account for the individual which infected the vertex in question. It was shown in [38] that this expectation value can be computed in terms of G_0 as

$$R_0 = T \frac{G_0'(1)}{G_0(1)} = T \left[\mu \left(1 + \frac{\sigma^2}{\mu^2} \right) - 1 \right]$$

where μ and σ are, respectively, the mean and standard deviation of the degree distribution. Let us briefly summarize the derivation for those readers who may be unfamiliar with these techniques.

First, the probability that a vertex of degree k has $n \leq k$ occupied edges attached to it is $\binom{k}{n} (1 - T)^{k-n} T^n$. This fact allows one to determine that the distribution of occupied degree in the network has probability generating function $G_0(x; T) = G_0(1 - T + Tx)$. The basic reproduction number is the expectation value of a closely related distribution. Namely, the expectation must be taken over only the infected vertices in the network. In Newman’s work it is assumed that the network of contacts is a random graph, an assumption implying that the desired distribution has probability generating function $G_1(x; T) = \frac{1}{\mu} G_0'(1 - T + Tx)$. A general property of probability generating functions is that the expected value of a distribution is simply the derivative of its generating function evaluated at $x = 1$. So, the basic reproduction number is calculated as $R_0 = G_1'(x; T)$, leading to the formula stated above.

In order to model vaccine allocation strategies based on exposure notification apps it is necessary to incorporate contact duration into the percolation model. To do this we consider the contact network to be weighted: each edge is further equipped with a weight w . The transmission probability T along an edge for the percolation process is assumed to depend on that edge’s weight, so that the single parameter is replaced with a distinct T_w for each weight w .

For our purposes in this paper the weight represents the number of time-steps over which the contact took place. If one assumes that in each time-step there is an independent probability T_1 of transmission, the transmission probability after w time-steps is $T_w = 1 - (1 - T_1)^w$.

However, the weights could represent any factor which influences the transmission probability along an edge, such as the nature of the contact, the setting, or the presence or absence of PPE. Unless stated otherwise, our analytical results hold for this general interpretation of weights.

In a weighted network we represent the configuration of edges around each vertex by a *generalized degree*, denoted \mathbf{k} . This is a vector having an entry for each of the possible weights appearing in the network. For a vertex of generalized degree \mathbf{k} , the entry corresponding to weight w , denoted k_w , is the number of contacts having weight w .

To generalize Newman’s results to the weighted setting we introduce a *multivariable generating function* for the distribution of generalized degrees. Since we are keeping track of the number of edges of each weight separately, the generating function needs a variable y_w for each weight w appearing in the network. Let $q_{\mathbf{k}}$ denote the fraction of vertices in the network having generalized degree \mathbf{k} and let \mathbf{y} denote the vector of variables y_w . Then the multivariable generating function of the network is

$$Q(\mathbf{y}) = \sum_{\mathbf{k}} q_{\mathbf{k}} \mathbf{y}^{\mathbf{k}}, \quad \mathbf{y}^{\mathbf{k}} = \prod_w y_w^{k_w}.$$

We derive the basic reproduction number using the exact same approach as Newman, except starting from the multivariable generating function. Specifically, the distribution of occupied degree has probability generating function,

$$G_0(x; \mathbf{T}) = Q(\mathbf{1} - \mathbf{T} + \mathbf{T}x) ,$$

where \mathbf{T} is the vector composed of the transmission probabilities T_w . Furthermore, we introduce

$$G_1(x; \mathbf{T}) = \frac{G'_0(x; \mathbf{T})}{G'_0(1; \mathbf{T})}$$

as the probability generating function for the number of other occupied edges attached to a vertex reached by following a random occupied edge (cf. [38]). By definition, the basic reproduction number R_0 is precisely the mean of this final distribution, i.e. $R_0 = G'_1(1; \mathbf{T})$.

One can express R_0 directly in terms of the generating function $Q(\mathbf{y})$. This is accomplished by introducing the differential operator $\nabla_{\mathbf{T}}$, defined as

$$\nabla_{\mathbf{T}} Q(\mathbf{y}) = \sum_w T_w \frac{\partial Q(\mathbf{y})}{\partial y_w} .$$

Then, we find

$$R_0 = \frac{\nabla_{\mathbf{T}}^2 Q(\mathbf{1})}{\nabla_{\mathbf{T}} Q(\mathbf{1})} .$$

Alternatively, this result can be rewritten in a form which is more amenable to computation,

$$R_0 = \frac{\text{Av}[(\sum_i T_{w(i)})^2] - \text{Av}[\sum_i T_{w(i)}^2]}{\text{Av}[\sum_i T_{w(i)}]} .$$

In this equation, the averages $\text{Av}[\dots]$ are taken over all vertices in the network and the sums \sum_i are taken over all contacts i of a vertex, with $w(i)$ being the weight of that contact. Note that this formula holds regardless of the interpretation of the weights as contact duration.

Modelling vaccination

We model vaccination as a stochastic process wherein a vertex of generalized degree \mathbf{k} has a probability $\nu(\mathbf{k})$ of being vaccinated. Vaccination modifies the original contact network by *removing vaccinated vertices*, as we assume that a vaccinated individual can neither become infected nor infect others. The model can readily incorporate imperfect efficacy and results in a simple rescaling of the main results shown below. Indeed, in the language of percolation theory vaccination is an instance of site percolation, and the combined spread of a disease and vaccination is modeled as a mix of bond and site percolation.

The formulas for the basic reproduction number R_0 can be adapted to give the expected post-vaccination reproduction number. This adaptation requires two changes: Firstly, each vertex of generalized degree \mathbf{k} has a probability $1 - \nu(\mathbf{k})$ of remaining in the network, that is, of being unvaccinated. Secondly, the changes to the generalized degree of the remaining unvaccinated individuals must be taken into account. Specifically, the probability that a weight w edge

leads to an unvaccinated vertex is,

$$\phi_w = 1 - \frac{\text{Av}[k_w v(\mathbf{k})]}{\text{Av}[k_w]} .$$

Hence, the expected post-vaccination reproduction rate is given by

$$\mathbb{E}[R^v] = \frac{\text{Av}[(1 - v(\mathbf{k}))((\sum_i \tilde{T}_{w(i)})^2 - \sum_i \tilde{T}_{w(i)}^2)]}{\text{Av}[(1 - v(\mathbf{k}))\sum_i \tilde{T}_{w(i)}]} ,$$

where $\tilde{T}_w = T_w \phi_w$, and the averages $\text{Av}[\cdot \cdot \cdot]$ on the right-hand side run through all vertices in the network.

To compare two vaccine strategies in a context of limited supply we introduce a novel measure of the *efficiency* of a vaccination strategy:

$$E^v = \frac{1 - R^v/R_0}{V} ,$$

where R^v is the post-vaccination reproduction number and V is the fraction of the population that receives the vaccine. In other words, the efficiency of a strategy is the percentage decrease in the reproduction number per percentage of the population vaccinated.

In a stochastic process of vaccination, V , R^v and E^v are all random variables. In particular, the vaccine coverage rate V has the expected value $\mathbb{E}[V] = \text{Av}[v(\mathbf{k})]$ and the variance $\text{Var}[V] = \frac{1}{N} \text{Av}[v(\mathbf{k})(1 - v(\mathbf{k}))]$, where N is the size of the population. Therefore, for computing the expectation value of the efficiency E^v in a large population, we may neglect the variance in V and treat it as a fixed number at its mean value. Hence, we can write

$$\mathbb{E}[E^v] = \frac{1 - \mathbb{E}[R^v]/R_0}{\mathbb{E}[V]} .$$

One can compare any vaccination strategy to the uniform one, under which vaccines are distributed uniformly across the population to achieve a target vaccine coverage without taking into account any form of contact or exposure heterogeneity. For a uniform vaccination probability v , one would have $\mathbb{E}[R^v] = (1 - v)R_0$ and $\mathbb{E}[E^v] = 1$. Strategies for which the efficiency is less than 1 are regarded as inefficient, since the same number of vaccines could have achieved a greater impact if they were allocated uniformly. On the other hand, strategies with efficiency greater than 1 are promising candidates for a significant impact.

The hot-spotting strategy

Here we propose a “hot-spotting” vaccination strategy, in reference to a fire-fighting practice that focuses on areas with intense fires [40]. This strategy aims to increase vaccine coverage among individuals with higher total exposure time, as determined by the Bluetooth apps. It operates without the central collection of any user data, as the selection of each user is decided locally by their device.

The hot-spotting strategy depends on a parameter β encoding the probability of success in a weighted coin-flip. Each user performs a virtual coin flip for each encounter stored in their encounter log over a fixed time period. The app selects those users who receive at least one success. Since the probability of obtaining at least one success on n weighted coin flips is $1 - (1 - \beta)^n$, we see that users with a greater number of recorded encounters have a greater probability of being selected. Note that the extent to which high total exposure time individuals are preferentially selected can be increased by requiring a greater number of successes.

In order to estimate the impact and efficiency of our hot-spotting strategy on the whole population, we must take into account that not every person in the population uses the app in question. To this end, we assume that there is an app usage rate of U and that the app users are homogeneously distributed in the population. We also assume that there is no preferential attachment between app users. This implies that for a random individual with the contact structure \mathbf{k} , which is not necessarily registered in the app network, the probability for them being an app user and also being selected by the app is given by

$$v(\mathbf{k}; \beta, U) = U \left(1 - \prod_w \gamma_w^{k_w} \right),$$

where

$$\gamma_w = 1 - U + U(1 - \beta)^w.$$

We simulate an implementation of hot-spotting wherein individuals receive a vaccine if and only if they have been selected by the app, although the results are qualitatively similar if the app only increases the likelihood of vaccine acceptance among these individuals.

Methods

Model network

The model weighted contact network is based on contact data collected in the well-known POLYMOD study that surveyed over 7000 individuals across 8 European countries (Belgium, Germany, Finland, Great Britain, Italy, Luxembourg, the Netherlands, and Poland) between May 2005 and September 2006 [9]. Respondents kept a log of all contacts made on a single day noting, among other features, how long the contact lasted and how frequently that contact is made. This data captures normal contact patterns and doesn't reflect any changes which can occur in response to an ongoing pandemic.

In the language of time-weighted networks, the inclusion of duration data means that the survey responses sample from the generalized degree distribution of the daily contact network. The network should capture the contacts made over the typical infectious period of the disease being modelled. For simplicity we choose a period of 14 days, as it aligns well with the frequency responses in the survey.

Using the daily contact data we must generate samples from the generalized degree distribution of the *fortnightly contact network*. We accomplish this by a bootstrapping technique. More precisely, for each contact recorded, respondents chose between 5 options concerning both the duration and the frequency of that contact as shown in Table 2. We can therefore represent each respondent's contacts on that day in a 5×5 matrix whose (i, j) entry is the number of contacts recorded with frequency key i and duration key j . For respondent n we denote this matrix by D_n .

Table 2. Possible responses on POLYMOD survey concerning frequency and duration of contact.

Keys	Frequency	Duration
(1)	daily	< 5 mins
(2)	1 – 2 times per week	5 – 15 mins
(3)	1 – 2 times per month	15 – 60 mins
(4)	less than once a month	1 – 4 hrs
(5)	first time	> 4 hrs

<https://doi.org/10.1371/journal.pone.0256889.t002>

Our goal is to extrapolate fortnightly contact matrices F_n from the daily contact matrices D_n . In our bootstrapping procedure, it is important that we distinguish between daily repeating contacts, which have frequency key (1), and *infrequent contacts*, which have frequency keys (2)-(5). We denote by I_n the matrix of infrequent contacts for respondent n . The matrix I_n is obtained from the same respondent's D_n by setting the first row of daily repeating contacts to 0.

The matrices F_n are created by sampling from the daily contact matrices. Each sample is generated as follows:

1. Sample 14 respondents, n_1, \dots, n_{14} .
2. Set $F'_n = D_{n_1} + I_{n_2} + \dots + I_{n_{14}}$.
3. Produce F_n from F'_n by dividing each entry in the second row by 3 and rounding it to the nearest integer.

In words, we sample 14 daily contact logs from the survey. Then, we add together the 1-day duration-frequency matrices for each of the samples, excluding the daily repeating contacts from all but the first sample. Then, assuming that a frequency-key-(2) contact is seen 3 times in a fortnight, we divide the second row by three to account for repeated counting of the same contact.

Finally, we assign to each type of contact in the matrix F_n a certain weight. Each unit of weight is approximately 10 minutes of contact time, which corresponds to the token exchange rate of the exposure notification apps. We chose these weights to be given as in the following weight matrix:

$$W = \begin{bmatrix} 0 & 12 & 36 & 120 & 480 \\ 0 & 3 & 9 & 30 & 120 \\ 0 & 1 & 3 & 10 & 40 \\ 0 & 1 & 3 & 10 & 40 \\ 0 & 1 & 3 & 10 & 40 \end{bmatrix}$$

Hence, we obtain a list of generalized degrees sampled from a virtual weighted contact network, in which the set of possible weights is $\{1, 3, 9, 10, 12, 30, 36, 40, 120, 480\}$.

Simulation

The simulations are based on 500 samples of fortnightly generalized degree distributions generated through the bootstrapping procedure described above. We interpret these 500 samples as an *observed cohort* of 500 individuals within a much larger population. As such, the recorded contacts are assumed to lie outside the observed cohort.

For these simulations we set the unit transmission probability at $T_1 = 0.000375$ yielding a basic reproduction number (in the absence of vaccination) of $R_0 = 1.501$. This value of T_1 was chosen as the corresponding value of R_0 is on the lower range of R_0 estimates for the 2009 influenza A/H1N1 pandemic in the USA [41]. The expected efficiency is, in fact, quite insensitive to the value of T_1 , with the change of τ in T_1 leading to a change in expected efficiency of $\mathcal{O}(\tau^2)$.

We select a set of parameters β for the vaccination probability function $v(\mathbf{k}; \beta, U)$ of the hot-spotting strategy (and a set of parameters ν for the uniform strategy) to target a homogeneous

distribution of the vaccinate rate V in the outcome over its range of possible values. 10,000 simulations are performed for each selected parameter. This accounts to a total of 990,000 simulations for the strategies with $U = 100\%$. For the strategies with $U = 20\%$, 40% , and 60% , we ran a total of 190,000, 390,000, and 590,000 simulations, respectively.

Each simulation consists of four steps:

1. The vaccination probabilities $v(k; \beta, U)$ for each individual, and the probabilities φ_w for each weight are calculated for the given parameters.
2. For each component k_w of the generalized degree of each individual, a binomial process is performed with the probability φ_w . The outcome of this process replaces k_w as the residual contact number.
3. For each individual, a Bernoulli trial is performed with the probability $v(k; \beta, U)$. If the outcome is a success, the individual is removed from the list with all their contacts, otherwise they remain.
4. The base reproduction number is computed in the residual list and saved as the post-vaccination reproduction number R^v . The vaccine coverage V is deduced from the length of the residual list. The efficiency E^v is calculated from R^v and V .

Results

Our simulation results show that the hot-spotting strategy is one to four times more efficient than the uniform vaccination strategy (Fig 1). The difference in efficiency between the two strategies is most pronounced when the vaccine coverage V is small, since in this case, high exposure individuals are likely to be vaccinated first under a hot-spotting strategy. As the vaccine coverage increases, hot-spotting becomes indistinguishable from the uniform strategy and its efficiency approaches 1. Indeed, when the vaccine coverage equals the app usage rate all app users are vaccinated. By our model assumptions, the app using population have the same contact patterns as the general population, and so vaccinating all app users is equivalent to uniformly vaccinating the same fraction of the population.

We also find that hot-spotting retains its lead in efficiency across a wide range of values for U , the rate of app usage in the population (Fig 1): vaccinating individuals with more exposure always provides more population protection for fewer vaccines expended, even if public health cannot reach all of the high-exposure individuals. The success of digital contact tracing technologies has been hampered by inadequate uptake rates [42, 43]. Our simulations show that our proposal doesn't suffer from this issue, at least from the perspective of efficient use.

These trends are also reflected in the impact of the two strategies on the post-vaccination reproduction number (Fig 2). Under the uniform strategy, the post-vaccine reproduction number declines in direct proportion to the vaccine coverage. But for the hot-spotting strategy, the post-vaccine reproduction number declines very rapidly for very low vaccine coverage before eventually converging to the same level as the uniform strategy, when enough individuals are vaccinated. The greatest relative impact is achieved by vaccinating 20-40% of the app users. Finally, when app usage rates are high enough, hot-spotting achieves herd immunity with fewer than half as many doses. In Fig 3 we show that the significantly lower herd immunity thresholds persist at a higher R_0 value of 2.2, which is in the middle range of estimates for the 2009 influenza A/H1N1 pandemic in the USA [41].

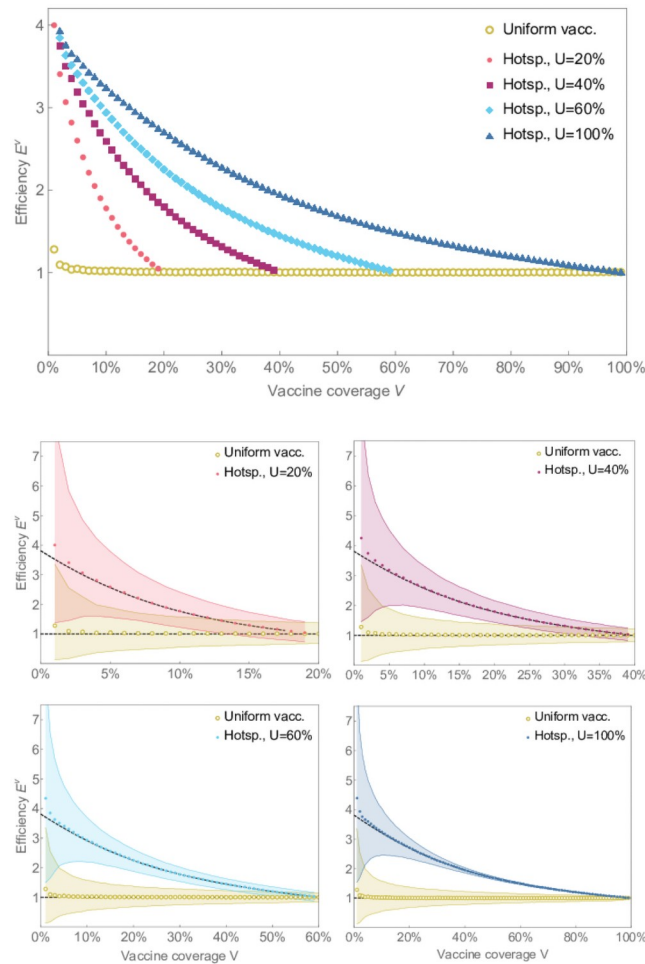


Fig 1. Efficiency of hot-spotting strategy. The efficiency for the hot-spotting strategy for $R_0 = 1.5$ with various app usage rates are compared with the uniform vaccination in relation to vaccine coverage. The dots represent the mean efficiency obtained from 10,000 simulations for each dot. The shaded regions show the intervals that capture 90% of all simulation results. The dashed lines in the bottom figures show the expected efficiency obtained analytically.

<https://doi.org/10.1371/journal.pone.0256889.g001>

In general, the necessary vaccine coverage V_H required to reach herd immunity is given by the solution of the equation

$$V_H = \frac{1}{E^v(V_H)} \left(1 - \frac{1}{R_0} \right).$$

For a vaccine allocation strategy with $E^v > 1$, this equation shows that V_H can be reduced through such strategy by a factor of $E^v(V_H)$.

App-using subnetwork

In our model we assumed that app users are uniformly distributed in the network and that there is no preferential attachment. The actual contact patterns of app users need not satisfy this condition in a given population and can have meaningful impacts on the efficiency of the hot-spotting strategy. For example, if very few high exposure individuals are app users the hot-spotting strategy might be less efficient than the uniform strategy. Or if there is very little mixing between the app using and non-app using populations.

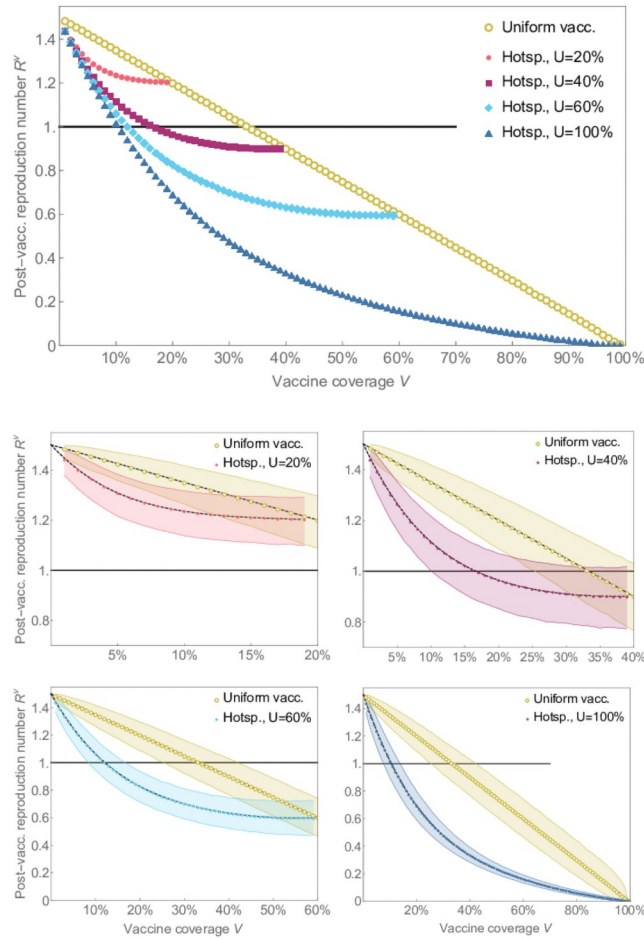


Fig 2. The post-vaccination reproduction numbers for the hot-spotting strategy. The post-vaccination reproduction numbers for the hot-spotting strategy for $R_0 = 1.5$ with various app usage rates are compared with the uniform vaccination in relation to vaccine coverage. The dots represent the mean reproduction number obtained from 10,000 simulations for each dot. The shaded regions show the intervals that capture 90% of all simulation results. The dashed lines in the bottom figures show the expected post-vaccination reproduction numbers obtained analytically.

<https://doi.org/10.1371/journal.pone.0256889.g002>

In order to model non-uniformly distributed app-using populations one can make use of mixing matrices (cf. [44]). More specifically, one needs the following information about the network: The distributions of generalized degree for both app users, $\{q_k^A\}$, and non-users $\{q_k^N\}$; and, for each weight w , a mixing matrix e_{ij}^w where each entry is the fraction of edges in the network having weight w and joining a vertex of type i to one of type j , $i, j \in \{A, N\}$. The assumptions made in our model are that the app using and non-app using population have the same generalized degree distribution, $q_k^A = q_k^N$, and that the mixing matrices for an app usage rate of U are $e_{AA}^w \propto U^2$, $e_{AN}^w \propto U(1 - U)$ and $e_{NN}^w \propto (1 - U)^2$.

One can readily extend our formula obtained above for the expected post-vaccination reproduction number $\mathbb{E}[R^v]$ to a non-uniformly distributed app-using subnetworks along the same lines as [44]. The more difficult part is finding data to determine the network variables for the model. For exposure notification apps operating under the Google/Apple framework there is currently no centralized collection of contact data and so no direct way for public health authorities to measure these variables. In principle such a functionality could be included in the future, though concerns over privacy must be addressed.

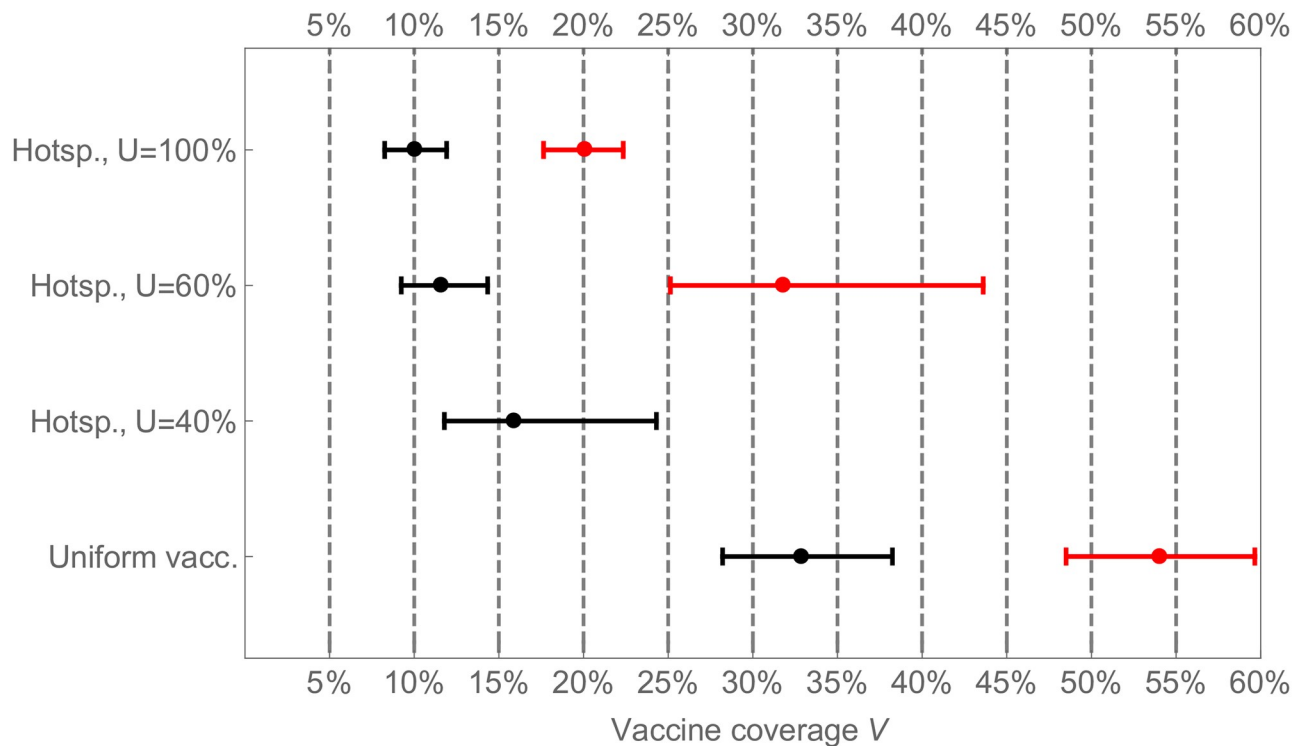


Fig 3. Necessary vaccine coverage to achieve herd immunity. The necessary vaccine coverage to bring the reproduction number from its initial value $R_0 = 1.5$ (black) or $R_0 = 2.2$ (red) to the herd immunity threshold $R^* = 1$ is demonstrated for the uniform and hot-spotting strategy for various app usage rates. The dots represent the mean value from our simulations, whereas the intervals show the range that captures 1σ (68.27%) of all simulation results. If $R_0 = 2.2$ and $U = 40\%$, vaccinating only the app users would not be enough to reach herd immunity.

<https://doi.org/10.1371/journal.pone.0256889.g003>

The Digital Global Health & Humanitarianism Lab carried out case studies on uptake of digital contact tracing apps in 5 countries: Iceland, Cyprus, Ireland, Scotland, and South Africa [45]. Their research indicated a number of individual and system-level factors influencing uptake, such as concerns around data collection, sense of community, accessibility and trust in institutions. While these results likely indicate a degree of preferential attachment, it is otherwise impossible to extract quantitative network statistics. This is especially true since the same social factors could lead to very different network structures in different communities. In particular, there is no a priori reason to assume that app users have more or less exposure on average than non-users.

In the absence of data to inform the structure of the app-using population we limited the scope of this work to the uniform case. We can, however, briefly comment on some features of the general case.

In the limit when *all* app users are vaccinated the residual network of unvaccinated individuals is exactly the subnetwork of non-users. The post-vaccination reproduction number therefore agrees with the basic reproduction number, R_0^N , of the non-user subnetwork. If there is a high degree of preferential attachment between non-users and/or they have much higher overall exposure it could be that $R_0^N > R_0$, leading to an expected efficiency of less than 1 for the hot-spotting strategy. In other words, in this scenario it would have been better to vaccinate the same number of people using the uniform strategy. This doesn't necessarily imply that the uniform strategy is always better than hot-spotting in this population. The hot-spotting strategy initially prioritizes the highest exposure app-users and so can still achieve a higher efficiency than the uniform strategy when vaccinating a smaller fraction of the app-using

population, assuming there are high exposure app-users. Conversely, in cases where $R_0^N < R_0$ the efficiency of the hot-spotting strategy is always more efficient than the uniform strategy.

Concluding remarks

An important limitation of our modelling is the underlying data used to construct the model network [9], which does not take into account any changes in contact patterns as a result of disease spread. Moreover, in both our theoretical work and our simulations we have assumed that there is no distinction between a potentially infectious contact and the contacts detected by Bluetooth exposure notification apps. Our percolation model is not equipped to analyze the time evolution of concurrent vaccination and disease spread. Instead, it is limited to describing the structural properties of the network of unvaccinated individuals. We have also limited our modelling to the hot-spotting strategy deployed in isolation, rather than as one part of a broader vaccine allocation strategy. Finally, we have only considered the impact of vaccination strategies on overall disease spread. Other factors such as mortality or health care system strain are, of course, also important features to consider when comparing vaccination strategies.

These simplifying assumptions could be relaxed in future work with more detailed agent-based simulations that test the generality of the new theory we have introduced. However, we note that a diverse collection of previous literature already finds that prioritizing individuals based on their contact patterns can be highly effective [8, 13–15, 46] at reducing the burden of infectious disease. We have built upon this literature by (1) proposing a measure E^v of the relative efficiency of different strategies, (2) showing that existing COVID-19 digital contact tracing technology allows the measurement of epidemiologically important quantities without violating privacy, (3) proposing a new vaccination strategy based on these measures that is technically feasible to implement, and (4) introducing novel analytical techniques and simulations that (5) demonstrate how hot-spotting is dramatically more efficient than uniform allocation.

Author Contributions

Conceptualization: Mark D. Penney, Yigit Yargic, Lee Smolin, Edward W. Thommes, Madhur Anand, Chris T. Bauch.

Data curation: Mark D. Penney.

Formal analysis: Mark D. Penney, Yigit Yargic.

Funding acquisition: Mark D. Penney, Lee Smolin, Madhur Anand, Chris T. Bauch.

Investigation: Mark D. Penney, Yigit Yargic, Madhur Anand, Chris T. Bauch.

Methodology: Mark D. Penney, Yigit Yargic, Edward W. Thommes.

Project administration: Madhur Anand.

Software: Yigit Yargic.

Supervision: Lee Smolin, Edward W. Thommes, Madhur Anand, Chris T. Bauch.

Visualization: Yigit Yargic.

Writing – original draft: Mark D. Penney, Yigit Yargic.

Writing – review & editing: Mark D. Penney, Yigit Yargic, Lee Smolin, Edward W. Thommes, Madhur Anand, Chris T. Bauch.

References

1. Jentsch Peter C and Anand Madhur and Bauch Chris T Prioritising COVID-19 vaccination in changing social and epidemiological landscapes: a mathematical modelling study. *The Lancet Infectious Disease* 2021. [https://doi.org/10.1016/S1473-3099\(21\)00057-8](https://doi.org/10.1016/S1473-3099(21)00057-8) PMID: 33811817
2. Preliminary guidance on key populations for early COVID-19 immunization;. <https://www.canada.ca/en/public-health/services/immunization/national-advisory-committee-on-immunization-naci/guidance-key-populations-early-covid-19-immunization.html>.
3. Caramelo F, Ferreira N, Oliveiros B. Estimation of risk factors for COVID-19 mortality-preliminary results. *MedRxiv*. 2020;.
4. Persad G, Peek ME, Emanuel EJ. Fairly prioritizing groups for access to COVID-19 vaccines. *Jama*. 2020; 324(16):1601–1602. <https://doi.org/10.1001/jama.2020.18513>
5. Lurie N, Saville M, Hatchett R, Halton J. Developing COVID-19 vaccines at pandemic speed. *New England Journal of Medicine*. 2020; 382(21):1969–1973. <https://doi.org/10.1056/NEJMp2005630>
6. Dushoff J, Plotkin JB, Viboud C, Simonsen L, Miller M, Loeb M, et al. Vaccinating to protect a vulnerable subpopulation. *PLoS Med*. 2007; 4(5):e174. <https://doi.org/10.1371/journal.pmed.0040174> PMID: 17518515
7. Matrajt L, Eaton J, Leung T, Brown ER. Vaccine optimization for COVID-19: who to vaccinate first? *medRxiv*. 2020;. <https://doi.org/10.1101/2020.08.14.20175257> PMID: 32817963
8. Mulberry, Nicola and Tupper, P and Kirwin, E and McCabe, C and Colijn, Caroline Vaccine rollout strategies: The case for vaccinating essential workers early *medRxiv* 2021
9. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*. 2008; 5(3):e74. <https://doi.org/10.1371/journal.pmed.0050074> PMID: 18366252
10. 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 [version 3; peer review: 2 approved] *Wellcome Open Res* 2020, 5:67. <https://doi.org/10.12688/wellcomeopenres.15842.3> PMID: 32685698
11. PRENTISS MARA G and Chu Arthur and Berggren Karl K Superspreading events without superspreaders: Using high attack rate events to estimate N_0 for airborne transmission of COVID-19. *MedRxiv* 2020
12. Chang S., Pierson E., Koh P.W. et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 589, 82–87 (2021). <https://doi.org/10.1038/s41586-020-2923-3> PMID: 33171481
13. Pastor-Satorras R, Vespignani A. Immunization of complex networks. *Phys Rev E*. 2002; 65:036104. <https://doi.org/10.1103/PhysRevE.65.036104>
14. Holme P. Efficient local strategies for vaccination and network attack. *EPL (Europhysics Letters)*. 2004; 68(6):908. <https://doi.org/10.1209/epl/i2004-10286-2>
15. Dushoff J, Plotkin JB, Viboud C, Simonsen L, Miller M, Loeb M, et al. Vaccinating to Protect a Vulnerable Subpopulation. *PLOS Medicine*. 2007; 4(5):1–7. <https://doi.org/10.1371/journal.pmed.0040174> PMID: 17518515
16. Stauffer D, Aharony A. Introduction to percolation theory. CRC press; 2018.
17. Sahimi M. Applications of percolation theory. CRC Press; 2003.
18. MacKay G, Jan N. Forest fires as critical phenomena. *Journal of Physics A: Mathematical and General*. 1984; 17(14):L757. <https://doi.org/10.1088/0305-4470/17/14/006>
19. McKenzie D, Kennedy MC. Power laws reveal phase transitions in landscape controls of fire regimes. *Nature Communications*. 2012; 3(1):1–6. <https://doi.org/10.1038/ncomms1731>
20. Grassberger P. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*. 1983; 63(2):157–172. [https://doi.org/10.1016/0025-5564\(82\)90036-0](https://doi.org/10.1016/0025-5564(82)90036-0)
21. Ingle M, Nash O, Nguyen V, Petrie J, Schwaber J, Szabo Z, et al. Slowing the spread of infectious diseases using crowdsourced data. *IEEE Data Engineering Bulletin*. 2020; p. 12.
22. <https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19/covid-alert.html> Accessed: 2021-07-19. COVID Alert is not available in Alberta, British Columbia, Yukon and Northwest Territories. We therefore use a figure of 28.4 million as the relevant population.
23. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Archiv_Kennzahlen/WarnApp_KennzahlenTab.html Accessed: 2021-07-19
24. <https://www.immuni.italia.it/dashboard.html> Accessed: 2021-07-19

25. <https://www.health.govt.nz/our-work/diseases-and-conditions/covid-19-novel-coronavirus/covid-19-data-and-statistics/covid-19-nz-covid-tracer-app-data> Accessed: 2021-07-19
26. <https://www.tracetogether.gov.sg/> Accessed: 2021-07-19
27. Fenner F., Henderson D. A., Arita I., Jezek Z., & Ladnyi I. D. Smallpox and its eradication (Vol. 6). Geneva: World Health Organization. (1988)
28. Gsell P. S., Camacho A., Kucharski A. J., Watson C. H., Bagayoko A., Nadlaou S. D., et al. Ring vaccination with rVSV-ZEBOV under expanded access in response to an outbreak of Ebola virus disease in Guinea, 2016: an operational and vaccine safety report. *The Lancet infectious diseases*, 17(12), 1276–1284. [https://doi.org/10.1016/S1473-3099\(17\)30541-8](https://doi.org/10.1016/S1473-3099(17)30541-8) PMID: 29033032
29. Holme Petter. Efficient local strategies for vaccination and network attack. *EPL (Europhysics Letters)* 68.6 (2004): 908. <https://doi.org/10.1209/epl/i2004-10286-2>
30. Vassallo L., Di Muro M. A., Sarkar D., Valdez L. D., & Braunstein L. A. Ring vaccination strategy in networks: A mixed percolation approach. *Physical Review E*, 101(5), 052309. (2020). <https://doi.org/10.1103/PhysRevE.101.052309> PMID: 32575220
31. Alvarez-Zuzek L. G., Di Muro M. A., Havlin S., & Braunstein L. A. Dynamic vaccination in partially overlapped multiplex network. *Physical Review E*, 99(1), 012302. (2019). <https://doi.org/10.1103/PhysRevE.99.012302> PMID: 30780375
32. Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of the Royal Society Interface*. 2007; 4(16):879–891. <https://doi.org/10.1098/rsif.2007.1100>
33. Keeling MJ, Eames KT. Networks and epidemic models. *Journal of the Royal Society Interface*. 2005; 2(4):295–307. <https://doi.org/10.1098/rsif.2005.0051>
34. Meyers LA, Pourbohloul B, Newman ME, Skowronski DM, Brunham RC. Network theory and SARS: predicting outbreak diversity. *Journal of theoretical biology*. 2005; 232(1):71–81. <https://doi.org/10.1016/j.jtbi.2004.07.026>
35. Meyers L. Contact network epidemiology: Bond percolation applied to infectious disease prediction and control. *Bulletin of the American Mathematical Society*. 2007; 44(1):63–86. <https://doi.org/10.1090/S0273-0979-06-01148-7>
36. Serrano MÁ, Boguná M. Percolation and epidemic thresholds in clustered networks. *Physical review letters*. 2006; 97(8):088701. <https://doi.org/10.1103/PhysRevLett.97.088701>
37. Newman MEJ, Strogatz SH, Watts DJ. Random graphs with arbitrary degree distributions and their applications. *Phys Rev E*. 2001; 64:026118. <https://doi.org/10.1103/PhysRevE.64.026118>
38. Newman ME. Spread of epidemic disease on networks. *Physical review E*. 2002; 66(1):016128. <https://doi.org/10.1103/PhysRevE.66.016128>
39. Johnson NL and Kotz S and Kemp AW *Distributions in Statistics, Vol. 1, Univariate Discrete Distributions*. Wiley, New York 1993
40. Murphy PJ, Woodard PM, Quintilio D, Titus SJ. Exploratory analysis of the variables affecting initial attack hot-spotting containment rate. *Canadian Journal of Forest Research*. 1991; 21(4):540–544. <https://doi.org/10.1139/x91-075>
41. White Laura Forsberg and Wallinga Jacco and Finelli Lyn and Reed Carrie and Riley Steven and Lipsitch Marc and Pagano Marcello. Estimation of the reproductive number and the serial interval in early phase of the 2009 influenza A/H1N1 pandemic in the USA. *Influenza and other respiratory viruses* 3.6 (2009): 267–276 <https://doi.org/10.1111/j.1750-2659.2009.00106.x> PMID: 19903209
42. Kim H, Paul A. Contact Tracing: a game of big numbers in the time of COVID-19. arXiv preprint arXiv:200410762. 2020;
43. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. *The Lancet Digital Health*. 2020;. [https://doi.org/10.1016/S2589-7500\(20\)30184-9](https://doi.org/10.1016/S2589-7500(20)30184-9) PMID: 32839755
44. Newman, Mark EJ Mixing patterns in networks *Physical review E* 67.2 (2003): 026126. <https://doi.org/10.1103/PhysRevE.67.026126> PMID: 12636767
45. Phillips, Jennie; Molnar, Petra; Babcock, Rebecca. Exploring User-Uptake of Digital Contact Tracing Apps <https://dghhlab.com/publications/>
46. Cohen R, Havlin S, ben Avraham D. Efficient Immunization Strategies for Computer Networks and Populations. *Phys Rev Lett*. 2003; 91:247901. <https://doi.org/10.1103/PhysRevLett.91.247901>