

RESEARCH ARTICLE

Test allocation based on risk of infection from first and second order contact tracing

Gabriela Bayolo Soler^{ID*}, Miraine Dávila Felipe^{ID}, Ghislaine Gayraud^{ID}

LMAC (Laboratory of Applied Mathematics of Compiègne), Université de technologie de Compiègne, Compiègne, France

* gabriela.bayolo-soler@utc.fr



Abstract

Strategies such as testing, contact tracing, and quarantine have been proven to be essential mechanisms to mitigate the propagation of infectious diseases. However, when an epidemic spreads rapidly and/or the resources to contain it are limited (e.g., not enough tests available on a daily basis), to test and quarantine all the contacts of detected individuals is impracticable. In this direction, we propose a method to compute the individual risk of infection over time, based on the partial observation of the epidemic spreading through the population contact network. We define the risk of individuals as their probability of getting infected from any of the possible chains of transmission up to length-two, originating from recently detected individuals. Ranking individuals according to their risk of infection can serve as a decision-making tool to prioritise testing, quarantine, or other preventive measures. We evaluate interventions based on our risk ranking through simulations using a fairly realistic agent-based model calibrated for COVID-19 epidemic outbreak. We consider different scenarios to study the role of key quantities such as the number of daily available tests, the contact tracing time-window, the transmission probability per contact (constant versus depending on multiple factors), and the age since infection (for varying infectiousness). We find that, when there is a limited number of daily tests available, our method is capable of mitigating the propagation more efficiently than some other approaches in the recent literature on the subject. A crucial aspect of our method is that we provide an explicit formula for the risk, avoiding the large number of iterations required to achieve convergence for the algorithms proposed in the literature. Furthermore, neither the entire contact network nor a centralised setup is required. These characteristics are essential for the practical implementation using contact tracing applications.

OPEN ACCESS

Citation: Bayolo Soler G, Dávila Felipe M, Gayraud G (2025) Test allocation based on risk of infection from first and second order contact tracing. PLoS ONE 20(4): e0320291. <https://doi.org/10.1371/journal.pone.0320291>

Editor: Alberto Aleta, University of Zaragoza, SPAIN

Received: September 16, 2024

Accepted: February 17, 2025

Published: April 7, 2025

Copyright: © 2025 Soler 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: All relevant data are within the paper, its [Supporting Information](#) files, and on GitHub. (https://github.com/gbayolo26/risk_estimation).

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

In the context of epidemics, contact tracing is the process of identifying individuals who have been in contact with other individuals diagnosed with a transmissible disease. The relevant contacts are those that would allow the transmission to happen, which depends on the

mode of transmission of the disease, and requires the detected individual to be infectious at the time of the encounter. Together with strategies such as testing and quarantining, contact tracing has been shown to be an essential mechanism in order to mitigate the spread of a disease, allowing to contain and delay outbreaks, see [1]. Ideally, detected individuals (those who receive a positive test result) are quarantined and their contacts are identified and tested as well. However, these interventions have an economic and social cost, and a scenario where all the contacts of detected individuals are tested is not realistic for diseases that start spreading quickly in the population (outbreak). Hence, when the resources are limited (e.g., amount of daily available tests), the question of how to cleverly allocate them to the population arises.

In this direction, we propose a method to compute the risk of infection of individuals in the population over time, based on the partial observation of the epidemic spreading through the population contact network. The risk of each individual is defined as her/his (marginal) probability of infection conditionally on the observed variables in the recent past, and the higher-risk individuals can get notified to be tested, quarantined, or applied any other preventive measures. Thus, the quantification of the infection risk is proposed here as a tool to allocate the available resources more rationally than just randomly. Similar intervention approaches have been shown to have a positive impact on mitigating epidemics, as seen in [2–12].

To be more precise about our approach, we consider an agent-based model on a fixed-size population where individuals admit a set of (discrete) characteristics influencing the transmission of the disease. The pairwise contacts between individuals are described by a *dynamical network* model, in which some connections are deleted and some others are created while time evolves. Then, we depict the disease spread in the population contact network by a stochastic *Susceptible-Infectious-Removed* (SIR) dynamic (including more than three classes), meaning that contagion can only happen (with a certain probability) when an infectious individual is connected with a susceptible one by an edge in this contact network. In particular, we consider a non-Markovian dynamic since the infection probability per interaction depends on the date of infection of the source. This probability is also a function of the previously mentioned characteristics of both individuals and the context of the interaction (e.g., in the household, workplace, or random).

Given this dynamical network model and the propagation process on this network, we suppose that the infectious statuses of individuals are (partially) observed through testing, as well as the underlying contact network, the factors having an impact on the transmission for the set of tested individuals, and their direct and secondary contacts. Actually, we consider *at risk* not only the first-degree contacts of detected individuals (1° contacts) but also their subsequent contacts (2° contacts). In the sequel, to distinguish the different groups of individuals under consideration, we call *index cases* the detected individuals while they are infectious, 1° *contacts* the individuals who are not detected but interacted with index cases while the latter were infectious, and, 2° *contacts* the individuals who are not detected but interacted with 1° contacts after the latter were in contact with index cases. In addition, 1° *interaction* and 2° *interaction* refer to the risky encounter between index cases and 1° contacts, and between 1° and 2° contacts respectively.

In real situations, the information about the tested individuals and their 1° and 2° contacts is provided either by individuals themselves, through manual contact tracing (MCT), or by a digital contact tracing (DCT), and is not necessarily centralized in the digital case. The advantages of extending the contact tracing strategies up to 2° contacts have been highlighted in the recent literature, see [7,13,14]. Then, to compute the risk, we consider a rather general probability of transmission per interaction, depending on the observed attributes of both individuals, characteristics of the interaction, and the infection time of the source (which is unknown

and estimated from the observations). Finally, we compute the probability for each individual at risk of having been infected by one of the individuals detected in the previous days (fixed time window), through chains of transmission of length one or two. It is worth noticing that we compute this marginal probability by summing over all the possible paths of transmission, providing explicit formulas, and avoiding independence assumptions and the *cycling back* phenomenon described in [S2 appendix](#).

Finally, we propose and simulate the following mitigation strategy: every day the probability of being infected is computed for the individuals at risk, and a fixed number of the highest-ranked individuals are tested; the newly detected individuals are put in quarantine the day after, and the process is repeated each day during an intervention period. In parallel of the detection by risk, symptomatic individuals are tested with a fixed probability per day since the beginning of their symptoms, the ones detected are quarantined, and their contacts are traced as described before. We evaluate this intervention through simulations in a series of different scenarios, where we study the impact of some of the parameters in the model such as the number of daily available tests, the time-frame for the time of infection and the proportion of detection of symptomatic individuals. We also study the influence of the probability function of transmission per interaction and different ways to estimate the time of infection of the source. We further investigate the influence of the mean detection time for individuals who develop symptoms, test sensitivity and specificity, and the quarantine adoption fraction. Additionally, we examine the role of super-spreaders within our model. We found that, in most cases, our test allocation method is capable of mitigating the propagation of the disease considerably faster than randomly selecting (RS) individuals to get tested, or the usual contact tracing (CT, i.e. ranking according to the number of interactions with detected individuals). Moreover, we found that with fewer daily available tests, our risk ranking is more efficient than an equivalent setting where the probabilities are computed under the mean-field (MF) hypothesis, see [\[2\]](#).

In the recent literature, there are many research works that study the effect of contact tracing combined with treatment and/or quarantine, as non-pharmaceutical interventions for infectious disease mitigation and control, see [\[15\]](#). The SARS-CoV-2 outbreak has considerably increased the scientific research motivation around strategies including contact tracing. In particular, DCT apps have attracted the attention of Public Health authorities and the scientific community as well, [\[16,17\]](#). DCT apps allow to collect the information automatically, and provide fast processing times but also come with drawbacks regarding privacy and the data protection standards ruling in most countries, see [\[18\]](#). This privacy protection issues have had to be carefully addressed in order to implement this type of strategies in practice. It is not our aim to discuss here how DCT apps should be implemented, however, it is worth noticing that the tracing of random interactions, included in our simulations, is only possible through DCT. The effectiveness of contact tracing interventions for the COVID-19 pandemic is studied in [\[19\]](#) and more recently in [\[17,20\]](#), providing reviews of what has been done on this topic based on empirical and simulated data.

Mathematical models play a crucial role in understanding the spread of epidemics, predicting their progression, guiding public health decision-making, and evaluating the effectiveness of intervention measures. While simpler models are easier to understand and implement, they often represent reality in a more abstract manner, capturing only essential features [\[21,22\]](#). Among the most widely recognized families of epidemic models are the compartmental SI models, which describe epidemic spread within a homogeneous population by classifying individuals into groups based on their epidemic status [\[23\]](#).

To achieve greater realism, these compartmental models have been extended to agent-based models that account for the social structure and individual interaction patterns within a

population. The pandemic caused by SARS-CoV-2 has significantly increased interest in these models, in particular for evaluating the impact of non-pharmaceutical intervention strategies such as quarantine, contact tracing, lockdown, and social distancing [24–26]. Agent-based models incorporating contact network structures further enhance realism by accounting for heterogeneities in social interactions, which are critical for understanding and mitigating the spread of infectious diseases. In this work, we adopt an agent-based modeling approach based on networks to provide a realistic framework for assessing our proposed contact tracing method.

Specifically, in our simulations we use the Oxford OpenABM-Covid19 model, which defines a contact network based on demographic characteristics of the UK population calibrated for the transmission of airborne diseases [27]. Furthermore, in this model, the spread of the COVID-19 epidemic follows an enriched SIR dynamic (11 possible disease statuses), stratified by age group and context of the interaction (e.g., household, workplace, random encounter). It also accounts for key epidemiological features of the COVID-19 outbreak. In particular, the model defines an infectiousness function that varies with time since infection and considers asymptomatic and presymptomatic states. By capturing essential aspects of real-world contact patterns and the epidemiology of COVID-19, this model provides a solid foundation for evaluating the proposed risk estimation method. It reflects not only the structural heterogeneity of populations, but also the dynamic and nuanced nature of epidemic transmission, making it highly suitable for our analysis.

Among the recent research studies looking at non-pharmaceutical interventions strategies, there are a few with the same aim as ours, that is, to integrate individual infection risk levels based on the observation of the interaction network and the test results in order to optimize the allocation of the available resources. In [7] and [4], the risk is computed using Monte Carlo methods. In [7], the authors estimate the individual infection probability up to 3° contact tracing, arguing that it improves the detection of asymptomatic patients in diseases with a high percentage of them. Here, instead, we derive an explicit formula for these probabilities taking into account up to 2° contacts, and hence avoiding the large computing power and the centralized information required by Monte Carlo methods. Other works such as [2] and [6] avoid the use of Monte Carlo methods by using the mean-field approximation to evaluate the individual risk of infection. However, the way the risk is propagated is “bidirectional,” meaning that it is not only “forward” in the direction of the transmission given the observations. Indeed, they suppose that individuals interchange their risk information at each time step if there is an edge between them, regardless of the previous path followed by the transmitted risk, falling sometimes in the above mentioned cycling back issue. Despite the similarities in the use of the MF hypothesis, it should be noticed that in [6] the network and propagation model are simpler than in [2]. The latter deals with more realistic models, including the OpenABM-Covid19 model used in our simulations. Moreover, in [2] a second method is developed, that estimates the individual infection risk as the posterior distribution conditional on the test observations through the Belief Propagation (BP) inference algorithms. Similar computations are achieved in [3] and [5] using Gibbs Sampling (GS) and Factorized Neighbors (FN) respectively. Another related methodology is presented in a series of works [9,10,12,28–32], where the authors achieve the risk computation using deep learning algorithms based on neural networks. It is also worth mentioning other machine learning approaches for risk estimation from DCT data that exploit different observations, such as Bluetooth energy measurements and exposure data, see [8,11].

A crucial aspect of our work is that we consider a rather realistic contact network and a detailed disease spread model, see [27]. Another significant aspect is our consideration of 2°

contact tracing, which provides a more accurate estimation of the risk compared to 1° contact tracing. Although our approach can be extended to 3° contacts and beyond, the calculations would get much heavier, and we argue that the gain in the effectiveness of the mitigation would not be significant, due to the uncertainty on the statuses of the intermediate individuals in the chains of transmission. One more core feature of our method is that to compute the risk, neither the whole contact network nor a centralized setup (contacts and individual information) is required. These characteristics are crucial for practical implementation using DCT applications, where the interchange of information between contacts across the entire network could be challenging due to privacy restrictions impacting a vast amount of personal data. These challenges can be exacerbated in centralized systems, see [5]. Furthermore, compared to the previously mentioned inference algorithms used to calculate the individual risk of being infected (i.e., BP, GS, FN), our risk calculation is simpler: while these algorithms integrate the observations at any time t by updating and re-propagating the risks step-by-step in a given time interval previous to t for every contact (up to any contact degree) of all the individuals in the population, we calculate directly the risk at t of individuals in contact (up to 2°) with someone detected by integrating the probability of any possible path of length up to 2 that might lead to the infection of these individuals. In this way, we avoid any cycling back phenomenon, and we do not need to update the risk of all individuals for every time step in the contact tracing time window, getting a very low level of messages interchange between individuals, see [5].

Methods

We aim at defining a practical, realistic, and efficiently implemented risk-based dynamic detection process allowing us to identify the most likely infected individuals. To take into account the heterogeneity of disease transmissions, we consider that the probability of infection per interaction depends on individual attributes (e.g., age, healthy habits) and infectiousness of the source (e.g., day since infection, type of symptoms) of individuals in contact and the characteristics of the interaction (e.g., place, duration, distance, protective measure).

In addition, to provide a sharper and earlier detection process of the most likely infected individuals, not only the direct contacts of detected individuals are considered to be at risk, but also their subsequent contacts (2° contact tracing). Compared with the usual 1° contact tracing, we expect to obtain a more accurate estimation of the risk of infection, and hence to detect more efficiently (in terms of the mitigation of the epidemic) individuals that are in general harder to detect due to the absence of symptoms (asymptomatic or pre-symptomatic individuals). It has become clear from numerous research studies that these latter individuals play an important role in SARS-CoV-2 transmission dynamics, see [13,14,33].

From a realistic point of view, one can estimate the individual risk of getting infected only from the observations (available information) that are provided by the tested individuals and their contacts through manual or digital CT. In view of all the above, our intervention approach relies on a dynamic risk evaluation for 1° and 2° contacts, and their risk is defined as the marginal conditional probability of getting infected given their past known 1° and 2° interactions, and the information provided by the tests.

In the following, we give a brief description of disease-spread models on social networks, we introduce some useful notations and finally, we define the risks of infection for 1° and 2° contacts.

Disease spread model on social networks

We consider a population consisting of N ($N \in \mathbb{N}$) individuals that stays constant over time, so neither births, deaths nor migrations are taken into account. Notice that constant population size is a convenient assumption for the notation, but small variations in the population size do not influence the procedures described in the sequel. Let us denote by $V = \{1, \dots, i, \dots, j, \dots, N\}$ the population under consideration.

Social structure model

At any discrete time t ($t \in \mathbb{N}$), the social structure (interactions between individuals at t) is represented by an undirected graph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$. We consider that $\mathcal{V} = \{(i, a^i) : i \in V \text{ and } a^i \in A\}$ corresponding to the set of vertices V (the individuals), supplemented by the set A of vertices attributes mentioned at the beginning of the section. Likewise, $\mathcal{E}_t = \{(i, j, c_t^{ij}) : (i, j) \in E_t \text{ and } c_t^{ij} \in C\}$ is the set of the edges E_t , describing the interactions at time t between the corresponding individuals, supplemented by C , the set of the characteristics of these interactions.

The time interval $[0 : T]$ corresponds to the period of study, where we assume that at the first time 0 there is already an ongoing outbreak (a small number of infectious individuals) and at the last time T the study ends. Here the time unit is one day. In the sequel, to refer to the discrete-time interval between any t_l and t_m , we use $[t_l : t_m]$ with the convention that the interval is empty when $t_l > t_m$.

Although the stochastic mechanism of the social network evolution over time is not of primary interest here, the sequence $\mathcal{G}_{[0:T]}$ is however viewed as a realization of a dynamic random network model over the time-period $[0 : T]$ (see [34]). As previously mentioned, it is reasonable to consider that the social structure is partially random (except the sub-graph corresponding to household interactions), and hence that some individual and edge attributes are governed by some specific probability distributions.

Infectious disease spreading on the network

Here, we consider an individual-based SIR dynamic spreading on the underlying social network. The possible individual statuses are only Susceptible (S), Infected (I) and Removed (R), and the only possible status evolution over time are $S \rightarrow I$ and $I \rightarrow R$, where R considered as an absorbing state. We denote by $X_t^i \in \{S, I, R\}$ the random variable corresponding to the status of individual i at time t .

While the dynamic social network is represented by undirected graphs, the transmission of the infectious disease is directed along an edge from an infectious individual (*source* or *donor*) to a susceptible one (*recipient*). We consider that the transmission probability depends on the infection time of the source, as well as on both individual and interaction attributes A and C . For any time $t \geq 0$, and any two distinct individuals i and j in V , seen respectively as possible source and recipient, recall that $a^i, a^j \in A$ denote the individual attributes, and $c_t^{ij} \in C$ the characteristics of the interaction. In addition, we consider another attribute related to the disease spreading, namely $b_t^i \in B$, which corresponds to the type of symptoms the individual i manifests at time t . We have in mind that the individual i could be for example asymptomatic, mild, or severe, and the parameter b_t^i allows us to establish how the severity of the symptoms influences the probability of transmission of i . Finally, we denote by τ_I^i and τ_R^i the not observed random variables representing the infection and removal times of individual i , taking values in $[0 : t] \cup \{+\infty\}$, where for convenience we set $\tau_I^i = \tau_R^i = \infty$ if i has not yet been infected. We suppose that the transmission probability that the individual i infects j at t depends on all the above quantities (parameters and random variables). Hence, we denote it by $\lambda_{a^i, a^j, t, b_t^i, \tau_I^i, \tau_R^i, c_t^{ij}}^{i \rightarrow j}$ and for the sake of simplicity we use the following simplified notation,

$$\lambda_{a^i, a^j, t, b^i, \tau_l^i, \tau_R^i, c_t^{ij}}^{i \rightarrow j} \equiv \lambda_{\tau_l^i, t}^{i \rightarrow j}.$$

Notice that

$$\lambda_{\tau_l^i, t}^{i \rightarrow j} = 0 \quad \text{if} \quad (i, j) \notin E_t \quad \text{or} \quad t \leq \tau_l^i \quad \text{or} \quad t \geq \tau_R^i. \quad (1)$$

Observations

Let us now describe the observation process on which the risk computation relies. We assume that testing individuals for the disease is possible from time $t = 1$, so for any $t \geq 1$ let us denote by \mathcal{D}_t^+ , \mathcal{D}_t^- the set of individuals receiving a positive, respectively negative result at time t . To focus on the most recent and relevant interactions, we introduce the parameter $\zeta \in \mathbb{N}$ corresponding to the contact tracing time-frame. Hence, at a given time $t \geq 0$, the set of observations is provided by the graph of interactions during the recent days $[t - \zeta : t]$, and the set of individuals with a positive and a negative result until $t - 1$. This set includes test results, interactions network, and both individual and interaction attributes. Notice that we keep the list of all the individuals detected since the beginning because we consider that after they get the infection, they stay immune for the period of study. In conclusion, the set of observations at t is defined as

$$\mathcal{O}_t^\zeta = \{\mathcal{G}_l\}_{l \in [t - \zeta : t]} \cup \{\mathcal{D}_l^+\}_{l \in [1 : t - 1]} \cup \{\mathcal{D}_l^-\}_{l \in [1 : t - 1]}.$$

In practice, the spreading of the disease on the network is not available. Unless tested or with reported symptoms, the status and infectiousness of individuals are unknown. In addition, there is some uncertainty in the detection process due to the sensitivity and specificity of the tests, and the co-circulation of other diseases causing similar symptoms. However, here we consider only perfect tests (i.e., test specificity and sensitivity equal to 1) and we assume that all the reported symptoms are a consequence of the disease under study.

In particular, the infection and removal times of individuals are never known even for the detected individuals. These quantities are necessary to compute the probability of transmission from a possibly infectious individual i to a presumably susceptible j at time t , as well as the severity coefficient b_t^i , therefore we approximate them. More precisely, we approximate the probability distributions of τ_l^i and τ_R^i , and for the sake of simplicity, we keep the notations τ_l^i and τ_R^i to represent the random variables issued from the approximated distributions. Depending on the proposed contact tracing approach, we approximate the distribution of τ_l^i by a Dirac measure $\delta_{\widehat{\tau}_l^i}$ or a generalized truncated geometric distribution, as seen later in Equations (6) and (7). Similarly, we approximate the distribution of τ_R^i by a Dirac distribution $\delta_{\widehat{\tau}_R^i}$. For convenience, the quantities $\widehat{\tau}_l^i$ and $\widehat{\tau}_R^i$ are defined in $\mathbb{N} \cup \{+\infty\}$ and serve as estimations of τ_l^i , τ_R^i . By default, at time 0, we set $\widehat{\tau}_l^i = \widehat{\tau}_R^i = +\infty$ for all $i \in V$. We briefly describe two distinct situations that we have at any time t .

1. If $i \in \mathcal{D}_t^+$, we update the values of the estimators $\widehat{\tau}_l^i$ and $\widehat{\tau}_R^i$ to finite values computed from the observations. The details on the definition of the finite candidates for $\widehat{\tau}_l^i$ are provided later. We set $\widehat{\tau}_R^i = \widehat{\tau}_l^i + \beta$, where β is a positive integer that is chosen greater than the mean duration of infectiousness ($\beta = 21$ in our simulations).
2. If $i \notin \mathcal{D}_t^+$, the values of $\widehat{\tau}_l^i$ and $\widehat{\tau}_R^i$ stay equal to their previous values.

We denote by \widehat{b}_t^i the estimation of the parameter b_t^i at time t . If $i \notin \mathcal{D}_{[1:t]}^+$, we consider $\widehat{b}_s^i = \text{“asymptomatic”}$ for $0 \leq s \leq t$. On the other hand, if i is detected at time t , that is $i \in \mathcal{D}_t^+$, we update \widehat{b}_l^i as the real value b_l^i for $l \in [\widehat{\tau}_l^i : \widehat{\tau}_R^i]$ since we assume that when an individual is detected, the severity of the symptoms experienced by this individual is known.

To keep the trace of the negative test results, we define at any t and for any individual j , the day θ_t^j as the last date before t on which j receives a negative test result,

$$\theta_t^j = \max \{l \in [1 : t] : j \in \mathcal{D}_l^-\}, \text{ and by convention } \max \phi = 0.$$

Hence, only the interactions of j that are posterior to θ_t^j are considered risky, meaning that, before a negative test result, the probability that j has been infected is zero.

Risk of infection via transmission chains

We propose two approaches to compute the risk, based on two different degrees of interactions. To differentiate both methods, we call them in the sequel 1° contact tracing (1° CT) and 2° contact tracing (2° CT). In the first approach, the risk of infection is based on 1° interactions, while the second proposes a more accurate risk of infection, defined from both 1° and 2° interactions.

For any time t and any individual $j \notin \mathcal{D}_{[1:t-1]}^+$, our aim is to estimate the probability of infection of j given the set of observations \mathcal{O}_t^ζ , that is

$$\mathbb{P} \left(X_t^j = I \mid \mathcal{O}_t^\zeta \right). \quad (2)$$

We introduce a new truncation parameter $\gamma \in \mathbb{N}$, such that $\gamma \leq \zeta$, corresponding to the infection time-frame of interest. More precisely, for any $j \notin \mathcal{D}_{[1:t-1]}^+$, we approximate the probability in (2) by “ $R_{\gamma,\zeta}^j(t)$,” which is defined as the probability of j being infected in the interval $[t-\gamma : t]$ given the set of observations at time t , that is

$$R_{\gamma,\zeta}^j(t) = \mathbb{P} \left(\bigcup_{l=t-\gamma}^t \{ \tau_l^j = I \} \mid \mathcal{O}_t^\zeta \right). \quad (3)$$

The risk given by Equation (3) can be expressed as the probability for individual j of having been infected by any of the possible sources of transmission in the time interval $[t-\gamma : t]$, given the observations. Hence, Equation (3) can be rewritten as

$$R_{\gamma,\zeta}^j(t) = 1 - \prod_{i \in V: i \neq j} \mathbb{P} \left(\bigcap_{l=t-\gamma}^t \{ Y_l^{ij} = 0 \} \mid \mathcal{O}_t^\zeta \right), \quad (4)$$

where

$$Y_l^{ij} = \begin{cases} 1, & \text{if } i \text{ infects } j \text{ at time } l, \\ 0, & \text{otherwise.} \end{cases}$$

Then, for any possible individual i considered as a source, we can use the law of total probability with respect to the date of infection of i , which leads to

$$\mathbb{P} \left(\bigcap_{l=t-\gamma}^t \{ Y_l^{ij} = 0 \} \mid \mathcal{O}_t^\zeta \right) = \sum_{d \in \Delta} \mathbb{P} \left(\bigcap_{l=t-\gamma}^t \{ Y_l^{ij} = 0 \} \mid \tau_l^i = d, \mathcal{O}_t^\zeta \right) \mathbb{P} \left(\tau_l^i = d \mid \mathcal{O}_t^\zeta \right), \quad (5)$$

with $\Delta = [1 : t - 1] \cup \{+\infty\}$. By independence of the events $\left\{Y_l^{ij} = 0 \mid \tau_l^i = d, \mathcal{O}_t^\zeta\right\}_{l \in [t-\gamma:t]}$ we have

$$\mathbb{P}\left(\bigcap_{l=t-\gamma}^t \left\{Y_l^{ij} = 0\right\} \mid \tau_l^i = d, \mathcal{O}_t^\zeta\right) = \prod_{l=t-\gamma}^t \mathbb{P}\left(Y_l^{ij} = 0 \mid \tau_l^i = d, \mathcal{O}_t^\zeta\right),$$

where we assume

$$Y_l^{ij} \mid \tau_l^i = d, \mathcal{O}_t^\zeta \sim \text{Ber}\left(\lambda_{d,l}^{i \rightarrow j} \mathbb{1}(l > \theta_l^i)\right).$$

The way we model $\tau_l^i \mid \mathcal{O}_t^\zeta$ depends on the contact tracing method under consideration and it is explained in the following sections.

Infection risk for 1° contact tracing. We describe now how to compute the risk defined by Equation (4) for any j such that $j \notin \mathcal{D}_{[1:t-1]}^+$ using the 1°CT approach that considers as possible sources of infection only the index cases. At a given time t , an individual i that has been detected up to t , is considered as an *index case* for any time l between the respective estimated infection and removal times. Thus, we define the set of index cases at time l given the set of observations \mathcal{O}_t^ζ as

$$\mathcal{I}_{l,t} = \{i \in \mathcal{D}_{[1:t-1]}^+ : l \in]\widehat{\tau}_l^i, \widehat{\tau}_R^i[\}.$$

For the 1°CT approach, we only take into account the interactions with index cases that occur in the interval $[t - \gamma : t]$. Consequently, the set of observations reduces to

$$\mathcal{O}_t^{\gamma, 1^\circ} = \left\{ \mathcal{G}_{l,t}^{1^\circ} \right\}_{l \in [t-\gamma:t]} \cup \left\{ \mathcal{D}_l^+ \right\}_{l \in [1:t-1]} \cup \left\{ \mathcal{D}_l^- \right\}_{l \in [1:t-1]} \subset \mathcal{O}_t^\zeta,$$

where $\mathcal{G}_{l,t}^{1^\circ} = (\mathcal{V}_{l,t}^{1^\circ}, \mathcal{E}_{l,t}^{1^\circ}) \subset \mathcal{G}_l$, $\mathcal{E}_{l,t}^{1^\circ}$ corresponds to the set edges $E_{l,t}^{1^\circ}$ complemented by their attributes and with $E_{l,t}^{1^\circ}$ being composed of the 1° interactions at l , that is

$$E_{l,t}^{1^\circ} = \left\{ (i, j) \in E_l : i \in \mathcal{I}_{l,t}, j \notin \mathcal{D}_{[1:t-1]}^+ \text{ and } \theta_l^i < l \right\}.$$

In addition, the set $\mathcal{V}_{l,t}^{1^\circ}$ corresponds to the vertices in $E_{l,t}^{1^\circ}$ complemented by their attributes.

As we mentioned before, the time of infection of index cases is inferred from the observations and is defined as

$$\mathbb{P}\left(\tau_l^i = d \mid \mathcal{O}_t^\zeta\right) = \mathbb{P}\left(\tau_l^i = d \mid \mathcal{O}_t^{\gamma, 1^\circ}\right) = \delta_{\widehat{\tau}_l^i}(d). \quad (6)$$

As a consequence, Equation (5) becomes

$$\mathbb{P}\left(\bigcap_{l=t-\gamma}^t Y_l^{ij} = 0 \mid \mathcal{O}_t^\zeta\right) = \mathbb{P}\left(\bigcap_{l=t-\gamma}^t Y_l^{ij} = 0 \mid \tau_l^i = \widehat{\tau}_l^i, \mathcal{O}_t^{\gamma, 1^\circ}\right).$$

Finally, for the 1°CT method, the risk given by (4) for j at t is defined as

$$R_Y^{j,1^\circ}(t) = 1 - \prod_{i \in V: i \neq j} \prod_{l=t-\gamma}^t \left(1 - \lambda_{\tau_l, l}^{i \rightarrow j} \mathbb{1}(l > \theta_l^j)\right).$$

Remind that, as considered in Equation (1), $\lambda_{\tau_l, l}^{i \rightarrow j} = 0$ if $\tau_l^i = \infty$.

Infection risk for 2° contact tracing. Here, we derive the computation of the risk defined by Equation (4) by considering as possible sources of transmission both index cases and 1° contacts. For any individual j such that $j \notin \mathcal{D}_{[1:t-1]}^+$, when the possible source is an index case, the risk computation is analogous to the one developed for the 1° CT method. On the other hand, when the possible source is a 1° contact i (such that $i \notin \mathcal{D}_{[1:t-1]}^+$), we use the parameter ζ ($\zeta \geq \gamma$) as the time-frame for the infection date of i , meaning that it lies in the interval of time $[t-\zeta : t-1]$. As a consequence, we are interested in 1° interactions that occur in $[t-\zeta : t-1]$ and in the 2° interactions that occur after a possible transmission due to a 1° interaction in the interval of time $[t-\gamma : t]$.

Hence, the set of observations \mathcal{O}_t^ζ reduces to $\mathcal{O}_t^{\zeta, \gamma, 2^\circ}$ ($\mathcal{O}_t^{\zeta, \gamma, 2^\circ} \subset \mathcal{O}_t^\zeta$),

$$\left\{\mathcal{G}_{l,t}^{1^\circ}\right\}_{l \in [t-\zeta:t]} \cup \left\{\mathcal{G}_{l,t}^{2^\circ}\right\}_{l \in [t-\gamma:t]} \cup \left\{\mathcal{D}_l^+\right\}_{l \in [1:t-1]} \cup \left\{\mathcal{D}_l^-\right\}_{l \in [1:t-1]},$$

where $\mathcal{G}_{l,t}^{1^\circ} = (\mathcal{V}_{l,t}^{1^\circ}, \mathcal{E}_{l,t}^{1^\circ})$ is defined in ‘Infection risk for 1° contact tracing’ Section, $\mathcal{G}_{l,t}^{2^\circ} = (\mathcal{V}_{l,t}^{2^\circ}, \mathcal{E}_{l,t}^{2^\circ}) \subset \mathcal{G}_l$, $\mathcal{E}_{l,t}^{2^\circ}$ corresponds to the set of edges $E_{l,t}^{2^\circ}$ complemented by their attributes, and with $E_{l,t}^{2^\circ}$ being composed of the 2° interactions at l , i.e.

$$E_{l,t}^{2^\circ} = \left\{(i, j) \in E_l : \exists k \in V \text{ s.t. } (k, i) \in \bigcup_{d \in [t-\zeta:l-1]} E_{d,t}^{1^\circ}, j \notin \mathcal{D}_{[1:t-1]}^+ \text{ and } \theta_l^j < l\right\}.$$

The set $\mathcal{V}_{l,t}^{2^\circ}$ is composed of the vertices in $E_{l,t}^{2^\circ}$ complemented by their attributes.

Let us now consider an individual i that has not been detected up to time $t-1$. We remind that, θ_l^i is defined as the last time, before t , of a negative result test for i . On the other hand, $t - \zeta$ is considered as the first possible time of infection of i . Hence, we denote by $\theta_{t,\zeta}^i = (\theta_t^i + 1) \vee (t - \zeta)$ the first possible time of infection for i , where the notation “ $t_1 \vee t_2$ ” stands for the maximal term between t_1 and t_2 . Then, to bring together all the possible sources of infections (index cases and others), we model the distribution of the time of infection $\mathbb{P}(\tau_l^i = d \mid \mathcal{O}_t^{\zeta, \gamma, 2^\circ})$ of any individual $i \in V$, and any time $d \in \Delta = [1 : t-1] \cup \{+\infty\}$ as follows,

$$\delta_{\tau_l^i}(d) \mathbb{1}(i \in \mathcal{D}_{[1:t-1]}^+) + g_{t,\zeta}^i(d) \mathbb{1}(i \notin \mathcal{D}_{[1:t-1]}^+), \quad (7)$$

where $g_{t,\zeta}^i(d)$ is the probability mass function of a truncated generalized geometric in $[\theta_{t,\zeta}^i : t-1] \cup \{+\infty\}$, defined as

$$\mathbb{1}(\theta_{t,\zeta}^i \leq d \leq t-1) p_{d,t}^i \prod_{l=\theta_{t,\zeta}^i}^{d-1} (1 - p_{l,t}^i) + \mathbb{1}(d = \infty) \prod_{l=\theta_{t,\zeta}^i}^{t-1} (1 - p_{l,t}^i).$$

We denote by $p_{l,t}^i$ the probability that i gets infected by some index case at time l , given the set of observations $\mathcal{O}_t^{\zeta,\gamma,2^\circ}$, that is

$$p_{l,t}^i = 1 - \prod_{k \in V: k \neq i} \mathbb{P}\left(Y_l^{ki} = 0 \mid \tau_l^k = \tau_l^k, \mathcal{O}_t^{\zeta,\gamma,2^\circ}\right).$$

For the 2° CT method, the risk of infection $R_{\gamma,\zeta}^{j,2^\circ}(t)$ for an individual j at t is defined as,

$$R_{\gamma,\zeta}^{j,2^\circ}(t) = 1 - \prod_{i \in V: i \neq j} \sum_{d \in \Delta} \prod_{l=t-\gamma}^t \mathbb{P}\left(Y_l^{ij} = 0 \mid \tau_l^i = d, \mathcal{O}_t^{\zeta,\gamma,2^\circ}\right) \mathbb{P}\left(\tau_l^i = d \mid \mathcal{O}_t^{\zeta,\gamma,2^\circ}\right).$$

Simulation results

Simulated data come from OpenABM-Covid19 model which code is available on <https://github.com/aleingrosso/OpenABM-Covid19>. Our code is available on https://github.com/gbayolo26/risk_estimation.

Simulated data

To test our proposed method on a proper data set, we generate the data using the OpenABM-Covid19 model introduced by [27]. This agent-based model simulates the spread of the COVID-19 disease on a sequence of contact networks representing the daily interactions within a population whose demographic structure is based upon UK census data. This model has several advantages: (1) it is rich enough to mimic a dynamic social contact network at the level of a real country, with possible large population size and a variety of individual information, in particular, the daily interactions between individuals come from three different networks depicting the contacts at home, at work and at random; (2) concerning the disease spreading, several infected statuses are available ranging from asymptomatic to pre-symptomatic statuses to mild or severe symptomatic, where the pre-symptomatic status refers to infectious individuals without symptoms; (3) it has several implementation advantages such as a very fast running time and the fact that new intervention methods, like ours, can be easily integrated into the existing code.

In the OpenABM-Covid19 model, the transmission probability takes into account the infectiousness of the source (day of infection, disease severity according to the status), the susceptibility of the recipient based on the age group, and the place where the interactions occur, putting more weight to the household interactions than the others. Indeed, the transmission probability $\lambda_{\tau_l^i,t}^{i \rightarrow j}$ that i infects j at time t is defined as

$$1 - \exp\left(-L f_A(a^j) f_B(b_t^i) f_C(c_t^{ij}) \int_{t-\tau_{l-1}^i}^{t-\tau_l^i} f_\Gamma(u; \mu, \sigma^2) du\right), \quad (8)$$

where

1. f_Γ accounts for the varying infectiousness over the course of the disease, and it is chosen as the density function of the Gamma distribution with mean μ and standard deviation σ ,
2. $f_B(b_t^i)$ is the severity of the individual i considered as a possible source at time t (i can be susceptible, asymptomatic, pre-mild, mild, pre-severe, severe or removed),

3. $f_A(a^j)$ is the relative susceptibility of the recipient j , which depends on the age group of j , and is normalized by the mean number of daily interactions by age group,
4. $f_C(c_t^{ij})$ is the strength of the interaction (if it is at home, work or random) between i and j at t ,
5. L scales the overall infection rate.

For more details on the OpenABM-Covid19 model, and in particular, on the functions f_A , f_B and f_C , the interested reader can refer to [S1 appendix](#) and [27].

Intervention strategy

The simulation starts at time $t = 0$ with N individuals. At the beginning, all individuals are susceptible (S), except for a small number N_0 of infectious individuals (“patients zero”). Every day, starting from $t = 1$, a proportion p_s , respectively p_m , of individuals with newly developed severe and mild symptoms are tested, detected and quarantined. Later, at a fixed date t_0 in $[1 : T]$, the intervention based on the risk calculation starts, and it is carried out daily until the mitigation of the epidemic or the end of the study. At any $t \geq t_0$, the intervention strategy based on the 2°CT method consists of tracing 1° and 2° contacts, computing their risk of infection, and ranking them according to their risk values. See [Fig 1](#) for an illustration of how the proposed 2°CT method works for a simple scenario of three days. Then, the first η individuals in the ranking are tested, and the newly detected ones become index cases and are quarantined. The default quarantine protocol stops the interactions in the occupation and random network, but those within the household are maintained. On a given day, it may happen that the number of traced individuals is smaller than η , in which case we randomly select and test individuals among those who have not been detected, until reaching the number of η daily available tests; for those detected, we set their time of infection at γ days prior to their detection (since we do not have information to infer their time of infection from previous contacts with detected individuals). As already mentioned, the tests are assumed to be perfect, and we suppose that the test results are available the same day on which the tests are performed.

Notice that the 1°CT method is a particular case of the 2°CT method, and hence, the intervention related to the 1°CT method is analogous to the one for the 2°CT method, except that only the 1° contacts are traced and ranked. The implementation of these methods could be automated via contact tracing applications (DCT), when this tool is available, as follows: (1) data collection, namely index cases record and share their contact data anonymously, 1° contacts are traced and their subsequent contacts (2° contacts) are identified; (2) risk calculation by the DCT app for each traced individual using the predefined risk formula, based on their exposure to index cases; (3) risk-based ranking and notification, meaning that individuals are ranked according to their risk values, and those with the highest risks are prioritized for testing and subsequent quarantine; and (4) real-time updates by the app of the risk scores as new cases are reported or additional contacts are traced.

To compute the risk of infection for the 1°CT and 2°CT methods, we use the transmission probability considered in [27] and defined by Equation (8), with \widehat{b}_i^j and $\widehat{\tau}_i^j$ in place of the true values b_i^j and τ_i^j . Due to the way the infectiousness is modeled, we have that $\lambda_{\tau,t}^{i \rightarrow j} \approx 0$ if $t - \tau \geq 15$. The latter, combined with a significant gain of the computational cost of our method, leads us to reduce the set of index cases as follows

$$\mathcal{I}_{l,t} = \{i \in V : l \in]\widehat{\tau}_l^i \vee (t - 21) : \widehat{\tau}_R^i[\}.$$

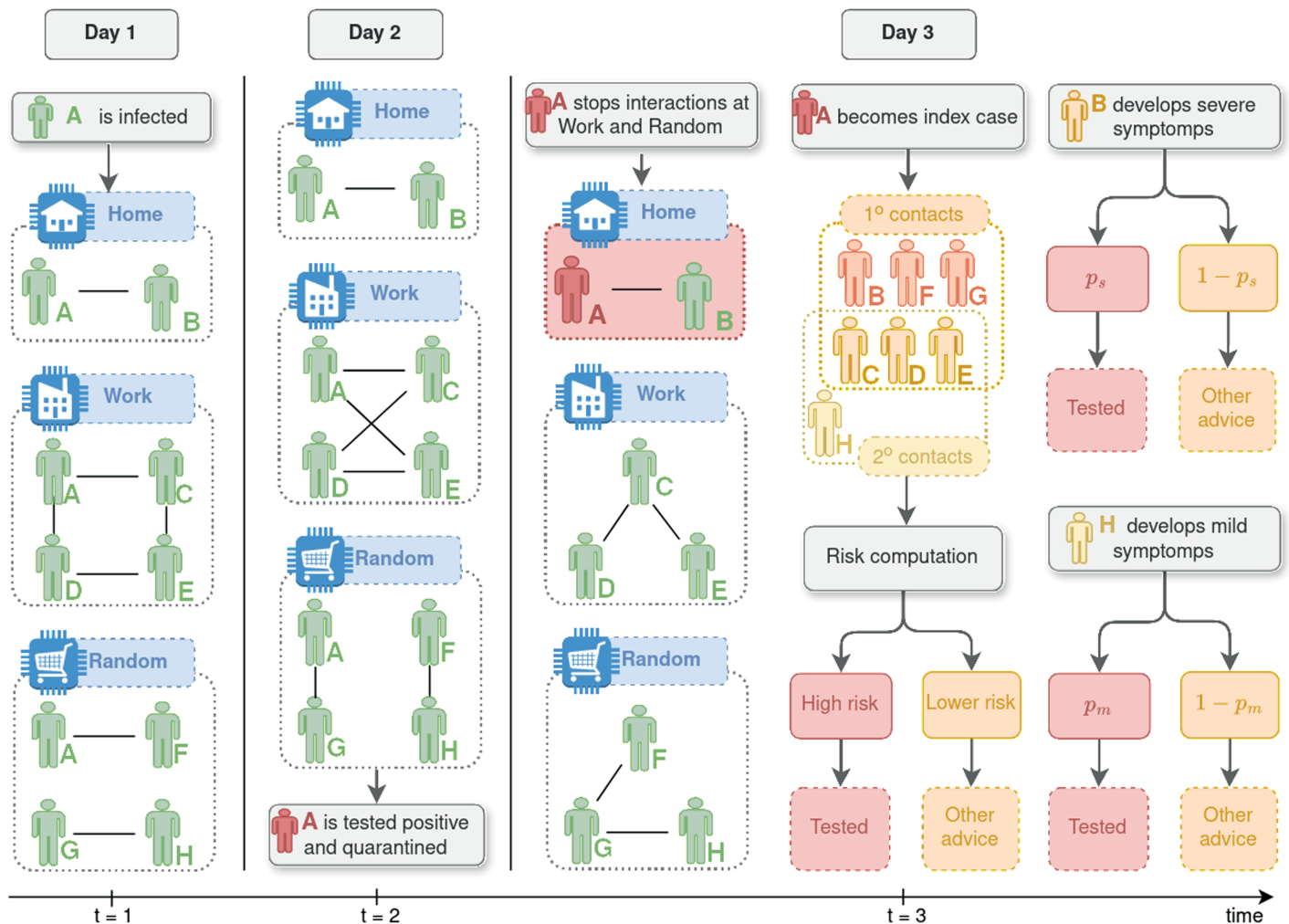


Fig 1. Interactions between individuals over three days. When the individual A is detected, A is quarantined, and their time of infection is estimated. The 2^o contact tracing method traces forward the first and second contacts in interaction with A after the estimated time of infection. The risk for these individuals is then computed, and those with the highest risk are selected for testing.

<https://doi.org/10.1371/journal.pone.0320291.g001>

Results

In this section, we present the results obtained using the intervention based on the risk, through the simulation of different scenarios. For all the simulations, the propagation of the epidemic is identical until t_0 , while it might change after t_0 , when the intervention method starts, depending on the particular scenario. In the figures presented later, each thin line represents the result obtained for the realization associated with one *seed*, while the thick lines correspond to the average of all the realizations.

Estimation of the time of infection. As seen before, the estimated time of infection of index cases has a direct impact and an indirect impact on the computation of the risk of infection since (1) the probability of transmission depends on it and (2) the selection of the risky interactions relies on it. The estimation of the time of infection is computed on the day of detection of the individual, and it remains constant over time after this day. Remind that, for any individual j who has not been detected, we have set $\hat{\tau}_j^i = \infty$.

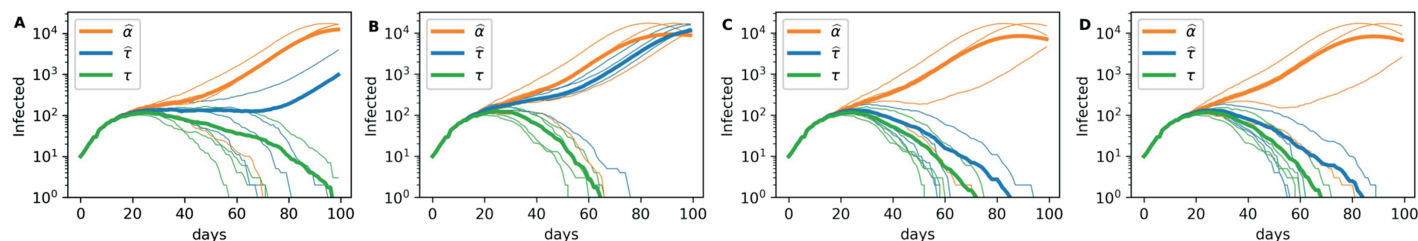


Fig 2. Effect of the estimators $\hat{\tau}_I^j$ (blue), $\hat{\alpha}_I^j$ (orange), and the real time of infection τ_I^j (green). (A) The 1°CT method with parameter $\gamma = 6$. (B) 2°CT with $\gamma = 6$ and $\zeta = 7$. (C) 2°CT with $\gamma = 6$ and $\zeta = 8$. (D) 2°CT with $\gamma = 6$ and $\zeta = 9$. The figures illustrate the impact of these methods on the spread of the epidemic, displaying the number of infectious individuals over $T = 100$ days in a population of size $N = 50K$. The intervention begins on day $t_0 = 12$, with $N_0 = 10$ initial patients (patient zero cases), $\eta = 125$ daily available tests, a proportion $p_s = 1$ of detected severe symptomatic individuals, and a proportion $p_m = 0.75$ of detected mild symptomatic individuals.

<https://doi.org/10.1371/journal.pone.0320291.g002>

At any time $t \geq t_0$, the individuals are tested because of their symptoms or because they have been traced and selected based on their risk of infection. Among the symptomatic individuals, let us denote by m ($m \in \mathbb{N}$) the expected number of days it takes to develop symptoms after the day of infection. In our simulations, we used $m = 6$ as in [27]. Then, if j is detected by symptoms at t we define the approximation of the time of infection as $(t - m) \vee (\theta_I^j + 1)$.

For any individual j detected by risk at t , we define $\hat{\tau}_I^j$ as follows,

$$\hat{\tau}_I^j = \hat{\tau}_I^{j,1^\circ} \mathbb{1}(\hat{\tau}_I^{j,1^\circ} \neq \infty) + \hat{\tau}_I^{j,2^\circ} \mathbb{1}(\hat{\tau}_I^{j,1^\circ} = \infty),$$

where $\hat{\tau}_I^{j,1^\circ}$ and $\hat{\tau}_I^{j,2^\circ}$ are defined respectively as the minimal time of the 1° and 2° interactions of j , lying within the time-frame $[t - \gamma : t]$, that is

$$\begin{aligned} \hat{\tau}_I^{j,1^\circ} &= \min \left\{ d \in [t - \gamma : t] : \exists i \in V \text{ s.t. } (i, j) \in E_{d,t}^{1^\circ} \right\}, \\ \hat{\tau}_I^{j,2^\circ} &= \min \left\{ d \in [t - \gamma : t] : \exists i \in V \text{ s.t. } (i, j) \in E_{d,t}^{2^\circ} \right\}, \end{aligned}$$

where, by convention we set $\min \emptyset = \infty$.

To evaluate the effectiveness of the estimator $\hat{\tau}_I^j$ in the mitigation strategy, we propose another estimator $\hat{\alpha}_I^j$ for the individuals detected by risk, which is constant for all index cases provided that they do not have a previous negative test result, that is

$$\hat{\alpha}_I^j = (t - \gamma) \vee (\theta_I^j + 1).$$

To compare the effect of $\hat{\tau}_I^j$, $\hat{\alpha}_I^j$, and τ_I^j (the real infection time of i), on the mitigation of the epidemic, we simulate the same intervention strategy with these three different times of infection and for different sets of parameter values. We depict in Fig 2 the number of infectious individuals in logarithmic scale through time, and for the same simulated trajectories, we display in Fig 3 the box-plots of the empirical distribution of the differences $\tau_I^j - \hat{\alpha}_I^j$ and $\tau_I^j - \hat{\tau}_I^j$.

The results in Fig 2 show that, for a broad range of parameter values, the use of the estimated infection time $\hat{\tau}_I^j$ is more effective in the mitigation of the epidemic than the constant estimator $\hat{\alpha}_I^j$. This improvement is more pronounced in panels C and D, in which $\hat{\tau}_I^j$ provides results that are almost as good as the ones obtained with τ_I^j , the true time of infection. If we take a look at the corresponding box-plots in Fig 3, we can see that the empirical distribution

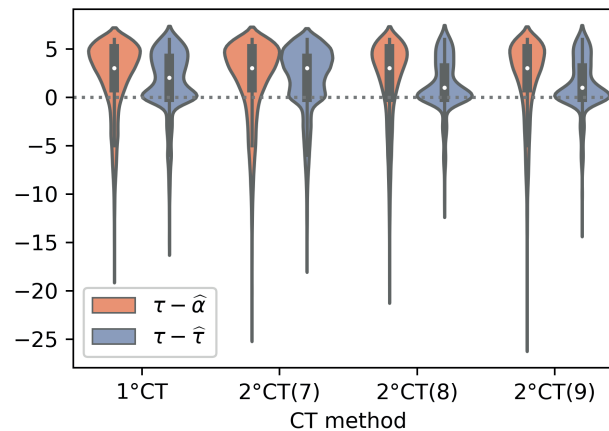


Fig 3. Comparison of infection time estimators. Box plots of the differences $\tau_i^i - \hat{\alpha}_i^i$ (orange) and $\tau_i^i - \hat{\alpha}_i^{\tilde{i}}$ (blue) for individuals detected by risk using the 1°CT and 2°CT methods, considering the parameter range $\zeta = 7, 8, 9$ and $\gamma = 6$. Each box plot was generated using four seeds, with parameters $T = 100$, $N = 50K$, $t_0 = 12$, $N_0 = 10$, $\eta = 125$, $p_s = 1$, and $p_m = 0.75$.

<https://doi.org/10.1371/journal.pone.0320291.g003>

of $\tau_i^i - \hat{\alpha}_i^{\tilde{i}}$ is more concentrated around the empirical median, and hence, has less variability, than the one of $\tau_i^i - \hat{\alpha}_i^i$. Moreover, for $\zeta = 8, 9$, the empirical median of $\tau_i^i - \hat{\alpha}_i^{\tilde{i}}$ is much closer to zero than the one of $\tau_i^i - \hat{\alpha}_i^i$. In these latter cases, the good performance of the estimator $\hat{\alpha}_i^{\tilde{i}}$ allows reaching the mitigation of the epidemic faster than the constant estimator (see panels C–D in Fig 2).

Time-frame for the time of infection. In a context of limited resources, providing an effective strategy for mitigating an epidemic goes through the detection at an early stage of the most likely infected individuals. Indeed, the detection of the latter before they become highly infectious is preferable. Hence, tuning the parameter γ , which corresponds to the considered infection time-frame for individuals at risk, plays a key role in providing an effective strategy for the mitigation of the epidemic. There is a trade-off between large and small values of γ . For large values of γ , one expects to detect more individuals since, by construction, the set of observations increases with time. On the other hand, small values of γ allow us to concentrate the efforts on the more recently infected individuals, before they propagate the disease, and discard those individuals that were infected a long time ago.

In Fig 4, we study the impact of the values of γ on the mitigation of the epidemic for 1°CT and 2°CT methods, showing the number of infectious individuals in logarithmic scale through time. For these simulations we keep ζ constant with respect to γ (the choice of the value of the parameter ζ is further discussed in ‘Comparison with other ranking methods’ Section). In Fig 5, we focus on the effect of γ on the early or late detection of the individuals by displaying the box-plots of the difference between the time of detection and the time of infection for the individuals detected by risk. Each box-plot in Fig 5 has been built from the same simulations presented in the Fig 4.

Fig 5 shows that, for both methods, the detection of infected individuals is faster with a small value of γ ; this enables a faster progression of the epidemic mitigation as seen in Fig 4. From both figures, we conclude that taking $\gamma = 6$ in the two methods makes them effective in detecting and quarantining individuals at an early stage of infection and improves the allocation of resources.

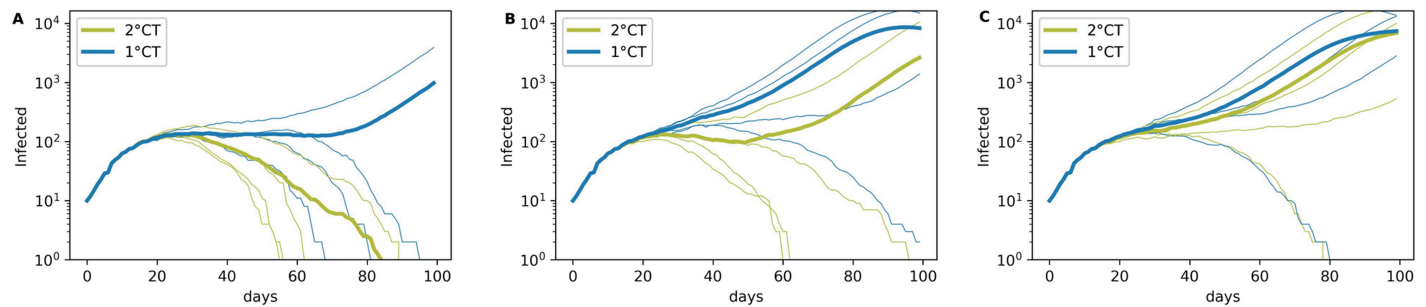


Fig 4. Effect of γ on epidemic spread. Effect of γ on epidemic spread for strategies 1°CT (blue) and 2°CT (yellow) when $\zeta = \gamma + 3$, with (A) $\gamma = 6$, (B) $\gamma = 10$, and (C) $\gamma = 14$. Each plot was generated using four seeds, with parameters $T = 100$, $N = 50K$, $t_0 = 12$, $N_0 = 10$, $\eta = 125$, $p_S = 1$, and $p_m = 0.75$.

<https://doi.org/10.1371/journal.pone.0320291.g004>

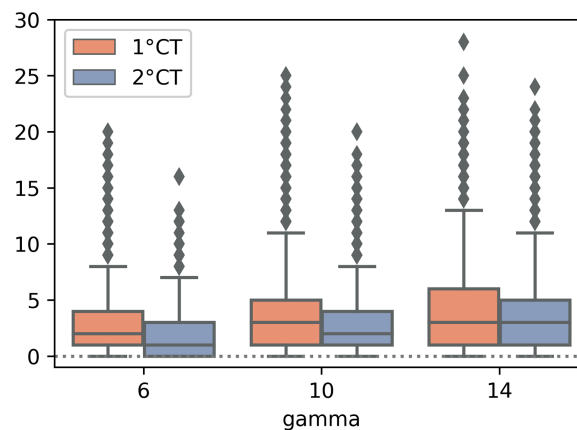


Fig 5. Box plots of infection and detection time differences. Box plots of the differences between the time of infection and the time of detection for individuals detected by risk using the 1°CT and 2°CT methods, with parameter values $\gamma = 6, 10, 14$, and for 2°CT with $\zeta = \gamma + 3$. The simulations were performed with $T = 100$, $N = 50K$, $t_0 = 12$, $N_0 = 10$, $\eta = 125$, $p_S = 1$, and $p_m = 0.75$, using four different seeds.

<https://doi.org/10.1371/journal.pone.0320291.g005>

Probability of transmission. The probability of transmission plays also a role in the ability of the contact tracing strategy to mitigate the epidemic. For the 1°CT and 2°CT methods, the probability of transmission is involved in the risk of infection calculated at t for an individual j and depends directly on the estimation of the time of infection (see Equation (8)).

Here we compare the 1°CT and 2°CT methods for a range of values of the pair “probability of transmission and real or estimated time of infection,” i.e., for $(\lambda_{\tau_i}^{i \rightarrow j}, \tau_i^j)$, $(\lambda_{\hat{\tau}_i}^{i \rightarrow j}, \hat{\tau}_i^j)$, (p, τ_i^j) and $(p, \hat{\alpha}_i^j)$, where $\lambda_{\tau_i}^{i \rightarrow j}$ is defined by Equation (8), $\hat{\tau}_i^j$, $\hat{\alpha}_i^j$ are defined later, τ_i^j is the true infection time of i and $p = 1/2$ is a constant. For all of these pairs, Fig 6 depicts the number of infectious individuals in logarithmic scale through time, for $\gamma = 6$ and $\zeta = 9$.

Both panels in Fig 6 show that the results given by the pair $(p, \hat{\tau}_i^j)$ (in orange) are much better than the ones obtained with $(p, \hat{\alpha}_i^j)$ (in yellow), confirming that it is crucial to accurately estimate the infection time. Moreover, the probability of transmission that depends on the individual and interaction attributes (in dark blue) considerably improves the results compared with the choice of a constant probability of transmission.

In Fig 6A, the results provided by the CT method are included; the CT method consists of ranking individuals according to their number of interactions with detected individuals

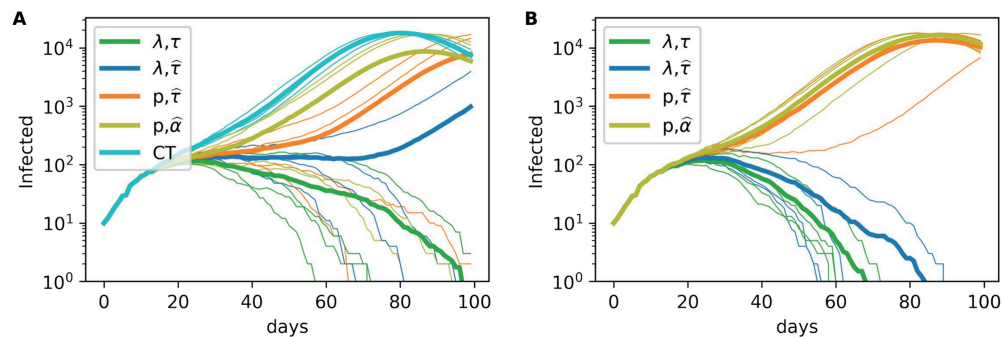


Fig 6. Effect of the transmission probability function on epidemic spread. Effect of the transmission probability function on the spreading of the epidemic for the (A) 1°CT and (B) 2°CT methods. Each plot was generated with $T = 100$, $N = 50K$, $t_0 = 12$, $N_0 = 10$, $\eta = 125$, $p_m = 0.75$, $p_s = 1$, $\gamma = 6$, and for 2°CT with $\zeta = 9$.

<https://doi.org/10.1371/journal.pone.0320291.g006>

in the time-frame $[t - \gamma : t]$. It is important to highlight that the CT method differs from the 1°CT method with $(p, \hat{\alpha}_t^i)$, since in the CT method the dates of the negative test results are not taken into account. Fig 6A shows that including the information on negative test results has a positive effect on the mitigation of the epidemic.

Comparison with other ranking methods. To evaluate the efficiency of the proposed 1°CT and 2°CT methods to mitigate an epidemic, we compare them with three other ranking strategies:

1. Random Selecting (RS): individuals not previously detected are ranked randomly.
2. Contact Tracing (CT): individuals not previously detected are ranked according to their number of interactions with detected individuals in the time-frame $[t - \gamma : t]$.
3. Mean-Field (MF): individuals not previously detected are ranked according to the mean-field risk approximation presented in [2].

We fix the parameters $\gamma = 6$ and $\zeta = 7, 8, 9$ (indicated in the legend as 2°CT(7), 2°CT(8), 2°CT(9), respectively). The values for the parameters in MF strategy are $\rho_{MF} = 5$ and $t_{MF} = 10$.

The mean-field procedure depends on two parameters: (1) ρ_{MF} , the mean time elapsed between the time of infection and the time of detection and (2) t_{MF} , a parameter called *integration time* on the MF method, meaning that given new observations, the probabilities are updated in the interval $[t - t_{MF} : t]$. For the following simulations we consider $\rho_{MF} = 5$ and $t_{MF} = 10$, as in [2]. In S2 appendix, we provide further details about the main differences in the computation of the mean-field risk and the 2°CT risk.

We compare the five strategies in Fig 7, in which we display the number of infectious individuals in logarithmic scale through time across a broad range of values for the parameters. In particular, we increase the number of daily available tests from the left panels to the right ones, and we increase the proportion of daily detected mild symptomatic individuals from top to bottom. As expected for all strategies, a higher value of p_m and/or η improves the mitigation of the epidemic in terms of the duration and the total number of infected individuals. The simulations show that our proposed methods (1°CT and 2°CT) improve considerably the results compared to the MF and the usual CT, which are all better than the RS strategy. The latter does not mitigate the epidemic even with a high number of daily available tests and a high value of p_m , while the MF and CT methods achieve the mitigation for a large value of η . We also study the 2°CT method for different time-frames in which the

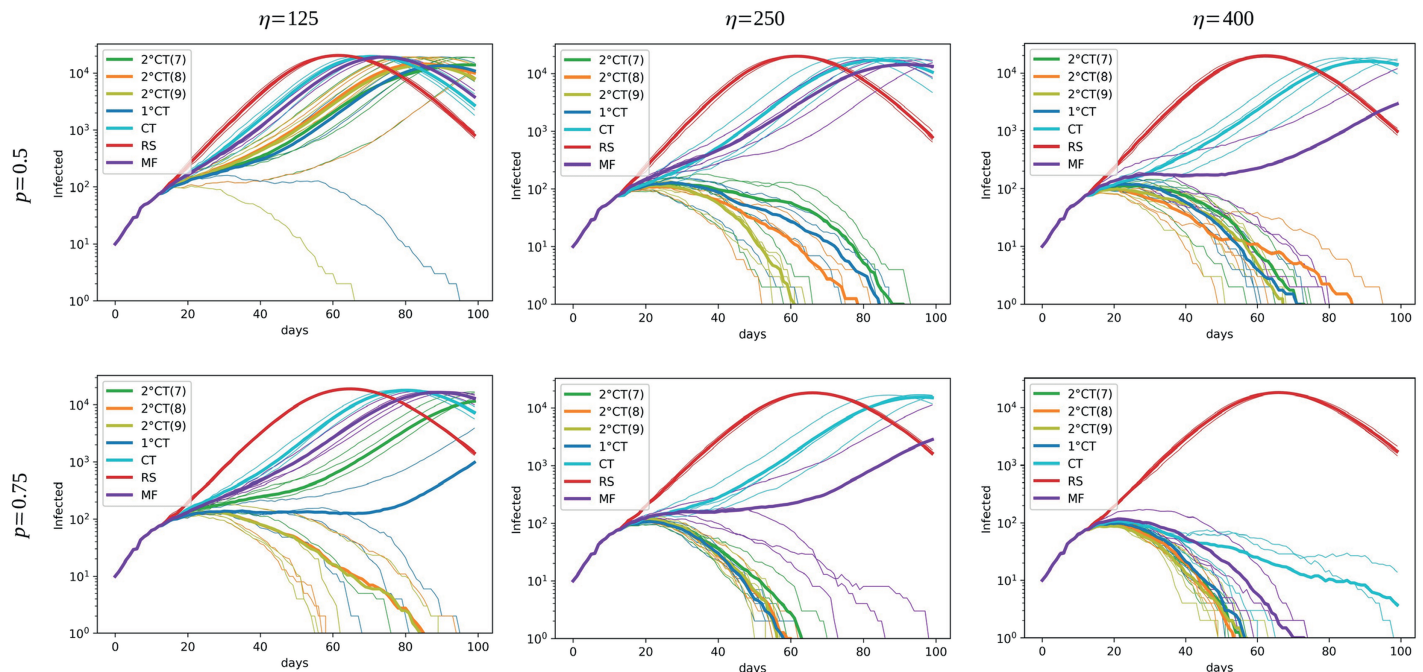


Fig 7. Effect of η and p_m on epidemic spread. Effect of the parameters η (the number of daily available tests, increasing from left to right) and p_m (the proportion of daily detected individuals with mild symptoms, increasing from top to bottom) on epidemic spread for the strategies 1°CT, 2°CT, CT, RS, and MF. In all simulations, we consider $T = 100$, $N = 50K$, $t_0 = 12$, $N_0 = 10$, and $p_S = 1$. The estimation of the infection time for the 1°CT and 2°CT strategies is given by $\hat{\tau}_i^j$. We fix the parameters $\gamma = 6$ and $\zeta = 7, 8, 9$ (indicated in the legend as 2°CT(7), 2°CT(8), and 2°CT(9), respectively). The parameter values for the MF strategy are $\rho_{MF} = 5$ and $t_{MF} = 10$.

<https://doi.org/10.1371/journal.pone.0320291.g007>

1°contact can get infected, that is in $[t - \zeta : t]$, where we consider $\zeta = 7$ in green, $\zeta = 8$ in orange and $\zeta = 9$ in yellow. Fig 7 shows that the results are improved as ζ increases. In particular, it should be noticed that the 2°CT method with $\zeta = 8$ and $\zeta = 9$ gives better results than the 1°CT method. However, the 1°CT method requires less individual information and therefore it is better in terms of privacy restrictions. From these results, a trade-off can arise between getting better results with computationally demanding (2°CT method) and preserving individual privacy with simpler and faster computation. Indeed, it is worth mentioning that for a high enough number of daily available tests and/or a high enough proportion of mild observed, the methods 1°CT and 2°CT have similar effects on the mitigation of the epidemic; hence in this case, we recommend the use of the 1°CT method than the 2°CT method.

Relaxing assumptions. To better align our model with real-world conditions, we relax several key assumptions, including a fixed delay between symptom onset and testing, perfect tests, and complete quarantine adherence. Previously, we assumed diagnostic tests with 100% sensitivity and specificity. While this simplification allowed us to focus on the structural aspects of our model, it does not reflect real-world conditions where diagnostic tests often have lower sensitivity and specificity. To address this, we have expanded our analysis to consider the practical implications of realistic test performance. Specifically, we evaluate how varying levels of sensitivity and specificity impact the effectiveness of our risk-based prioritization strategy. Likewise, we assumed before that all individuals identified for isolation or quarantine adhered fully to these measures. However, in real-world settings, not all individuals comply with quarantine recommendations. In this extended analysis, we relax this assumption by introducing variability in quarantine adoption rates. Specifically, we simulate

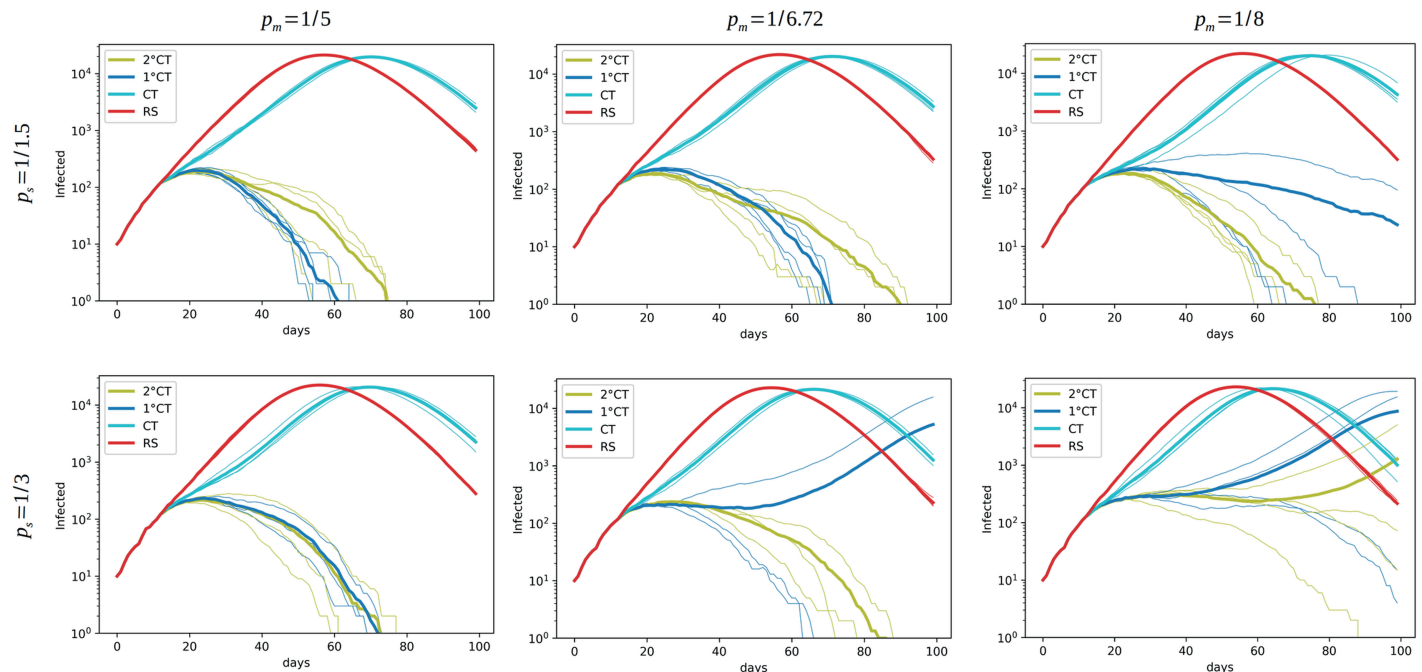


Fig 8. Effect of p_m and p_s on epidemic spread. Effect of the parameters p_m (the daily detection probability of individuals with mild symptoms, decreasing from left to right) and p_s (the daily detection probability of individuals with severe symptoms, decreasing from top to bottom) on epidemic spread for the strategies 1°CT, 2°CT, CT, and RS. Each plot is generated from simulations using four seeds, with parameters $T = 100$, $N = 50K$, $t_0 = 12$, $N_0 = 10$, and $\eta = 400$. The estimation of the infection time for the 1°CT and 2°CT strategies is given by $\widehat{\tau}_i^j$. We fix the parameters $\gamma = 6$ and $\zeta = 9$.

<https://doi.org/10.1371/journal.pone.0320291.g008>

scenarios where only a fraction of identified individuals adopt quarantine, reflecting different levels of public adherence.

Mean time to detection based on symptoms

Initially, the timing between symptom onset and testing was assumed to be fixed; here, we model it explicitly using a geometric distribution to capture the variability in detection times. This distribution reflects real-world scenarios in which testing time is influenced by factors such as severity of symptoms, access to healthcare, and individual behavior. By incorporating this variability, we aim to better capture the stochastic nature of testing delays and their impact on epidemic dynamics. More precisely, we introduce the following parameters,

- p_s and p_m correspond, for severe and mild individuals respectively, to the daily probabilities of being tested after symptoms onset, which results in a mean delay of $1/p_s$ and $1/p_m$ days between the first signs of the disease and the test.

To evaluate the impact of detection delays for individuals with mild (p_m) and severe (p_s) symptoms, we conducted simulations (see Fig 8) in which these parameters were systematically varied. Based on the review provided by [35], the optimal window for conducting RT-PCR testing is between the first and seventh days after symptom onset, with the highest positive result rate seen at a mean of 6.72 days. Accordingly, in our simulation we decreased p_m from 1/5 (left) to 1/8 (right) and we increased p_s from 1/1.5 (top) to 1/3 (bottom) to observe how variations in these parameters affect the performance of all the intervention strategies.

Our findings, displayed in Fig 8, indicate that increasing the mean time to detect individuals with mild and severe symptoms (i.e., decreasing p_s and p_m) significantly impacts all

strategies by extending the period of undetected transmission. This leads to more secondary infections and a longer epidemic trajectory.

However, when the mean time to detection decreases (as p_s and p_m increase), the proposed CT risk-based approaches demonstrate resilience in comparison with Random Selection and the usual Contact Tracing. By varying p_s and p_m , our simulations reveal a strong relationship between detection timeliness and epidemic control. Faster detection (higher p_s and p_m) significantly enhances the effectiveness of all strategies. Despite increasing detection delays, the 2°CT method consistently demonstrates superior performance compared to alternative strategies. Reducing the mean time to detection through improved testing accessibility and coverage is crucial. Faster identification and isolation amplify the benefits of risk-based strategies, optimizing the use of limited resources.

These results emphasize the importance of early detection in controlling epidemic spread. Our 1°CT and 2°CT approach remains robust under several detection timelines, but its effectiveness is maximized when detection is prompt, aligning with the important role of efficient public health interventions.

Sensitivity and specificity

Here we relax the assumption of perfect diagnostic tests by expanding our simulations to realistic values for sensitivity and specificity, and we evaluate how this impacts our proposed risk-based prioritization strategy.

As mentioned in [36–38], RT-PCR tests, which are considered the gold standard for the diagnosis of COVID-19, have a clinical sensitivity around 90% and specificity approximately 95%. But the performance of COVID-19 tests is significantly influenced by the severity of symptoms. Indeed, individuals with severe symptoms often have higher viral loads, leading to greater sensitivity in both molecular and antigen tests. Conversely, those who are asymptomatic tend to have lower viral loads, which reduces test sensitivity. For symptomatic individuals, RT-PCR tests demonstrate sensitivity near 100% and specificity of approximately 95.5%, while antigen tests show sensitivity around 96.4% and specificity close to 98.7% [37]. It is important to note, as highlighted in [37], that the sensitivity for symptomatic individuals is 100% when tests are administered after symptoms onset.

To address this questions, we conducted simulations with more realistic test characteristics (Fig 9). In our simulations, we modeled different scenarios based on the type of detection and corresponding test characteristics:

- **Testing after symptoms onset:** for individuals tested due to symptoms, we considered perfect sensitivity.
- **Risk-based testing:** For individuals detected through risk-based methods (most of them asymptomatic at the time of testing), we used sensitivity of 90% and specificity of 95%.

Imperfect sensitivity introduces false negatives, enabling undetected infections to spread, while imperfect specificity generates false positives, potentially misallocating limited testing resources and leading to unnecessary quarantines. Reduced sensitivity slightly undermines the efficiency of identifying and isolating infected individuals, especially when daily testing capacity is constrained, see Fig 9.

Nonetheless, our simulations reveal that even under realistic test performance, the proposed method of 2° CT remains effective in mitigating epidemic spread. However, achieving this level of mitigation requires increased resources, such as a higher number of daily tests as compared to the same scenario with perfect tests. In contrast, 1° CT struggles to achieve similar mitigation outcomes under the same conditions.

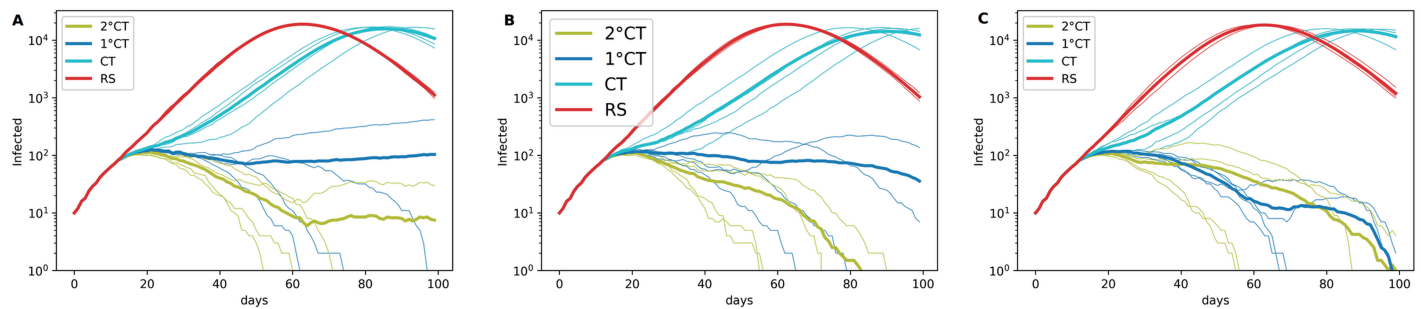


Fig 9. Effect of sensitivity and specificity on epidemic spread. Effect of sensitivity and specificity on epidemic spread for the strategies 1°CT, 2°CT, CT, and RS with (A) $\eta = 450$, (B) $\eta = 500$, and (C) $\eta = 600$. Each plot was generated from simulations using four seeds, with parameters $T = 100$, $N = 50K$, $t_0 = 12$, $N_0 = 10$, $p_s = 1$, and $p_m = 1/3$. The estimation of the infection time for the 1°CT and 2°CT strategies is given by $\hat{\tau}_j^i$. We fix the parameters $\gamma = 6$ and $\zeta = 9$.

<https://doi.org/10.1371/journal.pone.0320291.g009>

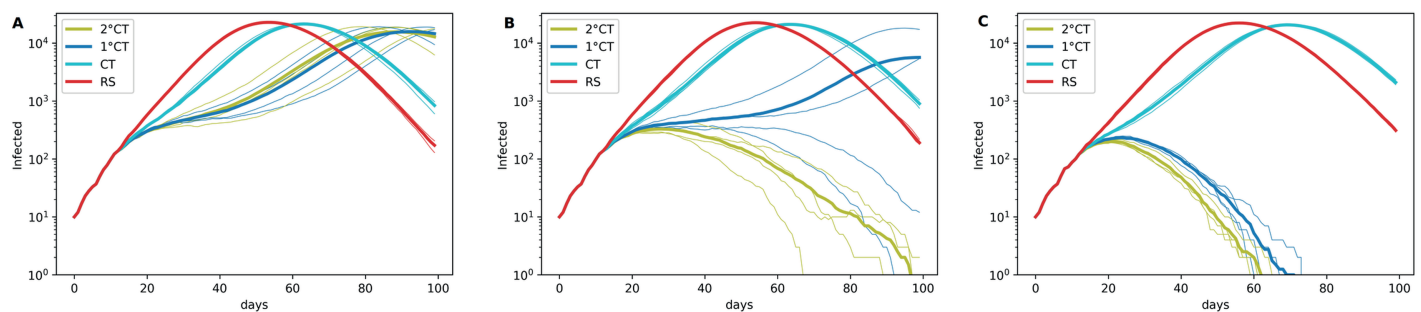


Fig 10. Effect of the quarantine adoption fraction q on epidemic spread. Effect of the parameter q (the fraction of individuals adopting quarantine), increasing from left to right with (A) $q = 0.9$, (B) $q = 0.95$, and (C) $q = 1$, on epidemic spread for the strategies 1°CT, 2°CT, CT, and RS. Each plot was generated from simulations using four seeds, with parameters $T = 100$, $N = 50K$, $t_0 = 12$, $N_0 = 10$, $\eta = 450$, $p_m = 1/5$, and $p_s = 1/3$. The estimation of the infection time for the 1°CT and 2°CT strategies is given by $\hat{\tau}_j^i$. We fix the parameters $\gamma = 6$ and $\zeta = 9$.

<https://doi.org/10.1371/journal.pone.0320291.g010>

As expected, realistic test performance slightly reduces the efficiency of our risk-based strategy, but does not compromise its relative advantage over alternative methods such as traditional contact tracing methods and random selection. In settings with limited testing capacity and imperfect diagnostics, the ability to prioritize based on risk might help compensating for the challenges posed by a reduced sensitivity and specificity.

Quarantine adoption fraction

To evaluate the effect of varying quarantine adoption rates, we introduce a parameter q corresponding to the probability for each individual to adhere to the quarantine recommendations when receiving a positive test result. By varying the values of q in our simulations, we explored how changes in the quarantine adoption rate impact the effectiveness of different intervention strategies, particularly our 2°CT risk approach.

Our findings, illustrated in Fig 10, show that reducing quarantine adoption rates significantly affects the performance of all strategies. However, the 2°CT risk-based approach demonstrates remarkable resilience when q remains greater than 0.95. Under these conditions, our method continues to outperform naive approaches by prioritizing high-risk individuals for testing and isolation, effectively mitigating epidemic spread despite partial quarantine adoption.

As q decreases below 0.95, the overall effectiveness of all strategies diminishes, highlighting the critical role of public adherence to quarantine measures. Lower adoption rates exacerbate the spread of infection, even when 1°CT and 2°CT ranking methods are employed.

These results underscore the necessity of both strategic prioritization and high levels of public cooperation to achieve meaningful epidemic control. The incorporation of variable quarantine adoption rates into our simulations provides a realistic assessment of the challenges and opportunities associated with public health interventions in real-world scenarios.

Role of super-spreaders. Super-spreaders, who infect a disproportionately high number of secondary cases, play a critical role in epidemic dynamics, see [39,40]. Below, we discuss how this phenomenon is considered in our model and provide evidence based on our simulations.

Our method indirectly accounts for super-spreaders by leveraging recent detections of infected individuals and promptly isolating them to decrease their number of contacts within the network structure. Specifically, the risk-based prioritization ranks individuals not only by direct connections to detected cases but also through indirect (up to two-step) transmission pathways. This approach naturally assigns higher priority to highly connected individuals, which are potential super-spreaders.

To evaluate the efficiency of the proposed 1°CT and 2°CT methods in identifying and mitigating the impact of super-spreaders, we compare them with other ranking strategies RS and CT, see Fig 11.

We conducted simulations comparing these methods to quantify their ability to address super-spreader dynamics.

- Frequency of secondary infections (Fig 11A):

We plotted the frequency of the number of infections caused by each infected individual (spreader) across four different seeds, using a logarithmic scale. Results demonstrate that our methods, particularly 2°CT, consistently identified and isolated super-spreaders earlier enough, leading to a significantly lower number of secondary cases as compared to the CT and RS approaches. In contrast, the CT and RS methods allowed super-spreaders to infect more individuals before being isolated, showcasing their relative inefficiency in mitigating the epidemic's spread.

- Effective reproduction number (Fig 11B):

We plotted the effective reproduction number ($R_{\text{effective}}$), that is the average number of individuals infected by active infectors at each time step t . This quantity plays a crucial role in understanding the influence of super-spreaders on epidemic dynamics. When super-spreaders are active, $R_{\text{effective}}$ tends to be higher due to their ability to amplify the spread of

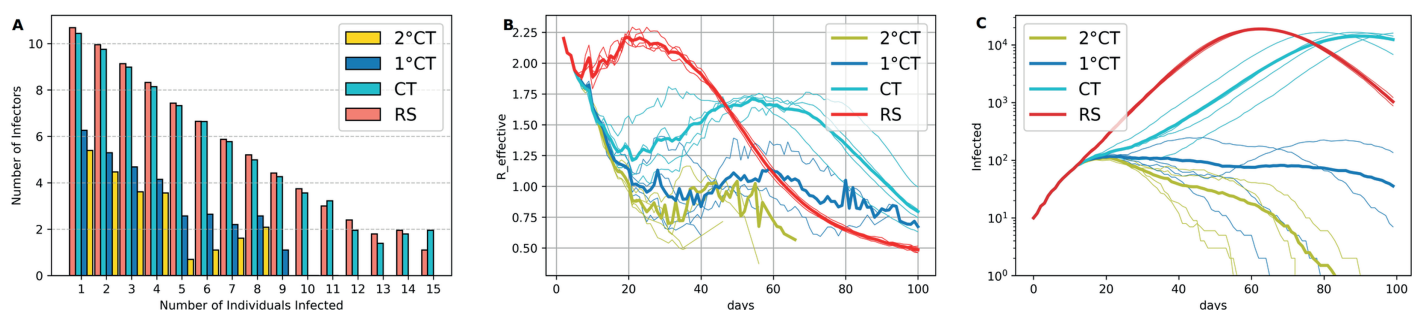


Fig 11. Effect of super-spreaders on ranking methods. Effect of super-spreaders on the ranking methods 1°CT, 2°CT, CT, and RS based on (A) the frequency of secondary infections caused by index cases, (B) the effective reproduction number over time ($R_{\text{effective}}$), and (C) the number of infectious individuals over time. Each plot was generated from simulations using four seeds, with parameters $T = 100$, $N = 50K$, $t_0 = 12$, $N_0 = 10$, $\eta = 500$, $p_m = 1/3$, $p_s = 1$, $q = 1$, sensitivity = 0.9, and specificity = 0.95. The estimation of the infection time for the 1°CT and 2°CT strategies is given by $\tilde{\tau}_i$. Parameters $\gamma = 6$ and $\zeta = 9$ are fixed across all simulations.

<https://doi.org/10.1371/journal.pone.0320291.g011>

infection through their extensive contact networks. If super-spreader events are not identified and mitigated, $R_{\text{effective}}$ may remain above the critical threshold of 1, allowing the epidemic to grow exponentially. Effective interventions, such as the proposed prioritization of high-risk individuals, can reduce the impact of super-spreaders by quickly isolating them and breaking chains of transmission.

- Number of infectious individuals over time (Fig 11C):

The number of infectious individuals is plotted over time, also on a logarithmic scale, for the same simulations corresponding to the two previous panels. As expected, better detection of individuals at increased risk of transmission allows more efficient mitigation of the epidemic.

Discussion

In this paper, we have introduced a method for the computation of the probability of infection of individuals in interaction, based on forward contact tracing, that considers at risk not only the direct contacts of detected individuals but also their subsequent contacts. We have called our method second-degree contact tracing (2°CT). The proposed method consists of estimating the individual infection risk by considering all possible chains of transmission up to second-degree contacts, coming from index cases. We propose a mitigation strategy that involves using the risk approximation to rank individuals and allocate the limited number of daily available tests accordingly. We have evaluated interventions based on our risk ranking through simulations of a fairly realistic agent-based model calibrated for COVID-19 epidemic outbreak (the Oxford OpenABM-Covid19 model). We have considered different scenarios to study the role of key quantities such as the number of daily available tests, the contact tracing time-window, the transmission probability per contact (constant versus depending on multiple factors), and the age since infection (for varying infectiousness). We found that, when there is a limited number of daily tests available, our method is capable of mitigating the propagation more efficiently than random selection, than the usual contact tracing (ranking according to the number of contacts with detected individuals), and than some other approaches in the recent literature on the subject. Additionally, our risk computation method can be easily adapted to the mitigation of other transmissible diseases spreading on contact networks.

One of the main difficulties in many forward contact tracing approaches for transmission diseases such as COVID-19, is to know the time of infection of the detected individuals. This quantity is in general not observed since in most cases individuals ignore from whom and when they got infected. However, inferring the time of infection is necessary for at least two reasons: (1) to know from which date the contacts of the detected individual should be traced, (2) to accurately assess the risk of infection in the case of varying transmissibility during the course of the disease. Given these arguments, we have considered age-dependent infectiousness, meaning that the probability of transmission from a source i depends on the time since infection of i . Moreover, we have proposed an efficient estimation of the time of infection for detected individuals, which is more accurate than considering a constant infection time, as it is often proposed in the literature. The results show how this estimation improves the contact tracing method in terms of the number of infectious individuals through time. Indeed, it allows to achieve in some cases almost as good results as considering the real date of infection.

Our 2°CT method encompasses the first degree contact tracing method, called here 1°CT method. We have found that with a limited number of available tests, the 2°CT method is more effective in the mitigation of an epidemic than the 1°CT method. However, with a

large enough number of available daily tests, both 1°CT and 2°CT methods provide similar results; in this case, we recommend the use of the 1°CT method because of a simpler and faster computation and better preserving individual privacy. Besides our results show that the proposed 1°CT and 2°CT methods can be very effective compared with the usual contact tracing, the mean-field risk approximation or the random selection of individuals to test.

By integrating test performance into our simulations and analyses, we strengthen the practical relevance of our method, which is designed to work with real-world constraints such as limited resources, imperfect diagnostics, and partial data. This makes it a robust decision-making tool for epidemic mitigation, even under non-ideal conditions. Future work could explore efficient approximations or scalable computational techniques to integrate test characteristics without incurring prohibitive complexity. Such advancements would enhance the precision of the risk-based method while preserving its practicality for real-world applications.

Despite its advantages, our intervention method has some limitations. Firstly, our risk estimation formula assumes perfect tests. Future work could explore the explicit calculation of the risk, or its efficient approximation, to integrate realistic test sensitivity and specificity. Increasing the population size beyond 50k, up to 100k as in [6], or to 500k as in [2], would be another direction for future work. Another assumption that we intend to include in a forthcoming version of our model is the uncertainty in the list of contacts of the traced individuals, and it would be of interest to study how this impacts the efficacy of the intervention. Likewise, the proposed model is primarily designed for the early stages of an epidemic, when the testing capacity is limited and vaccines are unavailable. Incorporating vaccination status in future extensions could enhance the model's applicability across various stages of an epidemic, broadening its utility in real-world scenarios. Finally, since several models exist, integrating our risk assessment with other risk approaches could enhance the predictive accuracy of the results, reduce uncertainty in prioritization decisions, and enable more targeted interventions tailored to specific regions, epidemic dynamics, or populations.

Supporting information

S1 Appendix. OpenABM-Covid19 model.
(PDF)

S2 Appendix. Comparison with mean-field.
(PDF)

Author contributions

Conceptualization: Miraine Dávila Felipe, Ghislaine Gayraud.

Formal analysis: Gabriela Bayolo Soler, Miraine Dávila Felipe, Ghislaine Gayraud.

Funding acquisition: Miraine Dávila Felipe, Ghislaine Gayraud.

Investigation: Gabriela Bayolo Soler, Miraine Dávila Felipe, Ghislaine Gayraud.

Methodology: Gabriela Bayolo Soler, Miraine Dávila Felipe, Ghislaine Gayraud.

Project administration: Gabriela Bayolo Soler.

Software: Gabriela Bayolo Soler, Miraine Dávila Felipe, Ghislaine Gayraud.

Supervision: Miraine Dávila Felipe, Ghislaine Gayraud.

Validation: Gabriela Bayolo Soler, Miraine Dávila Felipe, Ghislaine Gayraud.

Visualization: Gabriela Bayolo Soler.

Writing – original draft: Gabriela Bayolo Soler, Miraine Dávila Felipe, Ghislaine Gayraud.

Writing – review & editing: Gabriela Bayolo Soler, Miraine Dávila Felipe, Ghislaine Gayraud.

References

1. Brandt AM. The history of contact tracing and the future of public health. *Am J Public Health*. 2022;112(8):1097–9. <https://doi.org/10.2105/ajph.2022.306949>
2. Baker A, Biazio I, Braunstein A, Catania G, Dall'Asta L, Ingrosso A, et al. Epidemic mitigation by statistical inference from contact tracing data. *Proc Natl Acad Sci USA*. 2021;118(32):e2106548118. <https://doi.org/10.1073/pnas.2106548118>.
3. Herbrich R, Rastogi R, Vollgraf R. CRISP: a probabilistic model for individual-level COVID-19 infection risk estimation based on contact data. *arXiv*. 2022. <https://arxiv.org/abs/2006.04942>
4. Battle P, Bruna J, Fernandez-Granda C, Preciado VM. Adaptive test allocation for outbreak detection and tracking in social contact networks. *SIAM J Control Optim*. 2022;60(2):S274–93. <https://doi.org/10.1137/20m1377874>
5. Romijnders R, Asano YM, Louizos C, Welling M. No time to waste: practical statistical contact tracing with few low-bit messages. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, PMLR; 2023. pp. 7943–60. Available: <https://proceedings.mlr.press/v206/romijnders23a.html>.
6. Guttal V, Krishna S, Siddharthan R. Risk assessment via layered mobile contact tracing for epidemiological intervention. *medRxiv*; 2020. <https://www.medrxiv.org/content/10.1101/2020.04.26.20080648v1.full>
7. Bestvina I, Thornton W. Contact tracing infection risk estimate. Distributed simulation approach; 2021. Available: <https://www.viratrace.org/#/>.
8. Murphy K, Kumar A, Serghiou S. Risk score learning for COVID-19 contact tracing apps. In *Machine Learning for Healthcare Conference*. PMLR; 2021. pp. 373–90. Available: <https://proceedings.mlr.press/v149/murphy21a.html>.
9. Gupta P, Maharaj T, Weiss M, Rahaman N, Alsdurf H, Minoyan N, et al. Proactive contact tracing. *PLOS Digital Health*. 2023;2(3):e0000199. <https://doi.org/10.1371/journal.pdig.0000199>
10. Bengio Y, Gupta P, Maharaj T, Rahaman N, Weiss M, Deleu T, et al. Predicting infectiousness for proactive contact tracing. In *International Conference on Learning Representations (ICLR)*. Virtual Conference; 2021.
11. Sattler F, Ma J, Wagner P, Neumann D, Wenzel M, Schäfer R, et al. Risk estimation of SARS-CoV-2 transmission from bluetooth low energy measurements. *NPJ Digit Med*. 2020;3(1):129. <https://doi.org/10.1038/s41746-020-00340-0>
12. Alsdurf H, Belliveau E, Bengio Y, Deleu T, Gupta P, Ippolito D, et al. COVI White Paper. *arXiv*; 2020. <http://arxiv.org/abs/2005.08502>
13. Firth JA, Hellewell J, Klepac P, Kissler S, Kucharski AJ, Spurgin LG. Using a real-world network to model localized COVID-19 control strategies. *Nat Med*. 2020;26(10):1616–22. <https://doi.org/10.1038/s41591-020-1036-8>
14. Weigl JAI, Feddersen AK, Stern M. Household quarantine of second degree contacts is an effective non-pharmaceutical intervention to prevent tertiary cases in the current SARS-CoV pandemic. *BMC Infect Dis*. 2021;21(1):1262. <https://doi.org/10.1186/s12879-021-06818-w>
15. Eames KT, Keeling MJ. Contact tracing and disease control. *Proc Biol Sci*. 2003;270(1533):2565–71. <https://doi.org/10.1098/rspb.2003.2554>
16. Cebrian M. The past, present and future of digital contact tracing. *Nat Electron*. 2021;4(1):2–4. <https://doi.org/10.1038/s41928-020-00535-z>.
17. Leung KY, Metting E, Ebberts W, Veldhuijzen I, Andeweg SP, Luijben G, et al. Effectiveness of a COVID-19 contact tracing app in a simulation model with indirect and informal contact tracing. *Epidemics*. 2024;46:100735. <https://doi.org/10.1016/j.epidem.2023.100735>
18. Zwitter A, Gstrein OJ. Big data, privacy and COVID-19—learning from humanitarian expertise in data protection. *J Int Humanit Action*. 2020;5(4):1–7. <https://doi.org/10.1186/s41018-020-00072-6>
19. Jenniskens K, Bootsma MCJ, Damen JAAG, Oerbecke MS, Vernooij RWM, Spijker R, et al. Effectiveness of contact tracing apps for SARS-CoV-2: a rapid systematic review. *BMJ Open*. 2021;11(7):e0050519. <https://doi.org/10.1136/bmjopen-2021-0050519>

20. Pozo-Martin F, Sanchez MAB, Müller SA, Diaconu V, Weil K, Bcheraoui CE. Comparative effectiveness of contact tracing interventions in the context of the COVID-19 pandemic: a systematic review. *Eur J Epidemiol*. 2023;38(3):243–66. <https://doi.org/10.1007/s10654-023-00963-z>
21. Duan W, Fan Z, Zhang P, Guo G, Qiu X. Mathematical and computational approaches to epidemic modeling: a comprehensive review. *Front Comput Sci*. 2015;9:806–26. <https://doi.org/10.1007/s11704-014-3369-2>
22. Colizza V, Barthélemy M, Barrat A, Vespignani A. Epidemic modeling in complex realities. *C R Biol*. 2007;330(4):364–74. <https://doi.org/10.1016/j.crv.2007.02.014>
23. Britton T, Pardoux E, Ball F, Laredo C, Sirl D, Tran VC. *Stochastic Epidemic Models with Inference*, Vol. 2255. Berlin: Springer; 2019.
24. Simoy MI, Aparicio JP. Socially structured model for COVID-19 pandemic: design and evaluation of control measures. *Comput Appl Math*. 2022;41(1):14. <https://doi.org/10.1007/s40314-021-01705-1>
25. Xue J, Zhang M, Xu M. Modeling the impact of social distancing on the COVID-19 pandemic in a low transmission setting. *IEEE Trans Comput Soc Syst* 9(4):1122–31. <https://doi.org/10.1101/2020.05.06.20092841>
26. Wu F, Liang X, Lei J. Modelling COVID-19 epidemic with confirmed cases-driven contact tracing quarantine. *Infect Dis Model*. 2023;8(2):415–26. <http://dx.doi.org/10.1016/j.idm.2023.04.001>
27. Hinch R, Probert WJ, Nurtay A, Kendall M, Wymant C, Hall M, et al. OpenABM-Covid19—an agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. *PLoS Comput Biol*. 2021;17(7):e1009146. <https://doi.org/10.1371/journal.pcbi.1009146>
28. Biazio I, Braunstein A, Dall'Asta L, Mazza F. A Bayesian generative neural network framework for epidemic inference problems. *Sci Rep*. 2022;12(1):19673. <https://doi.org/10.1038/s41598-022-20898-x>
29. Shah C, Dehmamy N, Perra N, Chinazzi M, Barabási AL, Vespignani A, et al. Finding patient zero: learning contagion source with graph neural networks. *arXiv*; 2020. <https://arxiv.org/abs/2006.11913>
30. Čutura G, Li B, Swami A, Segarra S. Deep demixing: reconstructing the evolution of epidemics using graph neural networks. In 29th European Signal Processing Conference (EUSIPCO); 2021. pp. 2204–8.
31. Tomy A, Razzanelli M, Di Lauro F, Rus D, Della Santina C. Estimating the state of epidemics spreading with graph neural networks. *Nonlinear Dyn*. 2022;109(1):249–63. <https://doi.org/10.1007/s11071-021-07160-1>
32. Tan CW, Yu PD, Chen S, Poor HV. DeepTrace: learning to optimize contact tracing in epidemic networks with graph neural networks. *arXiv*; 2023. <https://arxiv.org/abs/2211.00880>
33. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*. 2020;368(6491):eabb6936. <https://doi.org/10.1126/science.abb6936>
34. Britton T. Epidemic models on social networks—With inference. *Stat Neerl*. 2020;74(3):222–41.
35. Dos Santos PG, Vieira HCVS, Wietholter V, Gallina JP, Andrade TR, Marinowicz DR, et al. When to test for COVID-19 using real-time reverse transcriptase polymerase chain reaction: a systematic review. *Int J Infect Dis*. 2022;123:58–69. <https://doi.org/10.1016/j.ijid.2022.06.037>
36. Miller TE, Garcia Beltran WF, Bard AZ, Gogakos T, Anahtar MN, Astudillo MG, et al. Clinical sensitivity and interpretation of PCR and serological COVID-19 diagnostics for patients presenting to the hospital. *FASEB J*. 2020;34(10):13877–84. <https://doi.org/10.1096/fj.202001700r>
37. Pekosz A, Parvu V, Li M, Andrews JC, Manabe YC, Kodosi S et al. Antigen-based testing but not real-time polymerase chain reaction correlates with severe acute respiratory syndrome coronavirus 2 viral culture, *Clin Infect Dis*. 2021;73(9):e2861–6. <https://doi.org/10.1093/cid/ciaa1706>
38. Prince-Guerra, J. L. Evaluation of Abbott BinaxNOW rapid antigen test for SARS-CoV-2 infection at two community-based testing sites—Pima County, Arizona, November 3–17, 2020. *MMWR Morb Mortal Wkly Rep*. 2021;70(3):100–5. <https://doi.org/10.15585/mmwr.mm7003e3>
39. Brainard J, Jones NR, Harrison FC, Hammer CC, Lake IR. Super-spreaders of novel coronaviruses that cause SARS, MERS and COVID-19: a systematic review. *Ann Epidemiol*. 2023;82:66–76. <https://doi.org/10.1016/j.annepidem.2023.03.009>
40. Illingworth CJ, Hamilton WL, Warne B, Routledge M, Popay A, Jackson C et al. Superspreaders drive the largest outbreaks of hospital onset COVID-19 infections. *elife*. 2021;10:e67308. <https://doi.org/10.7554/elife.67308>