

RESEARCH ARTICLE

A stochastic contact network model for assessing outbreak risk of COVID-19 in workplaces

Xi Guo ^{*}, Abhineet Gupta ^{*}, Anand Sampat , Chengwei Zhai 

One Concern, Inc., Menlo Park, CA, United States of America

 These authors contributed equally to this work.^{*} cguo@oneconcern.com (XG); abhineet@oneconcern.com (AG)

Abstract

The COVID-19 pandemic has drastically shifted the way people work. While many businesses can operate remotely, a large number of jobs can only be performed on-site. Moreover as businesses create plans for bringing workers back on-site, they are in need of tools to assess the risk of COVID-19 for their employees in the workplaces. This study aims to fill the gap in risk modeling of COVID-19 outbreaks in facilities like offices and warehouses. We propose a simulation-based stochastic contact network model to assess the cumulative incidence in workplaces. First-generation cases are introduced as a Bernoulli random variable using the local daily new case rate as the success rate. Contact networks are established through randomly sampled daily contacts for each of the first-generation cases and successful transmissions are established based on a randomized secondary attack rate (SAR). Modification factors are provided for SAR based on changes in airflow, speaking volume, and speaking activity within a facility. Control measures such as mask wearing are incorporated through modifications in SAR. We validated the model by comparing the distribution of cumulative incidence in model simulations against real-world outbreaks in workplaces and nursing homes. The comparisons support the model's validity for estimating cumulative incidences for short forecasting periods of up to 15 days. We believe that the current study presents an effective tool for providing short-term forecasts of COVID-19 cases for workplaces and for quantifying the effectiveness of various control measures. The open source model code is made available at github.com/abhineetgupta/covid-workplace-risk.

OPEN ACCESS

Citation: Guo X, Gupta A, Sampat A, Zhai C (2022) A stochastic contact network model for assessing outbreak risk of COVID-19 in workplaces. PLoS ONE 17(1): e0262316. <https://doi.org/10.1371/journal.pone.0262316>

Editor: Amitava Mukherjee, VIT University, INDIA

Received: August 13, 2021

Accepted: December 20, 2021

Published: January 14, 2022

Copyright: © 2022 Guo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The model code is available on Github at <https://github.com/abhineetgupta/covid-workplace-risk>. Weekly COVID-19 data from nursing homes is from <https://data.cms.gov/Special-Programs-Initiatives-COVID-19-Nursing-Home/COVID-19-Nursing-Home-Dataset/s2uc-8wpx>. County-level COVID-19 case rate data in the US is from <https://github.com/nytimes/covid-19-data>.

Funding: One Concern, Inc. provided support in the form of salaries for authors [XG, AG, AS, CZ], but did not have any additional role in the study design, data collection and analysis, decision to

Introduction

The current pandemic of COVID-19 has now led to more than 180 million cases and more than 3.9 million deaths worldwide as of 2021–06–30 [1]. The disease is caused by SARS-CoV-2—a type of respiratory virus. Compared to other coronaviruses like severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus (MERS-CoV), it is less deadly but has higher transmissibility [2]. Due to its airborne nature, long incubation, and heightened transmissibility, the pandemic has endangered

publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: This does not alter our adherence to PLOS ONE policies on sharing data and materials.

public health, especially those of workers in various types of settings. For example, there has been a considerable number of outbreaks in meat processing facilities [3–5], offices, warehouses, and university campuses [6–9].

Amidst this pandemic, government and businesses have taken significant measures to mitigate the outbreak risk in indoor workplaces including capacity reduction, remote working, mask wearing policy, social distancing, etc. The percentage of employers who reported a successful shift to remote work have increased from 73% to 83%, from June 2020 to January 2021 [10]. However, more than 60% of workers in the United States economy cannot work remotely and this percentage is even higher in emerging economies [11]. For those able to work remotely, a survey of executives suggests that COVID-19 may propel faster adoption of automation and artificial intelligence [12], work from home may increase from 20% pre-pandemic to 27% post-pandemic [13], and hybrid work arrangements with employees splitting time at home and in the office will be the new normal [10, 14].

While the workforce is adapting to these significant changes, the need for ensuring the safety of workers in the long-term is becoming more urgent. Business owners are concerned about reoccupying their workplaces, but their expectations are far from reality. In June 2020, 80% of executives expected a return back to normal by September 2020, and 88% by December 2020 but most of these businesses continue being fully remote even in April 2021 [15]. Meanwhile, executives have been hard-pressed to make the right mitigation decisions for safe return of their employees. In a survey of 100 executives from June 2020, 78% had implemented or planned to implement once a day temperature checks while only 26% favored contact tracing and 11% favored regular testing of employees, showing a preference for more convenient measures like temperature checking which is less effective due to the prevalence of pre-symptomatic and asymptomatic cases [15]. At the same time, information about COVID-19 transmission has continued to evolve making it difficult for businesses to choose the right mitigation measures to maximize safety and minimize disruptions. Businesses need to make near real-time decisions in response to local trends of disease circulation while being resilient against long-term changes in the pandemic. In order to make effective decisions for reoccupation of workplaces, facility managers, human resources personnel, and business decision makers need data-driven tools to help them assess outbreak risks within their facilities based on local community conditions, and the relative benefits of different mitigation measures.

There are several differences between modeling disease dynamics in a large population and modeling transmissions in a workplace with a small number of people. Parameters such as probability of transmission (β , R_0) are relatively easy to estimate in a large population compared to smaller groups of employees in workplaces. In indoor spaces with restricted access such as offices and warehouses, population is generally static, and properties of the environment and individual behavior have large impacts on the transmission [16, 17]. Furthermore, an infectious pathogen can often invade and be sustained in a large population when $R_0 > 1$, but in small and closed populations, the disease transmission can die off quickly due to either the depletion of susceptible population or removal of infectious population [18]. Due to these differences, the traditional compartmental models often do not fit the purpose of modeling the risk of outbreaks in small and closed population.

Several recent models have aimed to simulate the physics behind COVID-19 transmission within offices, classrooms, and restaurants [19, 20]. These studies estimate the transmission risk at the point of the transmission event, but do not provide estimation on the scale of outbreaks. Other studies have modeled human behavior in indoor spaces using agent-based or network-based models [21–23]. For example, a contact network model was developed to estimate the cumulative incidence of COVID-19 on the Princess Diamond Cruise ship [21]. These models utilize the unique contact network structure in their specific cases, like the interactions

between passengers, interaction between crew members, and interactions between passengers and crew members on a cruise ship. However, they are difficult to generalize for other locations and settings.

Recently, several models and web-based tools have also been developed to model outbreaks in large gatherings or indoor places [24–26]. The majority of these models are deterministic. A web-based tool called the COVID-19 Indoor Safety Guideline aims to develop safety guidelines based on room specifications, human behavior, age group, and virus variant [25]. The study demonstrated the impact of airflow, human behavior, and indoor environment on SARS-CoV-2 transmission, but it did not consider the impact of local incidence rate on the introduction of the outbreaks. Another application is developed to estimate the risk that at least one individual with SARS-CoV-2 is present in gatherings of different sizes in the United States [27]. Their assumption of community introduction following Bernoulli trials matches that of the present study. However, the tool does not provide the ability to assess outbreak risk in workplaces. An agent based stochastic model called Covasim has been created to project epidemic trends, explore intervention scenarios, and estimate resource needs [22]. The model is composed of a collection of open-source data-driven synthetic contact networks, like Synth-Pops and hybrid networks. However, the study did not include information about its applicability for assessing risks in workplaces with small population.

This work contributes to the body of knowledge in modeling COVID-19 outbreaks in indoor places. It incorporates secondary attack rate (SAR) as the transmission parameter into a stochastic contact network. By consistently accounting for uncertainties in the transmission chain, this model enables probabilistic short-term forecasting of the outbreak risk within workplaces with static populations. It also allows for customization based on local case rates, control measures such as mask wearing, behaviors of employees such as interactions and speaking activity, and environmental factors such as airflow and filtration. Additionally, we provide guidance for parameter values appropriate for general workplaces so that the model can be implemented widely.

The objectives of this study include—presenting a contact-network model that can be used to forecast cumulative incidences of COVID-19 in small populations; selecting probabilistic parameters associated with transmission dynamics and control measures to evaluate the impact of mitigation; and validating the model with real-world examples. The design and methodology of the stochastic contact-network model is presented in the Materials and Methods section. The validation against observations from 7 individual outbreak reports in workplaces worldwide, and against weekly reported data from over 8000 nursing homes in the United States is presented in the Results section. The observations closely match model estimates for both sets of comparisons. In the Discussion section, we highlight the primary limitations of the work, and motivations for future work. Finally, we present a summary of our conclusions.

Materials and methods

We present a stochastic contact network which emulates the introduction of COVID-19 cases in a facility and the daily contacts among people within the facilities as the routes of transmission. The nodes in the network represent each individual, and the edges represent their interactions or contacts. The occurrence of edges is stochastic and homogeneous. We only consider the infectious, susceptible, and exposed populations in the model. The model outputs the cumulative incidence within a facility after n days, where $n \leq 15$ is intended to be a short forecasting period.

Network structure

Our model adopted a star network topology [28, 29] to represent the contacts between the first generation and the susceptible employees. The central hub of the network represents the first-generation cases who arrive at the workplace. Depending on the local case rates, there could be one or several hubs, hence one or several independent star networks. Moreover, new case introductions can occur during the forecast period, hence new hubs may be added on each day. The edges connecting the central hub and other nodes indicate contact made between the first-generation cases and the susceptible employees on site. Each hub draw connections with nodes randomly based on a geometric distribution. The edges can be created for each day, remain for the next day, or only exist for one day depending on the probability of contacts being maintained for the next day. Successful transmission of the pathogen through edges (contacts) is governed probabilistically by the secondary attack rate. Networks for each first-generation case are mutually independent.

Model design

To model the transmission dynamic within a short-term period, we defined the components of a transmission chain as below.

- **Case introduction**—Introduction of COVID-19 cases in a facility is defined as the time when the first case arrived in the facility.
- **Transmission chain**—The process in which the virus spreads from one person to another. In our contact network, the transmission chain is represented by series of nodes (infectious, susceptible, and exposed individuals) and edges (contacts).
- **First-generation cases**—Individuals who are the starting point of a transmission chain. These include one or more index cases that are infected outside a facility and introduce the disease into the facility. This group is considered as the infected population.
- **Second-generation cases**—Individuals who are exposed to COVID-19 as a result of a successful transmission event after coming into contact with the first-generation cases. This group forms the exposed population. In the present model, we assume that these individuals do not become infectious within the short forecasting period. Hence, the transmission chain ends at second-generation cases.

The pseudocode in Algorithm 1 illustrates the design of the model. The first-generation cases who are infected from the community can arrive at the facility on any given day. We assume that these cases are introduced to the facility as Bernoulli trials using local daily new case rate as the success rate. Once first-generation cases arrive, we initiate random contacts between the infectious and the susceptible individuals. The daily contact size of the first-generation cases is randomly determined from a geometric distribution based on the average daily contacts. The employees contacted by the first-generation case(s) are randomly chosen from the population. Additionally, the model has a parameter to probabilistically determine how many contacts remain as contacts for the next day.

The transition from susceptible to exposed state takes place among the employees who make contacts with the first-generation cases based on the secondary attack rate (SAR). A randomized SAR drawn from a Beta distribution is assigned to each first-generation case. The average SAR for a workplace is determined based on mask compliance, airflow, filtration, speaking volume, and speaking percentage.

The required number of simulations to achieve convergence for the model depends on the parameter values. For a wide range of parameters, we found that the average cumulative incidence converges at $n_{\text{sim}} \geq 1000$.

Algorithm 1 Model Pseudocode. Stochastic contact network model for short-term forecasting within facilities

Require:

```

 $p_{\text{case}} \leftarrow$  daily new case rate
 $e \leftarrow$  total employees or nodes
 $n_{\text{days}} \leftarrow$  total number of modeling days
 $\alpha_{\text{SAR}} \leftarrow$  alpha parameter for SAR Beta distribution
 $\beta_{\text{SAR}} \leftarrow$  beta parameter for SAR Beta distribution
 $\bar{c} \leftarrow$  average daily contact size
 $p_{\text{remain}} \leftarrow$  probability that edges (contacts) are sustained for the next day
 $n_{\text{sim}} \leftarrow$  number of simulations
for all  $\text{sim} \in n_{\text{sim}}$  do
  Sample number of first-generation cases for  $n_{\text{days}}$ ,
   $\mathbf{g}_1 := (g_1^{(1)}, \dots, g_1^{(n_{\text{days}})}) \sim \text{Binomial}(e, p_{\text{case}})$ 
  Initialize empty array of node indices for first-generation cases,
   $\mathbf{g}_1 := ()$ 
  for all  $g_1 \in \sum \mathbf{g}_1$  do
    Sample SAR for each first-generation case,  $\text{SAR}_{g_1} \sim \text{Beta}(\alpha_{\text{SAR}}, \beta_{\text{SAR}})$ 
    Initialize empty array of previous day's contacts for each first-generation case prior to day 1,  $\mathbf{c}_{\text{remain}}^{(0)} := ()$ 
  end for
  for all  $d \in n_{\text{days}}$  do
    Sample unique indices of first-generation cases for day  $d$ ,  $\mathbf{g}_1 = \mathbf{g}_1 + \text{Uniform}(e, g_1^{(d)})$ 
    for all  $g_1 \in \mathbf{g}_1$  do
      Sample number of contacts for each first-generation case for day while ensuring they do not exceed total employees
       $n_{g_1}^{(d)} \sim \min\left(\text{Geometric}\left(\frac{1}{\bar{c}+1}\right), e\right)$ 
      Sample indices of new contacts,  $\mathbf{c}_{\text{new}}^{(d)} \sim \text{Uniform}(e, n_{g_1}^{(d)} - \text{size}(\mathbf{c}_{\text{remain}}^{(d-1)}))$ 
      Total secondary contacts for day,  $\mathbf{c}_{g_1}^{(d)} := \mathbf{c}_{\text{new}}^{(d)} + \mathbf{c}_{\text{remain}}^{(d-1)}$ 
      Sample indices of second-generation cases from contacts for day  $d$ ,  $\mathbf{g}_2^{(d)} \sim \text{Binomial}(\mathbf{c}_{g_1}^{(d)}, \text{SAR}_{g_1})$ 
      Sample contacts that remain consistent the next day,
       $\mathbf{c}_{\text{remain}}^{(d)} \sim \text{Binomial}(\mathbf{c}_{g_1}^{(d)}, p_{\text{remain}})$ 
    end for
  end for
  Total cumulative incidence,  $\text{cases}_{\text{sim}} := \text{Unique}(\mathbf{g}_1 + \sum_d \mathbf{g}_2^{(d)})$ 
end for

```

Model parameters and implementation

This section describes the model implementation, and the distribution of model parameters.

Case introduction. We assume that first-generation cases are introduced into a facility as Bernoulli trials using local daily new case rate as the success rate. Therefore, the probability of x infectious employees arriving on any given day at a facility with e employees, and the local daily new case rate p_{case} , $P(x)$ is calculated from a Binomial distribution as $P(x) = \text{Binomial}(x, e, p_{\text{case}})$. Because of the short modeling period, and possibility of test lagging and underreported cases, we assumed the case rate to be constant throughout the modeling period,

although a variable case rate can also be implemented. The assumption of Binomial distribution matches similar approaches used in other studies for modeling probability of case occurrence [24, 27, 30–32].

Secondary attack rate (SAR). Secondary attack rate (SAR) is defined as the probability of an infection being transferred to a naive individual [33]. It is approximated as the proportion of secondary cases induced in a contact group by the first case in the group [34–36]. In the present model, it is the probability of getting exposed among all secondary contacts of a first-generation case.

Multiple studies on COVID-19 outbreaks have reported a wide range of SAR across different settings [37–43]. High values of SAR have been observed in specific settings, such as— SAR = 53.3% during loud singing [39], SAR = 84.6% after a business meeting [44], and SAR = 38.8% when having meals together [45]. A meta-analysis of SAR for COVID-19 concluded that workplaces where close contacts are less intense and frequent tend to have lower values of SAR (0%–5.3%) than household settings (3.9%–54.9%) [43]. These values of SAR reflect transmission in naive populations, and we have not considered the reduction in transmission due to COVID-19 vaccination in this study.

The SAR associated with each infectious individual varies because of their viral load and course of infection [43, 46]. We incorporate this variability by sampling SAR for each first-generation case from the Beta distribution. The SAR assigned to each first-generation case remains constant throughout the forecast period as a simplifying assumption. The Beta distribution is selected because it is defined on the desired interval [0, 1] and is able to model long tails. The distribution has two parameters: α_{SAR} and β_{SAR} . We expect the SAR distribution in a population to be long-tailed, where a majority of infectious individuals exhibit low SAR or infect few people, while a minority of individuals exhibit high SAR [43, 47]. A long-tail distribution can be implemented in the Beta distribution by setting $\alpha_{SAR} = 1$. Then, the β_{SAR} is calculated from the average \overline{SAR} across multiple individuals, as shown below.

$$\beta_{SAR} = \alpha_{SAR} \frac{1 - \overline{SAR}}{\overline{SAR}} \tag{1}$$

Multiple studies have concluded that the secondary attack rates associated with the transmission of airborne pathogens vary based on particle emission and inhalation rates, airflow, and filtration [19, 25, 48]. Here we present an approach to modify SAR in a facility based on these parameters.

Based on one of the studies [19], and the assumptions described in S3 Appendix, the average SAR for airborne transmission of COVID-19 depends on inhaled virions as -

$$\overline{SAR} = 1 - \exp(-N')$$

and

$$\begin{aligned} N' &\propto \frac{S(1 - m_e)}{\lambda} \\ \Rightarrow \frac{N'_i}{N'_j} &= \frac{S_i(1 - m_{e_i})}{\lambda_i} \frac{\lambda_j}{S_j(1 - m_{e_j})} \end{aligned} \tag{2}$$

where subscripts i and j represent any two sets of parameters. N' is the total number of inhaled virions normalized by the average infectivity threshold. $m_e \in [0, 1]$ is the mask effectiveness. S is the virion emission at the source and is a function of speaking percentage and volume. λ is the decay rate of virions and is a function of airflow and filtration.

Source emission rate. We assume that the source emission rate S is a function of speaking percentage s_p , and speaking volume s_v , in decibels. It was observed from case reports that talking releases 46 times more virions than breathing [19]. Other studies have concluded that higher volume releases higher number of aerosolized particles carrying virions [25, 48]. We represent the increase in virions based on volume by the volume multiplier s_m . Then, source emission S is described as -

$$S = (1 - s_p) + 46s_p s_m \tag{3}$$

Based on measurements from 10 participants, increasing the speaking volume from 70 dB to 98 dB increased the particle emission rate from 6 to 53 particles per second [48]. Additionally, the particle emission rate is directly proportional to the root mean square amplitude of the vocalization. Since the decibel level depends on the logarithm of amplitude, the relationship with emitted particles or virions q can be described as -

$$\begin{aligned} s_v &\propto \log(q) \\ \Rightarrow s_v &= C_q \log(q) \end{aligned} \tag{4}$$

where C_q is an unknown constant, and using the above values, is computed as -

$$C_q = \frac{98 - 70}{\log(53/6)} \tag{5}$$

Assuming that the ratio of 46 virions corresponds to speaking at a reference volume of 60 dB, the activity multiplier s_m at an arbitrary volume level can then be computed as -

$$s_m = \frac{53^{(s_v - 60)/(98 - 70)}}{6} \tag{6}$$

Decay rate. The decay rate λ is a function of air exchange rate or airflow λ_a (air changes per hour or ACH), viral settling and deactivation $\lambda_d = 0.62/h$, and decay due to filtration λ_f [19].

$$\lambda = \lambda_a + \lambda_d + \lambda_f \tag{7}$$

Mask effectiveness. The mask effectiveness parameter m_e combines the contribution of masks from both source emission rate and breathing rate, and effectively reduces the number of inhaled virions as shown in Eq 2. Based on a meta-analysis on the impact of mask-wearing on COVID-19 transmission, the relative risk (RR) of mask wearing under non-healthcare setting is RR = 56%, and for healthcare setting is RR = 30% [49]. Then mask effectiveness is implemented in Eq 2 as the factor $m_e = 1 - RR$. More research is needed to better characterize mask effectiveness, and studies have shown high variability of mask effectiveness for cloth masks [50].

Reference SAR. In order to use Eq 2 to estimate \overline{SAR} for any set of speaking percentage, volume, airflow, filtration, and mask effectiveness, we need a reference \overline{SAR}_{ref} and its corresponding parameters as shown in Eq 8.

$$\begin{aligned} \overline{SAR} &= 1 - \exp(-N') \\ N'_{ref} &= -\log(1 - \overline{SAR}_{ref}) \\ \frac{N'}{N'_{ref}} &= \frac{S(1 - m_e)}{\lambda} \frac{\lambda_{ref}}{S_{ref}(1 - m_{e,ref})} \end{aligned} \tag{8}$$

Impact of Airflow, Speaking Percentage and Speaking Volume on SAR

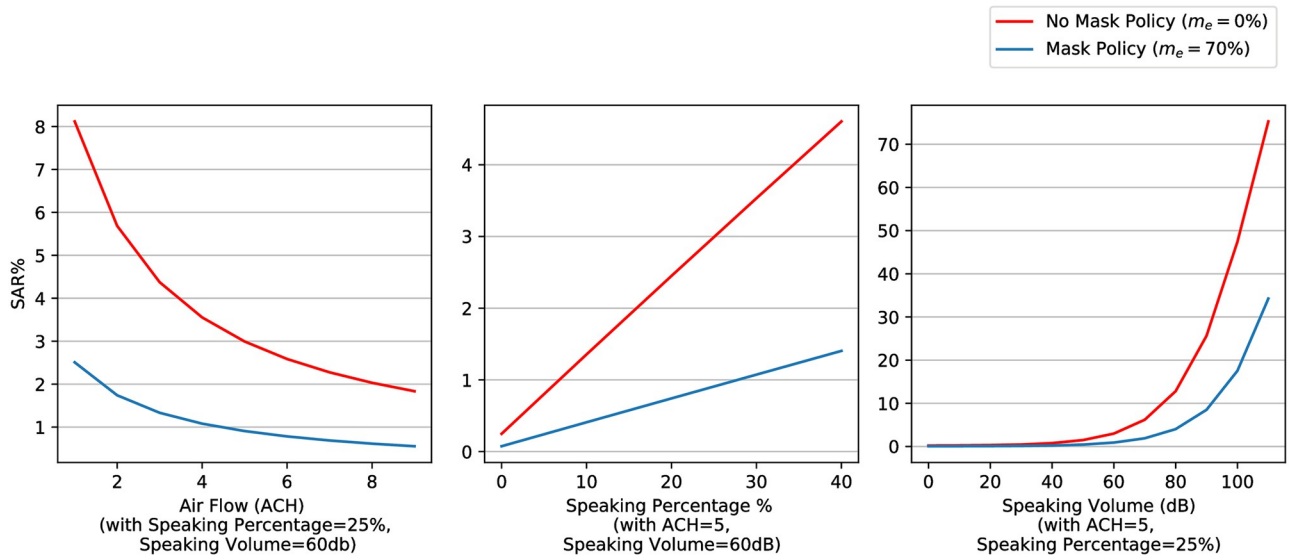


Fig 1. SAR variation. Impact of airflow, speaking volume, speaking percentage, and mask effectiveness on SAR.

<https://doi.org/10.1371/journal.pone.0262316.g001>

Different studies have presented SAR among non-household contacts in the range of 0%-5.3% [37, 42, 43, 51, 52]. In this study, we assume a reference SAR, $\overline{\text{SAR}}_{\text{ref}} = 5.1\%$ as the one associated with non-household unprotected contacts from the Germany outbreak study in a workplace setting [37], and assume the following associated parameters—airflow $\lambda_a = 2$ ACH, filtration $\lambda_f = 0$, speaking percentage $s_p = 25\%$, speaking volume $s_v = 65$ dB, and mask effectiveness $m_e = 0$. Fig 1 shows the resulting variation of SAR for different parameters. Determination of these parameters is based on expert judgement as little information exists in literature. The assumptions for selection of these parameters are described in S1 Appendix.

Daily contact size. Daily contact size is defined as the number of secondary contacts of each first-generation case on each day, and are randomly sampled from a Geometric distribution supported on $x \geq 0$. Given the average daily contacts \bar{c} , the probability of daily contacts $P(c = x)$ is defined as -

$$\begin{aligned}
 P(c = x) &= \text{Geometric}\left(x, \frac{1}{\bar{c}+1}\right) \quad \forall \quad x \geq 0 \\
 &= \left(1 - \frac{1}{\bar{c}+1}\right)^x \left(\frac{1}{\bar{c}+1}\right)
 \end{aligned}
 \tag{9}$$

The parameterization of the Geometric distribution is based on data collected by the citizen science-based BBC pandemic study [53, 54]. The workplace contacts in the BBC study align well with data from an independent POLYMOD study [55, 56]. Based on the BBC study, the average daily contacts at work are $\bar{c} = 6$. The data shows that daily workplace contacts are long-tailed and hence can be modeled with the Geometric distribution.

Additionally, we expect that employees contact some of the same people every day at work. The model incorporated this using a parameter for the probability that contacts remain the same the next day p_{remain} . We assumed $p_{\text{remain}} = 60\%$ based on an office study in Italy [57].

Preventive and control measures. The contact network model presented above can be effectively used to assess return-to-work policies in workplaces. Some examples of control measures and their impacts on the model are presented below and in Fig 1.

- **Work from home**—When a percentage of employees work remotely, it reduces the number of employees e in the workplace. This reduces the number of case introductions, and hence cumulative incidence. With some employees working remotely, contacts could also be reduced by the same proportion.
- **Mask wearing**—Masks reduce both the source emission and the breathing rate for virions, and is implemented in the model using the parameter m_e . Wearing masks reduces the number of virions N' , thus reducing average SAR, and hence cumulative incidence.
- **Ventilation improvement**—Increasing airflow and incorporating filtration increases the decay rate λ , resulting in reduction in average SAR.
- **Vocal loudness and speaking percentage**—Decreasing vocal loudness and the amount of talking leads to lower amounts of emitted virions S . This in turn reduces average SAR.
- **Social distancing**—Social distancing reduces the average daily contacts \bar{c} in the model, thus reducing cumulative incidence.

Individual outbreak reports from workplaces for model validation

One of the datasets used for validating our stochastic model was epidemiological reports on COVID-19 outbreaks in workplace-like situations, as shown in Table 1. Generally, information on environmental settings of these outbreaks were not reported, hence instead of selecting a single parameter value, we selected a range based on expert judgement. Table 1 summarizes the collected outbreaks and the parameters associated with them in the model. For the Henan Expressway outbreak, because the employees share the same residence, we used a fixed SAR = 20% which is similar to a household SAR, hence some of the inapplicable parameters are listed as N.A. in the table. Further description of each outbreak and the parameter selection process is provided in S1 Appendix.

For validation, the cumulative incidence from the model were compared with the observations. For each outbreak, simulation starts from when the case introduction was reported. Hence, we simulated cumulative incidence only for that subset of simulations with at least one case on the first day of the modeling period. We generated two model estimates for each outbreak—lower bound and upper bound. To generate lower bound estimates, we selected the

Table 1. Outbreak reports for workplace outbreaks. Model parameters associated with outbreak reports used for validation.

Outbreak Name	e	p_{case} (per 100k)	\bar{c}	λ_a (ACH)	s_v (dB)	s_p (%)	n_{days}	p_{remain}
1. Tianjin Office [58]	906	1	4–10	1.5–4	65	5–25	10	0.4
2. Korean Call Center [8]	216	1	4–10	1.5–4	70	55–66	14	0.4
3. San Diego VA Office [9]	100	1	4–10	1.5–4	65	5–40	8	0.4
4. Singapore Conference [59]	111	1	6–12	1.5–4	70	25–40	3	0
5a. South Dakota Meat Processing Plant (First Shift) [5]	1744	1	4–10	1–11	70	5–20	14	0.4
5b. South Dakota Meat Processing Plant (Second Shift) [5]	1459	1	4–10	1–11	70	5–20	14	0.4
6. Henan Expressway [60]	103	1	2–4	N.A.	N.A.	N.A.	13	0.4
7a. Major League Baseball Team with Social Distancing and Mask Protocols [61]	68	60	4–8	1.5–4	65	5–25	10	0.4
7b. Major League Baseball Team without Social Distancing and Mask Protocols [61]	68	60	4–10	1.5–4	65	5–25	10	0.4

<https://doi.org/10.1371/journal.pone.0262316.t001>

lower bound of average daily contacts and speaking percentage, and the upper bound of airflow, since this combination results in the lowest cumulative incidence; and vice-versa for the upper bound estimates. For instance, for the Tianjin Office outbreak, the lower bound model simulation is generated by the minimum average daily contact size $\bar{c} = 4$, the minimum speaking percentage $s_p = 5\%$, and the highest airflow $\lambda_a = 4$ ACH. Finally, if incidences were recorded on a daily basis in the reports, we additionally compared the temporal trends.

Weekly cases from nursing homes for model validation

Another dataset used for validating our model is the weekly COVID-19 data in nursing homes compiled and published since May 2020 by the Division of Nursing Homes; the Quality, Safety, and Oversight Group; and the Center for Clinical Standards and Quality at the Centers for Medicare and Medicaid Services [62].

For our analysis, we considered data associated with facilities prior to, and including the week when the first confirmed cases were reported. There are 8423 unique facilities in 2463 unique counties in the dataset for the analyzed period, totalling 89273 data entries. We assumed that resident population in a facility is equivalent to the number of beds, and staff population is 90% of the number of residents [63]. We also assumed that only staff population may be first-generation cases in nursing homes since in general, the resident population is expected to remain within the facility, while the staff population attends the facility each day similar to a workplace. We incorporated this as a modification to our stochastic model such that first-generation cases are sampled only from the staff population. We assumed the number of daily interactions in nursing homes is $\bar{c} = 6$ average contacts per person per day, and the average SAR is $\overline{\text{SAR}} = 5.1\%$. Relevant data attributes from the dataset, and data processing steps are further described in [S2 Appendix](#).

Results

In this section, we validate our model by comparing model estimates against observed data from outbreaks of COVID-19 in workplaces, and nursing home facilities.

Comparisons with outbreak reports in workplaces

[Table 2](#) summarizes the comparisons between the observations and the simulated cumulative incidences. The model outputs include the minimum and maximum cumulative incidence (shown as Range) across all simulations, the interquartile range (25th to 75th percentile), the median, and the percentile of the observation within the predicted range from both the lower bound and the upper bound models. All observations are within the ranges of model simulations, except for the Korean Call Center, the first shift of the South Dakota Meat Processing Plant, and the Major League Baseball Team outbreaks where observations exceed the lower bound estimates. Figures showing comparisons of temporal trends for 3 of the outbreaks are included in [S1 Appendix](#). Overall, the current model captures the case growth rate within 14 days post case introduction for all outbreaks. The daily observations lie within the range of the model simulations except for a few days in the Korean Call Center outbreak where the lower bound model simulations are higher than the observations. The comparisons demonstrate that all observations are within the range of cumulative incidences estimated from the upper bound parameters in the current model. This is expected since the real-world outbreaks collected here are mostly large-scale non-familial cluster cases, while sporadic cases are usually not reported. As a result, we believe that the contact network model presented in this study is reasonable for modeling short-term cumulative incidences in workplaces.

Table 2. Observations and model estimates for workplace outbreak reports.

Outbreak Name	I_{obs}	Lower Bound Model				Upper Bound Model			
		%-ile	range	IQT	median	%-ile	range	IQT	median
1. Tianjin Office	7	99.9	1–9	1–2	1	58.9	1–55	3–10	5
2. Korean Call Center	76	N.A.	1–34	2–8	4	97.9	1–135	10–38	22
3. San Diego VA Office	5	99.9	1–5	1–1	1	45	1–55	3–11	6
4. Singapore Conference	7	99	1–12	1–2	1	74	1–55	2–8	4
5a. SD Meat Processing (First Shift)	32	N.A.	1–7	1–2	1	88.9	1–102	5–21	12
5b. SD Meat Processing (Second Shift)	6	99.9	1–7	1–2	1	26.7	1–108	5–21	11
6. Henan Expressway	6	60	1–30	2–8	5	36	1–51	4–14	8
7a. Major League Baseball Team with Social Distancing and Mask Protocols	20	N.A.	1–8	1–2	1	99.8	1–28	2–6	4
7b. Major League Baseball Team without Social Distancing and Mask Protocols	20	N.A.	1–9	1–2	1	95.2	1–43	3–10	6

Observations and model outputs of cumulative incidence for outbreaks with $n_{sim} = 5000$ for each model. I_{obs} refers to the observed cumulative incidences; %-ile refers to the percentiles where the observations fall within the predicted range and N.A. indicates the observation is outside the range; IQT refers to the interquartile, or the 25th to 75th percentile range of the model output.

<https://doi.org/10.1371/journal.pone.0262316.t002>

Comparisons with weekly cases from nursing homes

In this section, we present the comparison of the model simulations with observations from the nursing home facilities. The comparisons were made for probability of case introduction, average 7-day cumulative incidence, and the distribution of cumulative incidence. Each run of the model included $n_{sim} \geq 10000$ simulations to ensure convergence of the outputs. Overall, the model estimates agreed well with the observations from nursing homes, and are described in detail below.

Probability of case introduction. We first compared the observed probability of case introduction in a facility with our model estimate. Since the nursing homes data is reported on a weekly basis, the estimated probability of at least one case being introduced in the facility in 7 days, $P_{est}(x \geq 1 | t = 7)$ is only dependent on the number of people in a facility, and the local case rate, and is computed as -

$$P_{est}(x \geq 1 | t = 7) = 1 - ((1 - p_{case})^{e_s})^7, \tag{10}$$

where e_s is the number of staff in each nursing home, and p_{case} is the daily new case rate in the county. As mentioned in the Materials and Methods section, the following analysis only considered data associated with facilities prior to, and including the week when the first confirmed cases were reported.

In order to calculate observed probabilities of case introduction, the data needs to be divided into groups. Since the estimated probability is a function of the local case rate, we binned the data by the county-level 7-day mean case rate and generated probabilities within each bin. The data were categorized into 40 bins with approximately 2200 data entries in each bin. Each entry corresponded to one facility for one week. Note that the same facility over different weeks can be categorized into different bins based on the case rate in the county for each corresponding week. Then within each bin, the probability of at least one case introduction, $P_{obs}(x \geq 1 | t = 7)$ is computed as -

$$P_{obs}(x \geq 1 | t = 7) = \frac{n_{x_{bin} \geq 1}}{N_{bin}}, \tag{11}$$

where $n_{x_{bin} \geq 1}$ is the number of data entries in a given bin with at least one confirmed case, and N_{bin} is the total number of entries in the bin.

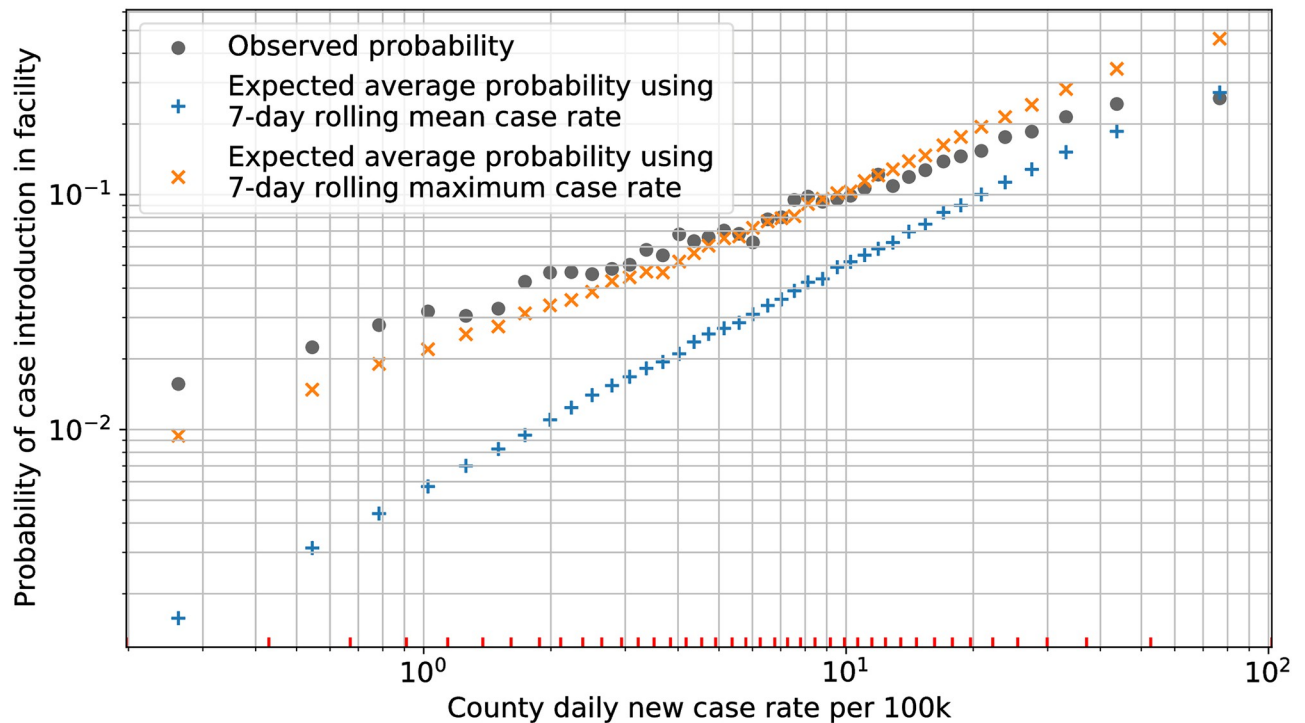


Fig 2. Probability of COVID-19 case introduction. Comparison of observed and expected probability of COVID-19 cases getting introduced in nursing homes based on the county-level case rates. Observations include all weeks prior to and including when the case introduction was reported in each facility. The red ticks on the x-axis mark the 7-day mean case-rate boundaries of each category bin.

<https://doi.org/10.1371/journal.pone.0262316.g002>

Fig 2 shows the comparison of the estimated probability, with the observed probability of case introduction in a facility within 7 days. The abscissa is for representation only and marks the average case rate within each bin. The estimated probabilities were computed separately for each facility and week, and then averaged across all data entries in each bin. Additionally, the estimated probability was computed using two different case rates for each data entry—the 7-day mean case rate, and the 7-day maximum case rate. The former represents the average of the daily new cases per 100,000 population during the previous 7-days from the reporting date of the facility, while the latter similarly represents the maximum.

We observe in Fig 2 that the observed probability of case introduction varies approximately linearly on the log-log scale with the county case rate. This matches the expected behavior for the Binomial distribution as presented in Eq 10, and hence corroborates that the assumption for Binomial distribution for case introduction is reasonable.

We also observe in the figure that the observed probability is higher than estimated probability when using the 7-day mean case rate, and matches closely with the estimate based on 7-day maximum case rate. In the dataset, the average 7-day maximum case rate within a category bin could be up to 6 times higher than its counterpart mean case rate. Since most of the cases were observed in nursing homes in the early stage of the pandemic, this comparison suggests that case rates in the United States could be under-reported due to limited testing capacity [64, 65]. Other factors contributing to the differences between observations and estimates could be our assumptions about the population in nursing homes, and that cases are introduced in facilities only from the staff. Test lagging, long incubation periods, and reporting delays could also contribute to these differences.

Average 7-day cumulative incidence. The probability of case introduction is only dependent on local case rates and population size, and does not impact the dynamics of in-facility transmission. In this section, we compare the observed weekly cases in facilities with the model estimates. The observations represent cases from both case introductions and in-facility transmission, and these contributions cannot be differentiated.

The same dataset and binning methodology was used as described for the previous comparison. Within each bin, the average of confirmed cases across all data entries was calculated. This average encompasses the zero cases observed in facilities during the weeks prior to their respective case introduction. This average value represents the average of observed 7-day cumulative incidence in the nursing homes corresponding to each category bin.

For each bin, a model was run with a case rate calculated using the geometric mean of the case rates across all data entries in the bin. The geometric mean is equivalent to taking the average in log-space and was selected because it is more robust when bin sizes are large, mainly in the first and the last category bins. Since each facility in a category bin can have a different population, the average number of beds across all data entries in a bin was used as an approximation for the average number of residents in a facility for that bin. For each bin, the average number of estimated cases was calculated by taking the average across all model simulations of the 7-day cumulative incidence.

Fig 3 shows the comparison of the average number of confirmed cases in nursing homes within each bin, with the average estimate of cases from model simulations for each bin. Within each bin, the model estimates are generated with two different case rates—geometric mean of the 7-day mean case rate, and geometric mean of the 7-day maximum case rate. The

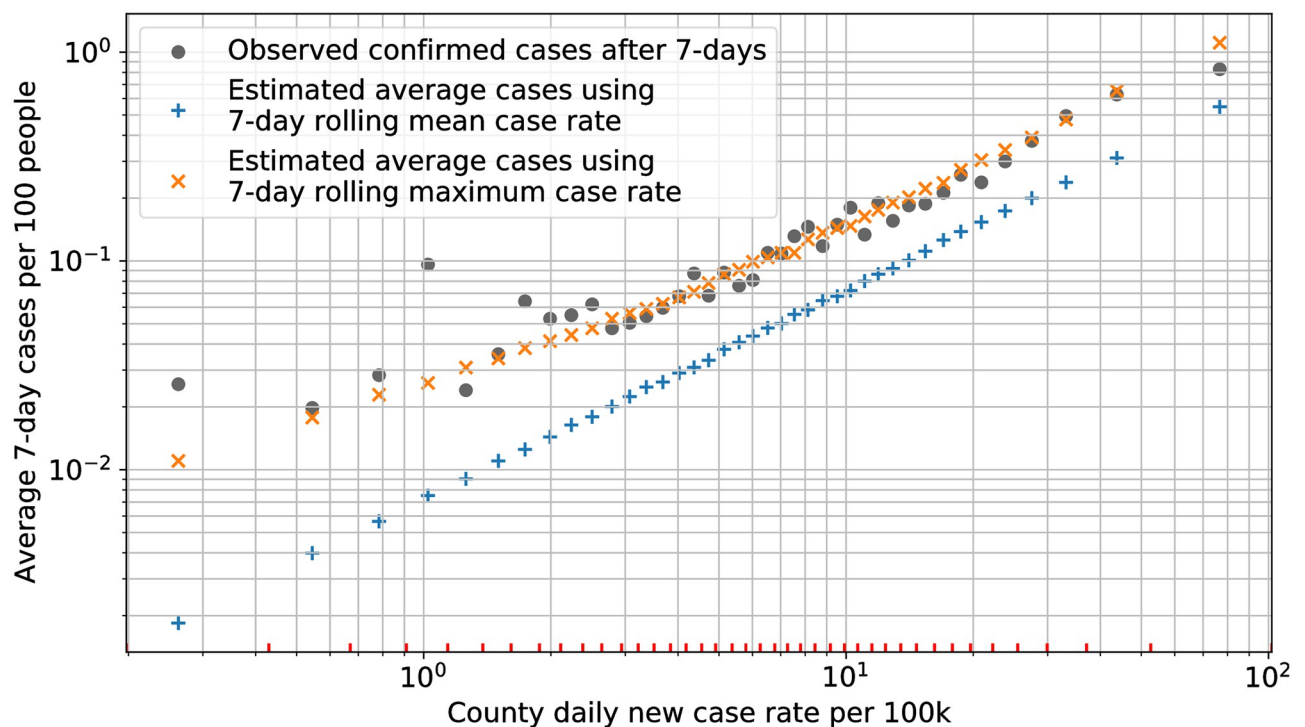


Fig 3. Average 7-day cumulative incidence. Comparison of average confirmed cases in nursing homes prior to and including the week of case introduction, with the model estimates of average cases in 1 week. The red ticks on the x-axis mark the 7-day mean case-rate boundaries of each category bin.

<https://doi.org/10.1371/journal.pone.0262316.g003>

observed data and both the model estimates are plotted on the same abscissa marking the geometric mean of the 7-day mean case rate for each bin.

Similar to the previous comparison, we observe from the figure that the average confirmed cases are higher than the estimated cases when using the 7-day mean case rate, and match closely with the estimate based on 7-day maximum case rate. The same factors as described previously are likely to contribute to this variation. Additionally, our assumptions about the daily interactions and the SAR parameters in nursing homes could also contribute to the differences.

Distribution of 7-day cumulative incidence. Finally, we compared the distribution of cases in nursing homes with the distribution of cases generated from model simulations. Given the closer correlation of observation averages and model averages for 7-day maximum case rates, this analysis was done only based on the 7-day maximum case rates.

The comparison of the distribution of cases between observations and model estimates was done only for facilities with at least one reported case. We divided these ≈ 8000 facilities into 8 category bins by their 7-day maximum case rate, with ≈ 1000 facilities in each bin. A separate model analysis was run for each category bin using the geometric mean of 7-day maximum case rate, and average number of beds for all facilities in the bin, as described previously. The distribution of confirmed cases in each unique facility within a bin during the week of case introduction were compared with the 7-day cumulative incidence outputs from the model only for those simulations that had at least one case at the end of the 7th day. The comparison for each of the bins is shown in Fig 4. We observe a strong correlation between the distribution of cases across facilities and the distribution of cases based on model simulations. This confirms that the model presented in this study provides a good basis for estimating the distribution of cases within a workplace over a 7-day period.

A similar comparison was done for the distribution of cumulative incidence after 14 days, with similar results, and shown in S2 Appendix. Based on our comparison with observations from nursing homes, we conclude that the model provides reasonable estimates for spread of COVID-19 within facilities for short forecast periods of up to 15 days.

Discussion

In this section, we present the key limitations of the current study, and motivations for future work.

Limitations of the model

The current model has the following limitations -

- The model only keeps track of the first and second generations of infections. However, evidence suggests that it is possible that multiple generations of transmission events could occur within a short period. In the report for the Tianjin Office outbreak [58], the third-generation cases occurred on the 7th day post case introduction. However, for most outbreaks, we expect limited transmission events within the intended short forecast period, as evidenced by the similarities in model outputs with observations. We have included an assessment of this assumption by comparing the results with a preliminary model that includes third-generation cases in S5 Appendix. Additionally, we expect that the underestimation of cumulative incidence by ignoring third-generation cases is somewhat balanced by the overestimation due to first-generation cases being infectious for the entire forecast period. Multiple studies show a highly variable infectious period distribution ranging from 5 to 20 days depending on the viable viral load [66–69].

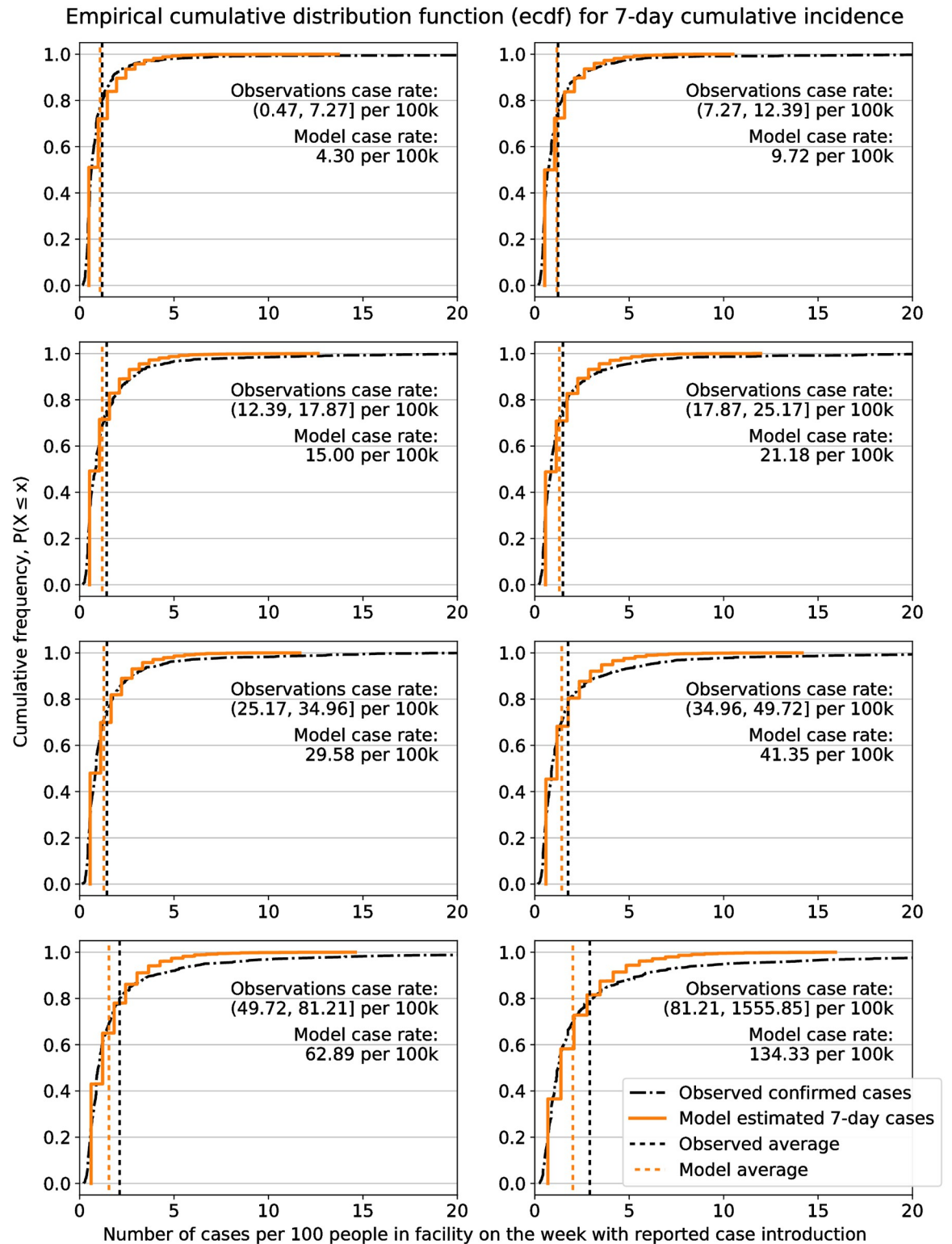


Fig 4. Distribution of 7-day cumulative incidence. Comparison of the distribution of cases per 100 people in nursing homes during the week of case introduction, and 7-day model simulations. Each subplot describes the distribution for facilities within one case rate category bin, and for the model based on corresponding mean case rate for that bin. Vertical dashed lines represent the average cases for the facilities and the simulations in each bin.

<https://doi.org/10.1371/journal.pone.0262316.g004>

- We did not differentiate among the droplet, aerosol, and direct contact modes of COVID-19 transmission [70, 71]. There is evidence of long-distance COVID-19 airborne transmission [72, 73]. However, the advantage of this abstraction is that it did not require consideration of different SAR for different interactions, and allowed us to adjust SAR based on airflow and filtration. Overall, since we are focusing on outbreaks in workplaces where people are expected to be in close distance less frequently than in household settings, they are more likely to be infected with aerosol particles that travel longer distances and stay longer in the air [74–77].
- The use of air changes per hour within facilities is a simplified proxy for estimating the effects of airflow. A study on airborne transmission risk in practical settings [20] demonstrated that the location of vents also plays a critical role in aerosol transmission risk.

Limitations of the data

- Underreporting, undertesting, and lagging of tests are common issues in reporting of COVID-19 data, especially prior to March 2020 when there was a lack of testing supplies for RT-PCR tests in the United States [78, 79]. The RT-PCR screening test of COVID-19 takes one to several days depending on the laboratory capacity [80, 81]. Additionally, it is likely that asymptomatic individuals are not getting tested. Therefore, the case rate data used in the model may not reflect the actual incidence rates.
- We did not explicitly assess asymptomatic cases in the current model. The proportion of asymptomatic cases can change the implementation of SAR distribution [82, 83]. However, this proportion is associated with numerous factors, including virus genetics [84], host factors [85, 86], generation time [32, 87], etc., and this information is not readily available in epidemiological reports of COVID-19 outbreaks.
- Detailed temporal data on outbreaks in workplaces is scarce. There have been a small number of published epidemiological reports with comprehensive information about major outbreaks, however to our knowledge most state-level public health departments in the United States do not specifically publish data for workplace outbreaks. Comprehensive data collection and publication for outbreak events in various workplaces would be highly valuable for policy makers, businesses, and researchers to facilitate disease surveillance and monitoring, and to evaluate the transmission risk in the workplace environment.

Future work

The model presented here provides a framework to comprehensively account for the uncertainties in the spread of COVID-19, so that the number of cases can be estimated for small populations and short durations. The underlying framework can be expanded further to incorporate other capabilities.

- Workplaces like restaurants and retail stores can be modeled by removing the limitation of a static population. As a result, instead of first-generation cases only being introduced in the staff population, they can also be introduced by external visitors, for example, customers in a restaurant.
- Heterogeneous interactions can be incorporated in the contact network, for example, to model different extent of interactions within and outside different departments, or between

different population groups like staff and residents in nursing homes; and to model events like meetings and meals where many people interact at the same time.

- Heterogeneity in SAR can be incorporated for different population groups, for example, by age, sex, presence of symptoms, and vaccination status, and for different modes of transmission, like droplet, airborne, or fomite.
- We modeled the long-tail behavior of SAR distribution using the Beta distribution. This choice of distribution can be evaluated as more information becomes available from contact tracing and in-depth analysis on SAR.
- Longer duration forecasts can be modeled by incorporating incubation period and time to recovery for exposed and infectious individuals. This would enable modeling third and subsequent generations of transmission, and intervention measures like quarantining and isolation during the forecast duration.
- SAR and other parameters can be changed to model the impact of emerging COVID-19 variants [88, 89], and other airborne diseases.

Conclusions

We presented a stochastic contact network model to forecast the spread of COVID-19 over short duration of ≤ 15 days, and for small populations within workplaces. The model incorporated cases arriving at a workplace from the local community, and the secondary transmission between the infectious and susceptible populations in the facility. Introduction of newly infected individuals was modeled probabilistically through a Bernoulli process with the success rate as the local daily new case rate. Transmission within the facility was determined by probabilistic realizations of both the number of secondary contacts for each infectious individual, and their corresponding secondary attack rates (SAR). By running multiple simulations with realizations from the probability distributions of these model parameters, we obtained a distribution of predicted cumulative incidence within a facility over the forecast period.

The short forecast duration was chosen in this study because we expect that interventions are often taken in workplaces soon after identification of cases, and as a result, longer duration forecasts may have higher variability and thus be less informative for policy and decision makers.

The model presented here was validated with respect to two different datasets. In the first part, we evaluated whether observations of cumulative incidence from 7 real-world outbreaks in workplaces could be estimated by the model. For all outbreaks examined, the observations lied within the range of model simulations. In the second part, we compared the model with weekly case reports from over 8000 nursing homes in the United States. We observed a strong correlation between the observed probability of first-generation individuals being introduced in the facility, and the average cumulative incidence after 7 days. We also compared the distribution of cumulative incidence after 7 days, and found a strong correlation with the distributions from model simulations.

We demonstrated that the contact network model presented in this study is a simplified but reasonable representation of the cumulative incidence of COVID-19 within workplaces for up to 15 days and populations as small as 50 individuals. We believe that it provides a consistent framework to account for uncertainties in disease transmission, serve as a data-driven tool for occupational safety, and can be expanded to model outbreaks for emerging COVID-19 variants and other airborne diseases.

The model described here is implemented in a web application, available at covid19.oneconcern.com. The open source model code implemented in python:3 is available at github.com/abhineetgupta/covid-workplace-risk.

Supporting information

S1 Appendix. Individual outbreak reports from workplaces for model validation.
(PDF)

S2 Appendix. Nursing homes dataset for model validation.
(PDF)

S3 Appendix. Relation of secondary attack rate with respect to source emissions and inhaled virions.
(PDF)

S4 Appendix. Model convergence.
(PDF)

S5 Appendix. Comparison with a model with additional generation.
(PDF)

Acknowledgments

We would like to thank Maura Sullivan, Arnold Kalmbach, Kyle Weston, Chris Martin, William Letton, and Cassie Springate for their help and guidance.

Author Contributions

Conceptualization: Xi Guo, Abhineet Gupta, Chengwei Zhai.

Data curation: Xi Guo, Abhineet Gupta, Chengwei Zhai.

Formal analysis: Xi Guo, Abhineet Gupta.

Methodology: Xi Guo, Abhineet Gupta, Chengwei Zhai.

Project administration: Abhineet Gupta, Anand Sampat.

Software: Xi Guo, Abhineet Gupta, Anand Sampat.

Supervision: Abhineet Gupta.

Validation: Xi Guo, Abhineet Gupta.

Visualization: Xi Guo, Abhineet Gupta.

Writing – original draft: Xi Guo, Abhineet Gupta, Anand Sampat, Chengwei Zhai.

Writing – review & editing: Xi Guo, Abhineet Gupta, Anand Sampat, Chengwei Zhai.

References

1. Ritchie H, Ortiz-Ospina E, Beltekian D, Mathieu E, Hasell J, Macdonald B, et al. Coronavirus Pandemic (COVID-19); 2021. Available from: <https://ourworldindata.org/coronavirus>.
2. Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, et al. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *The Lancet Infectious Diseases*. 2020; 20(9):e238–e244. [https://doi.org/10.1016/S1473-3099\(20\)30484-9](https://doi.org/10.1016/S1473-3099(20)30484-9) PMID: 32628905
3. Sternlicht A. Over 10,000 Tyson Employees Reportedly Test Positive For Covid. *Forbes*. 2020;.
4. Bradbury S. How coronavirus spread through JBS's Greeley beef plant. *The Denver Post*. 2020;.

5. Steinberg J, Kennedy ED, Basler C, Grant MP, Jacobs JR, Ortbahn D, et al. COVID-19 Outbreak Among Employees at a Meat Processing Facility—South Dakota, March–April 2020. *MMWR Morbidity and Mortality Weekly Report*. 2020; 69(31):1015–1019. <https://doi.org/10.15585/mmwr.mm6931a2> PMID: 32759914
6. Lerman R. Amazon says nearly 20,000 employees have caught the coronavirus. *Washington Post*. 2020;.
7. Fox MD. Response to a COVID-19 Outbreak on a University Campus - Indiana, August 2020. *MMWR Morbidity and Mortality Weekly Report*. 2021; 70. <https://doi.org/10.15585/mmwr.mm7004a3>
8. Park SY, Kim YM, Yi S, Lee S, Na BJ, Kim CB, et al. Coronavirus Disease Outbreak in Call Center, South Korea. *Emerging Infectious Diseases*. 2020; 26(8):1666–1670. <https://doi.org/10.3201/eid2608.201274> PMID: 32324530
9. Castellano J. How A COVID-19 Outbreak Unfolded At A San Diego VA Office; 2020. Available from: <https://www.kpbs.org/news/2020/may/21/how-covid-19-outbreak-unfolded-san-diego-va-office/>.
10. Caglar D, Couto V, Faccio E, Sethi B. Business needs a tighter strategy for remote work; 2021. Available from: <https://www.pwc.com/us/en/library/covid-19/us-remote-work-survey.html>.
11. Dua A, Cheng WL, Lund S, Smet AD, Robinson O, Sanghvi S. What 800 executives envision for the postpandemic workforce; 2020. Available from: <https://www.mckinsey.com/featured-insights/future-of-work/what-800-executives-envision-for-the-postpandemic-workforce>.
12. Lund S, Madgavkar A, Manyika J, Smit S, Ellingrud K, Meaney M, et al. The future of work after COVID-19. McKinsey Global Institute; 2021. Available from: <https://www.mckinsey.com/featured-insights/future-of-work/the-future-of-work-after-covid-19>.
13. Boland B, Smet AD, Palter R, Sanghvi A. Reimagining the office and work life after COVID-19; 2020. Available from: <https://www.mckinsey.com/business-functions/organization/our-insights/reimagining-the-office-and-work-life-after-covid-19>.
14. Lund S, Madgavkar A, Manyika J, Smit S. What's next for remote work: An analysis of 2,000 tasks, 800 jobs, and nine countries; 2020. Available from: <https://www.mckinsey.com/featured-insights/future-of-work/whats-next-for-remote-work-an-analysis-of-2000-tasks-800-jobs-and-nine-countries>.
15. Gambell T, Hughes J, Meigs B, Thanopoulou A. How US companies are planning for a safe return to the workplace; 2020. Available from: <https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/how-us-companies-are-planning-for-a-safe-return-to-the-workplace>.
16. Azuma K, Yanagi U, Kagi N, Kim H, Ogata M, Hayashi M. Environmental factors involved in SARS-CoV-2 transmission: effect and role of indoor environmental quality in the strategy for COVID-19 infection control. *Environmental Health and Preventive Medicine*. 2020; 25(1):66. <https://doi.org/10.1186/s12199-020-00904-2> PMID: 33143660
17. Liu F, Qian H, Luo Z, Zheng X. The impact of indoor thermal stratification on the dispersion of human speech droplets. *Indoor Air*. 2021; 31(2):369–382. <https://doi.org/10.1111/ina.12737> PMID: 32869358
18. Keeling MJ, Rohani P. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press; 2008. Available from: <http://www.jstor.org/stable/10.2307/j.ctvc4gk0>.
19. Prentiss M, Chu A, Berggren KK. Superspreading Events Without Superspreaders: Using High Attack Rate Events to Estimate N^0 for Airborne Transmission of COVID-19. *medRxiv*. 2020; p. 2020.10.21.20216895.
20. Shao S, Zhou D, He R, Li J, Zou S, Mallery K, et al. Risk assessment of airborne transmission of COVID-19 by asymptomatic individuals under different practical settings. *Journal of Aerosol Science*. 2021; 151:105661. <https://doi.org/10.1016/j.jaerosci.2020.105661> PMID: 32968325
21. Jenness SM, Willebrand KS, Malik AA, Lopman BA, Omer SB. Dynamic Network Strategies for SARS-CoV-2 Control on a Cruise Ship. *medRxiv*. 2020; p. 2020.08.26.20182766.
22. Kerr CC, Stuart RM, Mistry D, Abey Suriya RG, Rosenfeld K, Hart GR, et al. Covasim: an agent-based model of COVID-19 dynamics and interventions. *medRxiv*. 2021; p. 2020.05.10.20097469.
23. Xiao Y, Yang M, Zhu Z, Yang H, Zhang L, Ghader S. Modeling indoor-level non-pharmaceutical interventions during the COVID-19 pandemic: a pedestrian dynamics-based microscopic simulation approach. *arXiv:2006.10666 [physics]*. 2020;.
24. Weitz JS, Andris C. COVID-19 Event Risk Assessment Planning Tool; 2021. Available from: <https://covid19risk.biosci.gatech.edu/>.
25. Bazant MZ, Bush JWM. Beyond Six Feet: A Guideline to Limit Indoor Airborne Transmission of COVID-19. *medRxiv*. 2021; p. 2020.08.26.20182824.
26. Olsson C, Oremán J, Lindmark R, Tsykalova A, Bachar S, Dobro S, et al. microCOVID Project; 2021. Available from: <https://www.microcovid.org/>.

27. Chande A, Lee S, Harris M, Nguyen Q, Beckett SJ, Hilley T, et al. Real-time, interactive website for US-county-level COVID-19 event risk assessment. *Nature Human Behaviour*. 2020; 4(12):1313–1319. <https://doi.org/10.1038/s41562-020-01000-9> PMID: 33168955
28. Singer BJ, Thompson RN, Bonsall MB. The effect of the definition of “pandemic” on quantitative assessments of infectious disease outbreak risk. *Scientific Reports*. 2021; 11(1):2547. <https://doi.org/10.1038/s41598-021-81814-3> PMID: 33510197
29. Griffin RH, Nunn CL. Community structure and the spread of infectious disease in primate social networks. *Evolutionary Ecology*. 2012; 26(4):779–800. <https://doi.org/10.1007/s10682-011-9526-2>
30. Liu F, Li X, Zhu G. Using the contact network model and Metropolis-Hastings sampling to reconstruct the COVID-19 spread on the “Diamond Princess”. *Science Bulletin*. 2020; 65(15):1297–1305. <https://doi.org/10.1016/j.scib.2020.04.043> PMID: 32373394
31. Mentus C, Romeo M, DiPaola C. Analysis and Applications of Adaptive Group Testing Methods for COVID-19. *medRxiv*. 2020; p. 2020.04.05.20050245.
32. Mizumoto K, Kagaya K, Chowell G. Effect of a wet market on coronavirus disease (COVID-19) transmission dynamics in China, 2019–2020. *International Journal of Infectious Diseases*. 2020; 97:96–101. <https://doi.org/10.1016/j.ijid.2020.05.091> PMID: 32497812
33. Ridenhour B, Kowalik JM, Shay DK. Unraveling R0: Considerations for Public Health Applications. *American Journal of Public Health*. 2014; 104(2):e32–e41. <https://doi.org/10.2105/AJPH.2013.301704> PMID: 24328646
34. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases*. 2020; 20(8):911–919. [https://doi.org/10.1016/S1473-3099\(20\)30287-5](https://doi.org/10.1016/S1473-3099(20)30287-5) PMID: 32353347
35. Cauchemez S, Donnelly CA, Reed C, Ghani AC, Fraser C, Kent CK, et al. Household Transmission of 2009 Pandemic Influenza A (H1N1) Virus in the United States. *New England Journal of Medicine*. 2009; 361(27):2619–2627. <https://doi.org/10.1056/NEJMoa0905498> PMID: 20042753
36. Li W, Zhang B, Lu J, Liu S, Chang Z, Peng C, et al. Characteristics of Household Transmission of COVID-19. *Clinical Infectious Diseases*. 2020; 71(8):1943–1946. <https://doi.org/10.1093/cid/ciaa450> PMID: 32301964
37. Böhmer MM, Buchholz U, Corman VM, Hoch M, Katz K, Marosevic DV, et al. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *The Lancet Infectious Diseases*. 2020;. [https://doi.org/10.1016/S1473-3099\(20\)30314-5](https://doi.org/10.1016/S1473-3099(20)30314-5) PMID: 32422201
38. James A. High COVID-19 Attack Rate Among Attendees at Events at a Church—Arkansas, March 2020. *MMWR Morbidity and Mortality Weekly Report*. 2020; 69. <https://doi.org/10.15585/mmwr.mm6920e2>
39. Hamner L. High SARS-CoV-2 Attack Rate Following Exposure at a Choir Practice—Skagit County, Washington, March 2020. *MMWR Morbidity and Mortality Weekly Report*. 2020; 69. <https://doi.org/10.15585/mmwr.mm6919e6>
40. Zhang XA, Fan H, Qi RZ, Zheng W, Zheng K, Gong JH, et al. Importing coronavirus disease 2019 (COVID-19) into China after international air travel. *Travel Medicine and Infectious Disease*. 2020; 35:101620. <https://doi.org/10.1016/j.tmaid.2020.101620> PMID: 32222421
41. Shen Y, Xu W, Li C, Handel A, Martinez L, Ling F, et al. A Cluster of Novel Coronavirus Disease 2019 Infections Indicating Person-to-Person Transmission Among Casual Contacts From Social Gatherings: An Outbreak Case-Contact Investigation. *Open Forum Infectious Diseases*. 2020; 7(6):ofaa231. <https://doi.org/10.1093/ofid/ofaa231> PMID: 32613025
42. Burke RM. Active Monitoring of Persons Exposed to Patients with Confirmed COVID-19—United States, January–February 2020. *MMWR Morbidity and Mortality Weekly Report*. 2020; 69. <https://doi.org/10.15585/mmwr.mm6909e1>
43. Koh WC, Naing L, Chaw L, Rosledzana MA, Alikhan MF, Jamaludin SA, et al. What do we know about SARS-CoV-2 transmission? A systematic review and meta-analysis of the secondary attack rate and associated risk factors. *PLOS ONE*. 2020; 15(10):e0240205. <https://doi.org/10.1371/journal.pone.0240205> PMID: 33031427
44. Hijnen D, Marzano AV, Eyerich K, GeurtsvanKessel C, Giménez-Arnau AM, Joly P, et al. SARS-CoV-2 Transmission from Presymptomatic Meeting Attendee, Germany. *Emerging Infectious Diseases—Research Letters*. 2020; 26(8). <https://doi.org/10.3201/eid2608.201235> PMID: 32392125
45. Wu J, Huang Y, Tu C, Bi C, Chen Z, Luo L, et al. Household Transmission of SARS-CoV-2, Zhuhai, China, 2020. *Clinical Infectious Diseases*. 2020; 71(16):2099–2108. <https://doi.org/10.1093/cid/ciaa557> PMID: 32392331

46. Qiu X, Nergiz AI, Maraolo AE, Bogoch II, Low N, Cevik M. Defining the role of asymptomatic and pre-symptomatic SARS-CoV-2 transmission—a living systematic review. *Clinical Microbiology and Infection*. 2021; 0(0).
47. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature*. 2005; 438(7066):355–359. <https://doi.org/10.1038/nature04153> PMID: 16292310
48. Asadi S, Wexler AS, Cappa CD, Barreda S, Bouvier NM, Ristenpart WD. Aerosol emission and superemission during human speech increase with voice loudness. *Scientific Reports*. 2019; 9(1):2348. <https://doi.org/10.1038/s41598-019-38808-z> PMID: 30787335
49. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet*. 2020; 395(10242):1973–1987. [https://doi.org/10.1016/S0140-6736\(20\)31142-9](https://doi.org/10.1016/S0140-6736(20)31142-9) PMID: 32497510
50. Gandhi M, Marr LC. Uniting Infectious Disease and Physical Science Principles on the Importance of Face Masks for COVID-19. *Med*. 2021; 2(1):29–32. <https://doi.org/10.1016/j.medj.2020.12.008> PMID: 33521753
51. Chaw L, Koh WC, Jamaludin SA, Naing L, Alikhan MF, Wong J. Analysis of SARS-CoV-2 transmission in different settings, among cases and close contacts from the Tablighi cluster in Brunei Darussalam. medRxiv. 2020; p. 2020.05.04.20090043.
52. Cheng HY, Jian SW, Liu DP, Ng TC, Huang WT, Team fTCoi, et al. High transmissibility of COVID-19 near symptom onset. medRxiv. 2020; p. 2020.03.18.20034561.
53. Klepac P, Kucharski AJ, Conlan AJ, Kissler S, Tang ML, Fry H, et al. Contacts in context: large-scale setting-specific social mixing matrices from the BBC Pandemic project. medRxiv. 2020; p. 2020.02.16.20023754.
54. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*. 2020; 20(5):553–558. [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4) PMID: 32171059
55. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. POLYMOD social contact data; 2020. Available from: <https://zenodo.org/record/3874557#.YEVy37RKhTY>.
56. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Medicine*. 2008; 5(3). <https://doi.org/10.1371/journal.pmed.0050074> PMID: 18366252
57. Génois M, Vestergaard CL, Fournet J, Panisson A, Bonmarin I, Barrat A. Data on face-to-face contacts in an office building suggests a low-cost vaccination strategy based on community linkers. *Network Science*. 2015; 3(3):326–347. <https://doi.org/10.1017/nws.2015.10>
58. Zhang Y, Su X, Chen W, Fei CN, Guo LR, Wu XL, et al. Epidemiological investigation on a cluster epidemic of COVID-19 in a collective workplace in Tianjin. *Zhonghua Liu Xing Bing Xue Za Zhi = Zhonghua Liuxingbingxue Zazhi*. 2020; 41(5):648–652. PMID: 32162858
59. Pung R, Chiew CJ, Young BE, Chin S, Chen MIC, Clapham HE, et al. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. *The Lancet*. 2020; 395(10229):1039–1046. [https://doi.org/10.1016/S0140-6736\(20\)30528-6](https://doi.org/10.1016/S0140-6736(20)30528-6) PMID: 32192580
60. Pang Q, Li P, Li T, Li H. Investigation on a cluster epidemic of corona virus diseases 2019 (COVID-19) in expressway service area. *Anhui J Prev Med*. 2020; 26(2):130–132.
61. Murray MT, Riggs MA, Engelthaler DM, Johnson C, Watkins S, Longenberger A, et al. Mitigating a COVID-19 Outbreak Among Major League Baseball Players. *MMWR Morbidity and Mortality Weekly Report*. 2020; 69(42):1542–1546. PMID: 33090983
62. Centers for Medicare and Medicaid Services—Division of Nursing Homes/Quality, Safety, and Oversight Group/Center for Clinical Standards and Quality. COVID-19 Nursing Home Dataset; 2021. Available from: <https://data.cms.gov/Special-Programs-Initiatives-COVID-19-Nursing-Home/COVID-19-Nursing-Home-Dataset/s2uc-8wxxp>.
63. Gabrel CS. An Overview of Nursing Home Facilities: Data from the 1997 National Nursing Home Survey. *Advance Data*. 2000;(311):12. PMID: 10977763
64. Ketchum K, O'Connor L. COVID-19 testing problems started early, U.S. still playing from behind. *Modern Healthcare*. 2020;.
65. American Society of Microbiology (ASM). Supply Shortages Impacting COVID-19 and Non-COVID Testing; 2021. Available from: <https://asm.org/Articles/2020/September/Clinical-Microbiology-Supply-Shortage-Collecti-1>.

66. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. *Nature*. 2020; 581(7809):465–469. <https://doi.org/10.1038/s41586-020-2196-x> PMID: 32235945
67. Ferretti L, Ledda A, Wymant C, Zhao L, Ledda V, Abeler L, et al. The timing of COVID-19 transmission. *medRxiv*. 2020; p. 20.
68. Pan Y, Zhang D, Yang P, Poon LLM, Wang Q. Viral load of SARS-CoV-2 in clinical samples. *The Lancet Infectious Diseases*. 2020; 20(4):411–412. [https://doi.org/10.1016/S1473-3099\(20\)30113-4](https://doi.org/10.1016/S1473-3099(20)30113-4) PMID: 32105638
69. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*. 2020; 26(5):672–675. <https://doi.org/10.1038/s41591-020-0869-5> PMID: 32296168
70. Vella F, Senia P, Ceccarelli M, Vitale E, Maltezou H, Taibi R, et al. Transmission mode associated with coronavirus disease 2019: A review. *European review for medical and pharmacological sciences*. 2020; 24:7889–7904. PMID: 32744718
71. Karia R, Gupta I, Khandait H, Yadav A, Yadav A. COVID-19 and its Modes of Transmission. *SN Comprehensive Clinical Medicine*. 2020; 2(10):1798–1801. <https://doi.org/10.1007/s42399-020-00498-4> PMID: 32904860
72. Nissen K, Krambrich J, Akaberi D, Hoffman T, Ling J, Lundkvist Å, et al. Long-distance airborne dispersal of SARS-CoV-2 in COVID-19 wards. *Scientific Reports*. 2020; 10(1):19589. <https://doi.org/10.1038/s41598-020-76442-2> PMID: 33177563
73. Setti L, Passarini F, De Gennaro G, Barbieri P, Perrone MG, Borelli M, et al. Airborne Transmission Route of COVID-19: Why 2 Meters/6 Feet of Inter-Personal Distance Could Not Be Enough. *International Journal of Environmental Research and Public Health*. 2020; 17(8):2932. <https://doi.org/10.3390/ijerph17082932> PMID: 32340347
74. Morawska L, Cao J. Airborne transmission of SARS-CoV-2: The world should face the reality. *Environment International*. 2020; 139:105730. <https://doi.org/10.1016/j.envint.2020.105730> PMID: 32294574
75. Qian H, Miao T, Liu L, Zheng X, Luo D, Li Y. Indoor transmission of SARS-CoV-2. *Indoor Air*. 2020; p. ina.12766. PMID: 33131151
76. Noorimotlagh Z, Jaafarzadeh N, MartÃ-nez SS, Mirzaee SA. A systematic review of possible airborne transmission of the COVID-19 virus (SARS-CoV-2) in the indoor air environment. *Environmental Research*. 2021; 193:110612. <https://doi.org/10.1016/j.envres.2020.110612> PMID: 33309820
77. Azimi P, Keshavarz Z, Laurent JGC, Stephens B, Allen JG. Mechanistic transmission modeling of COVID-19 on the Diamond Princess cruise ship demonstrates the importance of aerosol transmission. *Proceedings of the National Academy of Sciences*. 2021; 118(8). <https://doi.org/10.1073/pnas.2015482118> PMID: 33536312
78. Lau H, Khosrawipour T, Kocbach P, Ichii H, Bania J, Khosrawipour V. Evaluating the massive underreporting and undertesting of COVID-19 cases in multiple global epicenters. *Pulmonology*. 2021; 27(2):110–115. <https://doi.org/10.1016/j.pulmoe.2020.05.015> PMID: 32540223
79. Hamidi S, Ewing R, Sabouri S. Longitudinal analyses of the relationship between development density and the COVID-19 morbidity and mortality rates: Early evidence from 1,165 metropolitan counties in the United States. *Health & Place*. 2020; 64:102378. <https://doi.org/10.1016/j.healthplace.2020.102378> PMID: 32738578
80. University of Colorado Health. Testing for COVID-19; 2020. Available from: <https://www.uchealth.org/services/infectious-diseases/coronavirus-covid-19/covid-19-testing/>.
81. Labcorp. How to Get My COVID-19 Test Result; 2021. Available from: <https://www.labcorp.com/coronavirus-disease-covid-19/individuals/test-results>.
82. Byambasuren O, Cardona M, Bell K, Clark J, McLaws ML, Glasziou P. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: Systematic review and meta-analysis. *Official Journal of the Association of Medical Microbiology and Infectious Disease Canada*. 2020;. <https://doi.org/10.3138/jammi-2020-0030>
83. Pollock AM, Lancaster J. Asymptomatic transmission of covid-19. *BMJ*. 2020; 371:m4851. <https://doi.org/10.1136/bmj.m4851>
84. Wang R, Chen J, Hozumi Y, Yin C, Wei GW. Decoding Asymptomatic COVID-19 Infection and Transmission. *The Journal of Physical Chemistry Letters*. 2020; 11(23):10007–10015. <https://doi.org/10.1021/acs.jpcclett.0c02765> PMID: 33179934
85. Lai CC, Liu YH, Wang CY, Wang YH, Hsueh SC, Yen MY, et al. Asymptomatic carrier state, acute respiratory disease, and pneumonia due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): Facts and myths. *Journal of Microbiology, Immunology, and Infection = Wei Mian Yu Gan Ran Za Zhi*. 2020; 53(3):404–412. <https://doi.org/10.1016/j.jmii.2020.02.012> PMID: 32173241

86. Yu C, Zhou M, Liu Y, Guo T, Ou C, Yang L, et al. Characteristics of asymptomatic COVID-19 infection and progression: A multicenter, retrospective study. *Virulence*. 2020; 11(1):1006–1014. <https://doi.org/10.1080/21505594.2020.1802194> PMID: 32722990
87. Tao J, Zhang X, Zhang X, Zhao S, Yang L, He D, et al. The time serial distribution and influencing factors of asymptomatic COVID-19 cases in Hong Kong. *One Health*. 2020; 10:100166. <https://doi.org/10.1016/j.onehit.2020.100166> PMID: 32904481
88. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole Á, et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell*. 2021; 184(1):64–75.e11. <https://doi.org/10.1016/j.cell.2020.11.020> PMID: 33275900
89. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*. 2020; p. 2020.12.21.20248640.