medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248186; this version posted December 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Vaccine Prioritisation Using Bluetooth Exposure Notification Apps

Mark D Penney[1,2], Yigit Yargic[1,3], Lee Smolin[1], Edward W Thommes[4,5,6], Madhur Anand[7], and Chris T. Bauch[2]

[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*

After vaccinating health care workers and vulnerable groups against COVID-19, authorities will need to decide how to vaccinate everyone else. Prioritising individuals with more contacts can be disproportionately effective, in theory, but identifying these individuals is difficult. Here we show that the technology underlying Bluetooth exposure notification applications, such as used for digital contact tracing, can be leveraged to prioritise vaccination based on individual contact data. Our approach is based on the insight that these apps also act as local sensing devices measuring each user's total exposure time to other users, thereby enabling the implementation of a previously impossible strategy that prioritises potential superspreaders. Furthermore, by generalising percolation theory and introducing a novel measure of vaccination efficiency, we demonstrate that this "hot-spotting" strategy can achieve herd immunity with up to half as many vaccines as a non-targeted strategy, and is attractive even for relatively low rates of app usage.

Since the first recorded cases in late 2019, COVID-19 has spread like wildfire across the globe. In order to contain the spread of the disease, an array of non-pharmaceutical interventions have been deployed, including hygienic measures, isolation of infected individuals, quarantine of their exposed individuals, social distancing, and even large scale closures of businesses and institutions [41]. This stimulated efforts to develop COVID-19 vaccines in record time [24]. In late 2020, the first phase of COVID-19 vaccination began with a small number of health-care workers and elderly individuals receiving the Pfizer/BioNTech vaccine. These groups were recommended for vaccine prioritisation [4, 3] based on the high rate of mortality and other severe outcomes experienced by elderly individuals [13] and the principle of reciprocity, which states that those who accepted the greatest risks to mitigate the effects of the pandemic, should be vaccinated first [36].

Once the most vulnerable individuals have been vaccinated, public health authorities will need to decide how to vaccinate the rest of the population[27, 21, 12]. In 2021, data on whether available COVID-19 vaccines prevent both transmissibility and disease in vaccinated individuals will become increasingly available [18]. Hence, public health authorities may want to consider pivoting to a strategy that interrupts transmission of SARS-CoV-2 most effectively. And in all likelihood, the supply of vaccines and/or the capacity to administer them will continue to be limited in early 2021[24], necessitating additional prioritisation decisions.

In this work we study how to prioritise COVID-19 vaccines in the second phase, where vaccination of vulnerable groups has been achieved. We assume a goal of achieving herd immunity through interrupting transmission as efficiently as possible, and we explore a scenario where COVID-19 vaccines can prevent transmissibility as well as disease, as applies to many other existing vaccines against respiratory pathogens [10].

Prioritisation of vaccines is often based on age or other demographic factors, and we know that contact patterns differ between demographic groups. However, as demonstrated in empirical studies of contact patterns [30], heterogeneity within demographic groups is much greater than the differences between them. Because individuals with more contacts have the potential to fuel the spread more than others, prioritising individuals for vaccination according to the risk they pose to further spread has been found to be highly effective in previous theoretical mod-

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248186; this version posted December 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

els [35, 19, 16, 14]. However, such exposure prioritisation 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 Bluetooth exposure notification apps have opened up a new avenue for public health interventions. The core functionality of these 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 which measure the duration of exposure: an epidemiologically significant quantity.

Though other digital contact tracing technologies exist, Bluetooth-based solutions have been chosen when the protection of user's personally identifiable information was a core concern. The data in the encounter log can only be used by public health authorities in a manner which does not require the centralized collection of user data. In fact, in our proposal each user's device makes a decision *purely locally*, based on the number of encounters reported, as to whether or not they will be prioritised for vaccination. By preferentially vaccinating individuals with greater total exposure to others our proposed strategy is significantly more efficient than traditional prioritisation strategies.

The success of digital contact tracing technologies has been hampered by inadequate uptake rates [23, 11]. Our proposal doesn't suffer from this issue. We show that the efficiency depends only on the fraction of the app-using population which receive the vaccine, with the greatest relative reduction occurring with the vaccination of roughly 20-40% of app users.

Our modelling approach is based on percolation theory: a collection of analytical techniques, coming from statistical physics [39], which has been successfully applied to material sciences [37] and to the spread of forest fires [25, 26] and infectious diseases [17]. Here we extend the theory in two directions: Firstly, we introduce a novel measure to compare vaccine prioritisation strategies; and secondly, we incorporate heterogeneity in the risk of infection (total exposure time) between contacts. These tools allow us to show that the improved efficiency gained by piggy-backing vaccine prioritisation 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 COVID-19.

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 [20]. 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.

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 individual contact. Instead, the number of tokens logged over a given timeframe is a measure of the user's total exposure time to others: the sum, over each contact, of the duration of that contact. To model this feature of Bluetooth exposure notification systems and how it impacts vaccination prioritisation strategies, we present new theory in the next section that incorporates both the number and duration of contacts. This theory predicts that the total exposure time of contacts has a direct causal influence on the spread of infectious disease.

Country	Downloads per 100 people
Canada [2]	20
Germany [8]	29
Italy [6]	17
New Zealand [1]	49
Scotland [5]	27
Singapore [7]	56

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

Percolation on Weighted Networks

Many infectious diseases spread through close contact, and contact patterns in human populations display a high degree of heterogeneity [9, 22]. 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 [29, 28, 38]. In the application of percolation theory to infectious diseases, the sum total of human contacts forms a network across the entire population. The vertices of this network are the individuals in the modelled population. An edge joins two individuals if they come into potentially infectious contact with one another in a typical time window matching the infectious period of the disease. The transmissibility of the infectious disease is described by the transmission probability, denoted T.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248186; this version posted December 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

In a series of foundational works [32, 33], Newman developed analytical techniques based on probability generating functions. Through these techniques one can derive formulas for key epidemiological quantities in terms of summary statistics of the degree distribution of the contact network, that is, the distribution of the number of contacts made by individuals in the modelled population. In particular, Newman proved that the basic reproduction number, R_0, depends on the transmission probability T, the average degree, and the variance of the degree distribution. As a corollary of these results, in particular the presence of the variance, we learn a fundamental insight: heterogeneity in contact patterns influences the spread of infectious disease.

The extent to which contact patterns differ between individuals goes beyond the number of contacts they have. Another key variable is the duration of each contact, as over the infectious period of the disease a typical individual will accumulate a large time in contact with a few regular contacts, and a much smaller amount of time with each of a potentially larger number of irregular contacts. Studies have shown that longer contact duration increases the transmission probability of COVID-19 [34]. Both the epidemiological significance of contact duration and the functionality of Bluetooth exposure notification systems imply the need to incorporate contact duration into the percolation framework.

We have generalized Newman's analytic framework to understand percolation on *weighted networks* (see Methods). In our model, each edge, which represents a contact between individuals, is further equipped with a weight representing the duration. The transmission probability is assumed to depend on the duration of contact. This leads to replacing T with a distinct T_w for each weight.

We derive a formula showing that total exposure time of individuals is a driving factor in the spread of disease. As we discussed above, Bluetooth exposure notification apps measure each user's total exposure time to other users of the app. Our model therefore predicts that total exposure time is an epidemiologically relevant quantity with a direct causal relationship to the reproduction number. Therefore, vaccine prioritisation via exposure notification apps is both plausible and potentially effective. To further examine the role of total exposure time in epidemic dynamics, in the next section we develop a quantitative framework for comparing prioritisation strategies.

Because of heterogeneities in contact patterns, vaccination of different individuals will have a varying impact on the rate of spread. In this regard, heterogeneity is a resource: preferential vaccination of those with greater total exposure time should lead to a greater reduction of disease spread from the same number of vaccine doses.

In order to compare different strategies we must further expand the percolation toolkit to incorporate vaccination. We model vaccination as a stochastic process which removes vertices from the original contact network. We assume that the strategies under consideration prioritise individuals according to the number and duration of their contacts, which is the case for strategies leveraging Bluetooth exposure notification apps. In the attached methods we derive formulas for the reproduction number on the residual network of susceptible individuals for a general vaccination strategy of this type.

Assessing Vaccination Efficiency

To compare two vaccine prioritisation strategies it is not enough to consider the post-vaccination reproduction number without taking into account the number of individuals vaccinated by each strategy. This is especially true if one is concerned with how best to use a limited vaccine supply. To that end we introduce a novel measure to compare vaccine prioritisation strategies. We define the *efficiency* of a vaccination strategy to be:

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

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 achieved in the reproduction number per percentage of the population vaccinated. It is important to note that since vaccination is modelled as a stochastic process, both R^v and V, and hence the efficiency, will vary across different realizations of the vaccination process.

The baseline strategy against which we compare prioritisation strategies is the uniform strategy, under which vaccines are distributed uniformly across the population to achieve a target vaccine coverage without taking into account any form of contact heterogeneity. The efficiency of the uniform strategy equals 1 since the expected reproduction number decreases linearly with the vaccine coverage from its original value. Strategies for which the expected 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. In the next section we introduce our exposure prioritisation strategy based on Bluetooth exposure notification apps whose efficiency is much greater than 1. We will refer to this vaccination strategy as "hot-spotting", in reference to a fire-fighting practice that focuses on areas with intense fires [31].

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248186; this version posted December 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

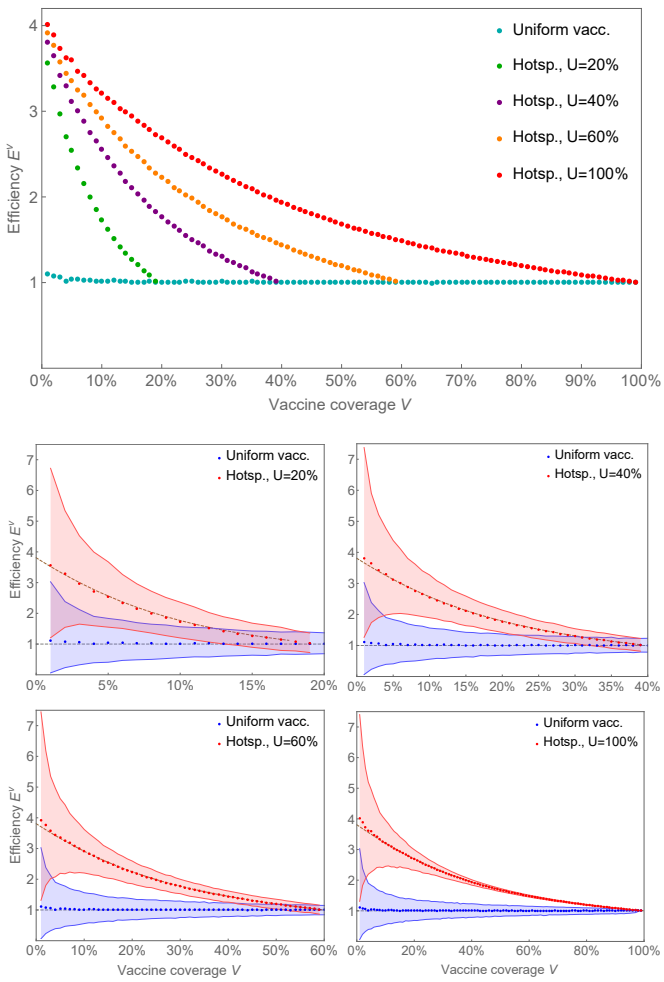

Figure 1: The efficiency for the hot-spotting strategy for $R_0 = 1.5$ with various app usage rates are compared with the uniform vaccination in relation to vaccine coverage. The dots represent the mean efficiency obtained from 10,000 simulations for each dot. The shaded regions show the intervals that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

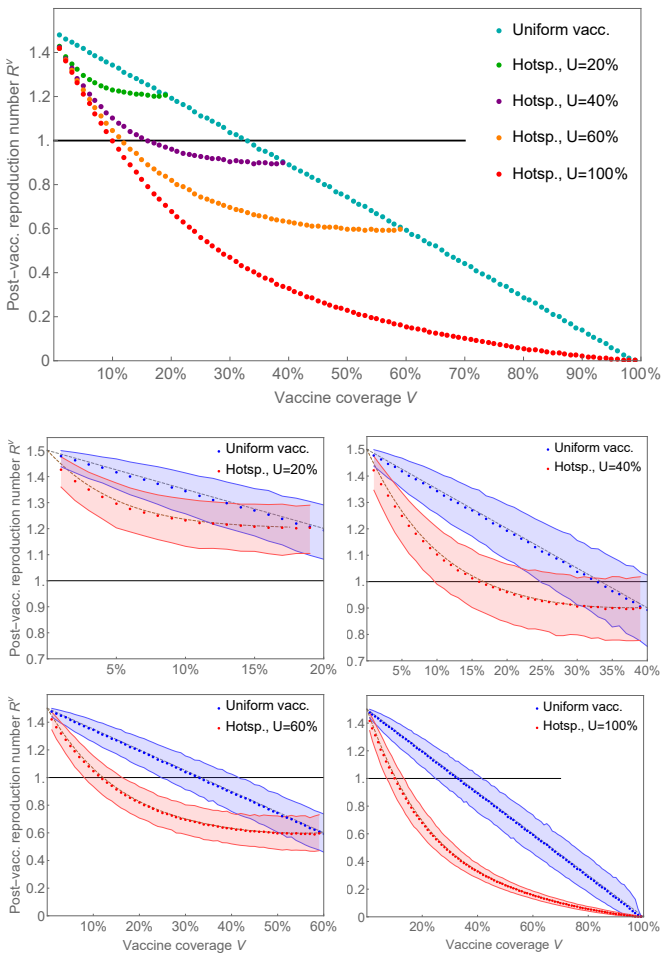

Figure 2: The post-vaccination reproduction numbers for the hot-spotting strategy for $R_0 = 1.5$ with various app usage rates are compared with the uniform vaccination in relation to vaccine coverage. The dots represent the mean efficiency obtained from 10,000 simulations for each dot. The shaded regions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

The Hot-spotting Strategy

Here we propose a "hot-spotting" vaccination strategy that prioritises app-using individuals for vaccination according to their total exposure time. The strategy operates without the central collection of any user data, as the prioritisation of each user is decided locally by their device.

The strategy depends on a parameter β encoding the probability of success in a weighted coin-flip. Each user performs a coin flip for each encounter stored in their encounter log over a fixed time period. Vaccines are prioritised to those 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 total exposure time have a greater probability of being prioritised for vaccination. Note that the extent to which high total exposure time individuals are prioritised can be increased by requiring a greater number of successes.

We simulate this strategy on a simulated network derived from a diary-based population contact survey (see Methods) [30]. In our model, individuals using the app form a subnetwork of the weighted network describing the population's contact patterns. The number of entries in a user's encounter log is their weighted degree in the app-user subnetwork. The results of the prioritisation strategy depend on the rate of app usage in the popula-

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248186; this version posted December 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

tion, denoted U. In particular, the vaccine coverage V (proportion of individuals vaccinated) is bounded by the usage rate U.

In Figure 1 we show the simulation results for the efficiency as a function of vaccine coverage assuming a basic reproduction number of $R_0 = 1.5$. The choice of basic reproduction number has little impact on the efficiency, as we show in the Supplementary Information by considering $R_0 = 2.2$ as well. For very low vaccine coverage the expected efficiency is near 4 for all U values, in agreement with our theoretically derived value of 3.8. There is also a large variance in efficiencies due to small number statistics. As one expects from a strategy prioritising high exposure individuals, the efficiency decreases as vaccine coverage increases. This is because high exposure individuals are likely to be vaccinated first. Nonetheless, across the full range of vaccine coverage rates our proposed strategy is *one to four times more efficient than the uniform vaccination strategy.*

We observe that the expected efficiency curves for differing usage rates are simply the same curve rescaled. This implies that the important quantity to consider for assessing the efficiency is not the overall vaccine coverage V but rather the fraction of the app-using population which are vaccinated.

Figure 2 shows the simulation results for the post-vaccination reproduction number, assuming again an initial $R_0 = 1.5$. The convexity of the curves is a witness to the superior efficiency of our proposed strategy. We see that as the fraction of the app-using population which are vaccinated increases the slope decreases, in alignment with the decreasing efficiency. 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 Figure 3 we show that the significantly lower herd immunity thresholds persist across a range of R_0 values, and applies even when only 40% of the population uses the app.

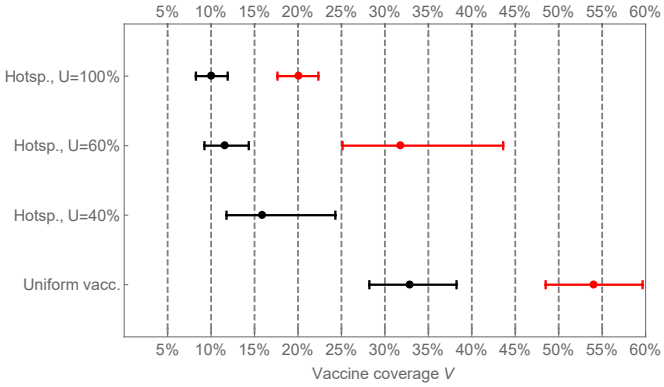

Figure 3: The necessary vaccine coverage to bring the reproduction number from its initial value $R_0 = 1.5$ (black) or $R_0 = 2.2$ (red) to the herd immunity threshold $R^v = 1$ is demonstrated for the uniform and hot-spotting strategy for various app usage rates. The dots represent the mean value from 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.

Concluding Remarks

Implementing our proposal in practice is surprisingly easy on the technical front, although it requires the participation of key players. The exposure prioritisation scheme itself is straightforward and requires no modifications of the core Bluetooth exposure notification framework. Since the Google/Apple system is implemented at the operating system level it is Google and Apple who must incorporate this new functionality. Public health authorities who wish to use hot-spotting must also ensure that their vaccine allocation systems are coordinated with the functionality of the app. Our proposal is, moreover, not strictly limited to digital contact tracing solutions based on Bluetooth. The hot-spotting strategy can be added on to any technology which creates an encounter log for each user.

As for any vaccine prioritisation strategy, hot-spotting comes with ethical concerns, including questions of access, equity, and vaccine program objectives. In addition, if not carefully integrated as part of a holistic approach to public health measures, a naive implementation of this approach could lead to perverse incentives for people to increase their exposure or circumvent the system. We are proposing hot-spotting as part of a broader approach that addresses equity and is implemented in a way that pre-empts harmful behavioural responses [40].

An important limitation of our approach is the underlying data used to construct the model network, which was collected before the pandemic [30]. 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. Finally, our modelling does not predict the time evolution of the epidemic in response to vaccination.

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 prioritising individuals based on their number of contacts can be highly effective [35, 19, 16, 14]. 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

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248186; this version posted December 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

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. We conclude that hot-spotting could enable public health authorities to greatly reduce the social and economic costs of COVID-19 until vaccine supply catches up with demand.

Acknowledgements

We acknowledge support from the National Sciences and Engineering Research Council of Canada (NSERC), NSERC COVID-19 Alliance Grant (MDP, MA, CTB) and Ontario Ministry of Colleges and Universities COVID-19 Rapid Response Grant (MA, CTB).

Patent pending, United States Provisional Application No. 63/122,587.

References

[1] COVID-19: NZ COVID Tracer app data. 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: 2020-12-16.

[2] Download COVID Alert today. https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19/covid-alert.html. COVID Alert is not available in Alberta, British Columbia, Yukon and Northwest Terrorities. We therefore use a figure of 28.4 million as the relevant population. Accessed: 2020-12-16.

[3] How CDC is making COVID-19 vaccine recommendations. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/recommendations-process.html. Accessed: 2020-12-16.

[4] 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. Accessed: 2020-12-16.

[5] Protect Scotland app compatible with English and Welsh app. https://www.gov.scot/news/protect-scotland-app-compatible-with-english-and-welsh-app/.

[6] The numbers of immuni. https://www.immuni.italia.it/dashboard.html. Accessed: 2020-12-16.

[7] TraceTogether, safer together. https://www.tracetogether.gov.sg/. Accessed: 2020-12-16.

[8] Übersicht zu aktuellen und früherer Zahlen und Fakten zur Corona-Warn-App. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Archiv_Kennzahlen/WarnApp_KennzahlenTab.html.

[9] Shweta Bansal, Bryan T Grenfell, and Lauren Ancel Meyers. When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of the Royal Society Interface*, 4(16):879–891, 2007.

[10] Nicole E Basta, M Elizabeth Halloran, Laura Matrajt, and Ira M Longini Jr. Estimating influenza vaccine efficacy from challenge and community-based study data. *American Journal of Epidemiology*, 168(12):1343–1352, 2008.

[11] Isobel Braithwaite, Thomas Callender, Miriam Bullock, and Robert W Aldridge. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. *The Lancet Digital Health*, 2020.

[12] Kate M Bubar, Stephen M Kissler, Marc Lipsitch, Sarah Cobey, Yonatan Grad, and Daniel B Larremore. Model-informed covid-19 vaccine prioritization strategies by age and serostatus. *medRxiv*, 2020.

[13] Francisco Caramelo, Nuno Ferreira, and Barbara Oliveiros. Estimation of risk factors for covid-19 mortality-preliminary results. *MedRxiv*, 2020.

[14] Reuven Cohen, Shlomo Havlin, and Daniel ben Avraham. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.*, 91:247901, Dec 2003.

[15] Odo Diekmann, Johan Andre Peter Heesterbeek, and Johan AJ Metz. On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *Journal of mathematical biology*, 28(4):365–382, 1990.

[16] Jonathan Dushoff, Joshua B Plotkin, Cecile Viboud, Lone Simonsen, Mark Miller, Mark Loeb, and David J. D Earn. Vaccinating to protect a vulnerable subpopulation. *PLOS Medicine*, 4(5):1–7, 05 2007.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248186; this version posted December 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

[17] Peter Grassberger. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, 63(2):157–172, 1983.

[18] Susanne H Hodgson, Kushal Mansatta, Garry Mallett, Victoria Harris, Katherine RW Emary, and Andrew J Pollard. What defines an efficacious covid-19 vaccine? a review of the challenges assessing the clinical efficacy of vaccines against sars-cov-2. *The lancet infectious diseases*, 2020.

[19] Petter Holme. Efficient local strategies for vaccination and network attack. *EPL (Europhysics Letters)*, 68(6):908, 2004.

[20] M Ingle, O Nash, V Nguyen, J Petrie, J Schwaber, Z Szabo, M Voloshin, T White, and H Xue. Slowing the spread of infectious diseases using crowdsourced data. *IEEE Data Engineering Bulletin*, page 12, 2020.

[21] Peter Jentsch, Madhur Anand, and Chris T Bauch. Prioritising covid-19 vaccination in changing social and epidemiological landscapes. *medRxiv*, 2020.

[22] Matt J Keeling and Ken TD Eames. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307, 2005.

[23] Hyunju Kim and Ayan Paul. Contact tracing: a game of big numbers in the time of COVID-19. *arXiv preprint arXiv:2004.10762*, 2020.

[24] Nicole Lurie, Melanie Saville, Richard Hatchett, and Jane Halton. Developing COVID-19 vaccines at pandemic speed. *New England Journal of Medicine*, 382(21):1969–1973, 2020.

[25] Gary MacKay and Naeem Jan. Forest fires as critical phenomena. *Journal of Physics A: Mathematical and General*, 17(14):L757, 1984.

[26] Donald McKenzie and Maureen C Kennedy. Power laws reveal phase transitions in landscape controls of fire regimes. *Nature Communications*, 3(1):1–6, 2012.

[27] Jan Medlock and Alison P Galvani. Optimizing influenza vaccine distribution. *Science*, 325(5948):1705–1708, 2009.

[28] Lauren Meyers. Contact network epidemiology: Bond percolation applied to infectious disease prediction and control. *Bulletin of the American Mathematical Society*, 44(1):63–86, 2007.

[29] Lauren Ancel Meyers, Babak Pourbohloul, Mark EJ Newman, Danuta M Skowronski, and Robert C Brunham. Network theory and SARS: predicting outbreak diversity. *Journal of theoretical biology*, 232(1):71–81, 2005.

[30] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*, 5(3):e74, 2008.

[31] Peter J Murphy, Paul M Woodard, Dennis Quintilio, and Stephen J Titus. Exploratory analysis of the variables affecting initial attack hot-spotting containment rate. *Canadian Journal of Forest Research*, 21(4):540–544, 1991.

[32] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64:026118, Jul 2001.

[33] Mark EJ Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.

[34] Oon Tek Ng, Kalisvar Marimuthu, Vanessa Koh, Junxiong Pang, Kyaw Zaw Linn, Jie Sun, Liang De Wang, Wan Ni Chia, Charles Tiu, Monica Chan, et al. SARS-CoV-2 seroprevalence and transmission risk factors among high-risk close contacts: a retrospective cohort study. *The Lancet Infectious Diseases*, 2020.

[35] Romualdo Pastor-Satorras and Alessandro Vespignani. Immunization of complex networks. *Phys. Rev. E*, 65:036104, Feb 2002.

[36] Govind Persad, Monica E Peek, and Ezekiel J Emanuel. Fairly prioritizing groups for access to covid-19 vaccines. *Jama*, 324(16):1601–1602, 2020.

[37] Muhammad Sahimi. *Applications of percolation theory*. CRC Press, 2003.

[38] M Ángeles Serrano and Marián Boguná. Percolation and epidemic thresholds in clustered networks. *Physical review letters*, 97(8):088701, 2006.

[39] Dietrich Stauffer and Ammon Aharony. *Introduction to percolation theory*. CRC press, 2018.

[40] John D Sterman. Learning from evidence in a complex world. *American journal of public health*, 96(3):505–514, 2006.

[41] Ashleigh R Tuite, David N Fisman, and Amy L Greer. Mathematical modelling of COVID-19 transmission and mitigation strategies in the population

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248186; this version posted December 17, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

of Ontario, Canada. *CMAJ*, 192(19):E497–E505, 2020.

[42] Jeroen JA van Kampen, David AMC van de Vijver, Pieter LA Fraaij, Bart L Haagmans, Mart M Lamers, Nisreen Okba, Johannes PC van den Akker, Henrik Endeman, Diederik AMPJ Gommers, Jan J Cornelissen, et al. Shedding of infectious virus in hospitalized patients with coronavirus disease-2019 (COVID-19): duration and key determinants. *MedRxiv*, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248186; this version posted December 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Methods

We model the spread of infectious disease in a population as a percolation process on the network of contacts. Each vertex in this network represents an individual person and each edge links two people who come into contact. Our particular approach is a generalization of Newman's *probability generating function* framework [33] to weighted networks.

Percolation Theory

The key mathematical object in the original formalism is the probability generating function of the degree[1] distribution,

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k \ ,$$

where p_k is the fraction of vertices in the network having degree k. This function is useful for writing analytic expressions for moments of the distribution; for example, the average degree is given by $\mu = G_0'(1)$.

An infectious disease spreads in the network through occupied edges. In the original model, each edge has a uniform probability T for being occupied. The basic reproduction number[2] R_0 corresponds to *the expected number of other occupied edges attached to a vertex at the end of a random occupied edge*. This number is given by

$$R_0 = T\,\frac{G_0''(1)}{G_0'(1)} \ .$$

In terms of the mean μ and the standard deviation σ of the degree distribution, this equation is equivalent to

$$R_0 = T\left[\mu\left(1+\frac{\sigma^2}{\mu^2}\right)-1\right] \ .$$

Weighted Percolation

In order to model vaccine prioritisation 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 along an edge for the percolation process is assumed to depend on that edge's weight, so that the single parameter T 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. 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. Given that this is quite a flexible formalism it can be applied to a broad range of modelling tasks.

For weights representing contact duration there is a natural candidate for the transmission probabilities T_w. 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

$$T_w = 1-(1-T_1)^w \ .$$

Once again, T_w is the probability for a weight-w edge to be occupied. The infection spreads through occupied edges.

In a weighted network we represent the configuration of edges around each vertex by a *generalized degree*, denoted $\boldsymbol{k}$. This is a vector having an entry for each of the possible weights appearing in the network. For a vertex of generalized degree $\boldsymbol{k}$, the entry corresponding to weight w, denoted k_w, is the number of contacts having weight w. There are two important quantities we can extract from $\boldsymbol{k}$: First, the degree of a vertex, that is, the number of distinct contacts, is simply $\sum_w k_w$. Second, the *weighted degree* of a vertex, that is, the sum of the weights on each edge, is $\sum_w w k_w$. When w represents contact duration, the weighted degree is the total exposure time of that individual to others.

To generalize Newman's results to the weighted setting we introduce a *multivariable generating function* for the distribution of generalized degrees. There is one variable y_w for each weight w appearing in the network, and we write $\boldsymbol{y}$ for the vector of these variables. Let $q_{\boldsymbol{k}}$ denote the fraction of vertices in the network having generalized degree $\boldsymbol{k}$. Then the multivariable generating function of the network is

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

The generating function $G_0(x)$ for the degree distribution in a weighted network is related to this function by $G_0(x) = Q(\mathbf{1}x)$, where $\mathbf{1}$ is a vector with each entry being 1.

In order to derive the basic reproduction number, we need the distribution of *occupied degree*, i.e., the number of occupied edges attached to a vertex. This distribution is given by

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

where $\boldsymbol{T}$ is the vector composed of the transmission prob-

[1] The degree of a vertex is the number of edges attached to it.

[2] In heterogeneous populations, R_0 is defined as "the expected number of secondary cases produced, in a completely susceptible population, by a typical infected individual during its entire period of infectiousness" [15].

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248186; this version posted December 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

abilities T_w. Furthermore, we introduce

$$G_1(x;\boldsymbol{T}) = \frac{G_0'(x;\boldsymbol{T})}{G_0'(1;\boldsymbol{T})}$$

as the probability distribution of the number of other occupied edges attached to a vertex reached by following a random occupied edge (cf. [33]). By definition, the basic reproduction number R_0 is precisely the mean of this final distribution, i.e. $R_0 = G_1'(1;\boldsymbol{T})$.

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

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

Then, we find

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

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

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

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

Total Exposure Time

Let us now focus on the case when the weights represent contact duration and $T_w = 1 - (1 - T_1)^w$. In this case, the weighted degree of a vertex is the total exposure time of that individual to others, measured in discrete time steps. To better understand the impact of the total exposure time on the reproduction number, consider the approximation $T_w \approx wT_1$, which is valid when $wT_1 << 1$ for all w. Under this approximation, we find the heuristic formula

$$R_0 \approx T_1 \left[\hat{\mu}\left(1 + \frac{\hat{\sigma}^2}{\hat{\mu}^2}\right) - \frac{\mathrm{Av}\left[\sum_i w(i)^2\right]}{\hat{\mu}}\right] .$$

Here, $\hat{\mu}$ and $\hat{\sigma}$ are the mean and the standard deviation of the total exposure time. Our model therefore predicts that the total exposure time has a direct influence on the spread of disease.

Vaccination on Contact Networks

We model vaccination as a stochastic process wherein a vertex of generalized degree $\boldsymbol{k}$ has a probability $v(\boldsymbol{k})$ 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.

We introduce the *post-vaccination reproduction number* R^v as the reproduction number in the non-vaccinated sub-population after a vaccination process. Since vaccination is modelled as a stochastic process, R^v is a random variable.

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, some vertices are removed themselves. Each vertex of generalized degree $\boldsymbol{k}$ has a probability $(1 - v(\boldsymbol{k}))$ of remaining in the residual network. Secondly, some contacts of the remaining vertices are removed. A weight-w contact has the probability

$$\varphi_w = 1 - \frac{\mathrm{Av}\left[k_w v(\boldsymbol{k})\right]}{\mathrm{Av}\left[k_w\right]}$$

of remaining non-vaccinated. Hence, the expected post-vaccination reproduction rate is given by

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

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

Efficiency of Vaccine Prioritisation

In general, one expects that prioritising the vaccination of individuals with greater exposure to others achieves a greater reduction in R^v than vaccinating the same number of less-connected individuals. This translates to slowing the infection more effectively and a reduction in the number of infections. Vaccine prioritisation becomes particularly important when only a small supply of vaccines is available. In that situation, adopting a strategy that prioritises highly connected individuals for vaccination can save the lives of more people.

To compare different prioritisation strategies we are concerned with not only how much they reduce the spread of disease, but also how many vaccines it takes for that outcome. To that end, we introduce a novel measure of prioritisation efficiency,

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

where V is the fraction of the population which receives the vaccine.

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 $\mathbb{E}[V] = \mathrm{Av}[v(\boldsymbol{k})]$

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248186; this version posted December 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

and the variance $\mathrm{Var}[V] = \frac{1}{N}\,\mathrm{Av}\left[v(\boldsymbol{k})\,(1 - v(\boldsymbol{k}))\right]$, 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

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

For a uniform vaccination probability v without any prioritisation, one would have $\mathbb{E}\left[R^v\right] = (1 - v)\,R_0$ and $\mathbb{E}\left[E^v\right] = 1$. Through a realistic vaccine prioritisation strategy, it is possible to achieve a significantly higher efficiency, while the exact numbers depend on the underlying network.

The Hot-spotting Strategy

We propose a *hot-spotting strategy* that can be implemented simply on the existing Bluetooth based exposure notification apps. This strategy is based on a Bernoulli process for each exposure registered in the app with a chance of β for success. The app users with at least one success are given the priority for vaccination. Hence, if an app user has a total exposure number n in the app, they will have a chance of $1 - (1 - \beta)^n$ for being selected for prioritised vaccination in this strategy.

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. Then, for a random individual with the contact structure $\boldsymbol{k}$, which is not necessarily registered in the app network, the probability for them being an app user and also being selected for prioritised vaccination is given by

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

where

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

Model Network

The well-known POLYMOD study surveyed over 7000 individuals across 8 European countries [30]. 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. 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.

To model COVID-19, the network should capture the contacts made over the typical infectious period of the disease. This varies significantly between patients, with one study estimating a median of 8 days after the onset of symptoms [42]. 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.

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

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

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, $n_1, \ldots, n_{14}$.
2. Set $F'_n = D_{n_1} + I_{n_2} + \ldots + I_{n_{14}}$.
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

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248186; this version posted December 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

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 presented in the main text 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 the chosen data of 500 samples, we fix the unit transmission probability at $T_1 = 0.000375$ to get the basic reproduction number $R_0 = 1.501$. We select a set of parameters β for the vaccination probability function $v(\boldsymbol{k}; \beta, U)$ of the hot-spotting strategy (and a set of parameters v 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 $v(\boldsymbol{k}; \beta, U)$ 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 $v(\boldsymbol{k}; \beta, U)$. 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248186; this version posted December 17, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Information

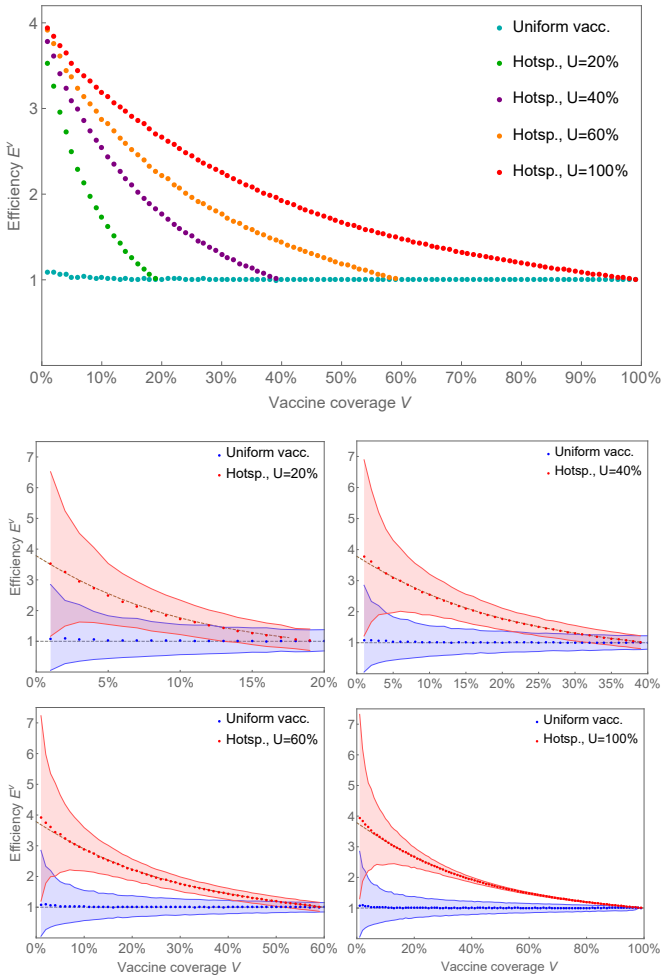

Figure 4: The efficiency for the hot-spotting strategy for $R_0 = 2.2$ with various app usage rates are compared with the uniform vaccination in relation to vaccine coverage. The dots represent the mean efficiency obtained from 10,000 simulations for each dot. The shaded regions show the intervals that capture 90% of all simulation results. The dashed lines in the bottom figures show the expected efficiency obtained from the formula in Methods.

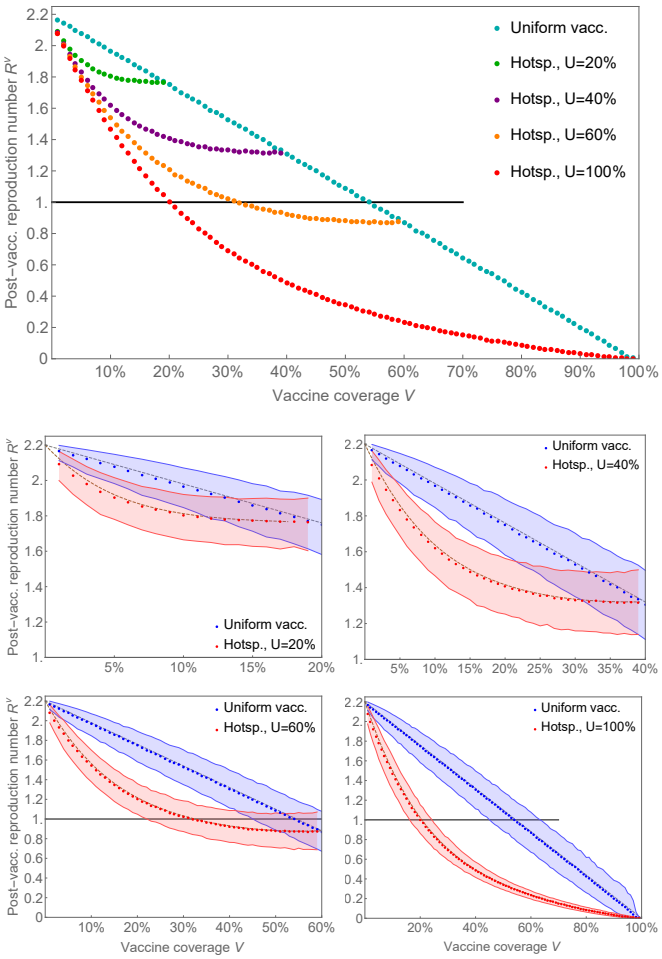

Figure 5: The post-vaccination reproduction numbers for the hot-spotting strategy for $R_0 = 2.2$ with various app usage rates are compared with the uniform vaccination in relation to vaccine coverage. The dots represent the mean efficiency obtained from 10,000 simulations for each dot. The shaded regions show the intervals that capture 90% of all simulation results. The dashed lines in the bottom figures show the expected post-vaccination reproduction numbers obtained from the formula in Methods.