

Analysis of overdispersion in airborne transmission of Covid-19

Swetaprovo Chaudhuri ^{*1}, Prasad Kasibhatla², Arnab Mukherjee¹, William Pan³, Glenn Morrison⁴, Sharmistha Mishra⁵, and Vijaya Kumar Murty^{6,7}

¹Institute for Aerospace Studies, University of Toronto

²Nicholas School of Environment, Duke University

³Global Institute of Health, Duke University

⁴Gillings School of Global Public Health, University of North Carolina

⁵Department of Medicine, Division of Infectious Diseases, University of Toronto

⁶The Fields Institute for Research in Mathematical Sciences

⁷Department of Mathematics, University of Toronto

Abstract

Superspreading events and overdispersion are hallmarks of the Covid-19 pandemic. To gain insight into the nature and controlling factors of these superspreading events and heterogeneity in transmission, we conducted mechanistic modeling of SARS-CoV-2 transmission by infectious aerosols using real-world occupancy data from a large number of full-service restaurants in ten US metropolises. Including a large number of factors that influence disease transmission in these settings, we demonstrate the emergence of a stretched tail in the probability density function of secondary infection

*Corresponding author, email: schaudhuri@utias.utoronto.ca

numbers indicating strong heterogeneity in individual infectivity. Derived analytical results further demonstrate that variability in viral loads and variability in occupancy, together, lead to overdispersion in the number of secondary infections arising from individual index cases. Our analysis, connecting mechanistic understanding of SARS-CoV-2 transmission by aerosols with observed large-scale epidemiological characteristics of Covid-19 outbreaks, adds an important dimension to the mounting body of evidence with regards to the determinants of airborne transmission of SARS-CoV-2 by aerosols in indoor settings.

1 Introduction

Superspreading events and overdispersion are now well-established characteristics of the Covid-19 pandemic, similar to SARS and other outbreaks of respiratory viruses [1]. The documented outbreaks at the Skagit Valley Chorale [2, 3], at a restaurant in Guangzhou [4, 5] and at a call center in Korea [6] are examples of superspreading events where one infectious index case led to tens of individuals infected within a few hours, at an order of magnitude higher than the basic reproduction number $2 \leq \mathcal{R}_0 \leq 3.6$ for the original SARS-CoV-2 variant [7]. The term "superspreading" has generally been ascribed to any event (index case and exposures) that led to more than the average number of secondary transmissions, and thus in probabilistic terms could refer to any number of secondary cases to the right of the expectation [1]. As such, it has been proposed that superspreading events are not exceptional events but an expected feature of an expected right-sided tail of the distribution of the basic reproduction number. When this right-hand tail is further skewed, with greater variability than expected leading to an uneven distribution, the term overdispersion is applied (statistical definition). In the context of communicable diseases, overdispersion refers to a non-random pattern of clustering, and which often include a large number of zero cases and a small number of larger outbreaks [8]. This pattern of overdispersion can be applied at the level of an event or index case (individual-level variation) [8], or in the context of networks (population-level via onward transmission chains) [9]: both fall under the broader epidemiological umbrella of heterogeneity. In the context of the former, data

suggest that such superspreading events are characterized by overdispersion in SARS-CoV-2 transmissions, with 10-20% of index cases responsible for 80% of secondary cases [10].

Understanding the nature and characteristics of superspreading events is therefore key to understanding SARS-CoV-2 spread. Agent based modeling by Sneppen et al. [11] suggested that non-repeating, random contacts such as those in restaurants, bars etc. is a dominant contributor to SARS-CoV-2 spread. Chen et al. [12] argue that while social, and micro-environmental factors affect transmissibility, overdispersion could result from an intrinsic characteristic of certain viruses. The knowledge gap that remains is the extent to which each of these factors contribute to the observed distribution of secondary SARS-CoV-2 transmissions per index case, by connecting variability in each of the components with our mechanistic understanding of SARS-CoV-2 transmission.

Emerging data strongly suggest the importance of airborne transmission of SARS-CoV-2 by respiratory aerosols [13, 14, 15, 16, 17]. A large number of small respiratory droplets (when size $< 100\mu m$ at point of exhalation) often referred as aerosols, as suggested by Prather et al. [18], can remain airborne in the liquid or semi-solid state [19], encapsulating the SARS-CoV-2 virus. As a result, the virus can remain infectious within the aerosols for substantial length of time [20]. Several studies have analyzed airborne transmission in specific micro-environments. Bourouiba et al. [21] identified that respiratory droplets and aerosols exhaled during violent expiratory events can travel long distances co-flowing with the moist air jet. Abkarian et al. [22], analyzed exhaled air flow during speech and how certain phonetics produce a train of puffs. Chen et al. [23] showed that for talking and coughing, short range airborne route dominates transmission of respiratory infections. Analyzing a respiratory droplet/aerosol laden cough jet, modeling by Chaudhuri et al. [19] showed that aerosols (droplet-nuclei) of initial size less than $50\mu m$ pose highest infection risk and variation in the corresponding viral load could lead to large variation in the number of secondary infections. Using well-mixed assumptions and the Wells-Riley model [24, 25], Buonanno et al. [26] proposed a quantitative risk assessment for specific micro-environments with asymptomatic infectious cases and suggested that instead of specific superspreaders it is a combination of several factors, including emission and exposure that lead to highly probable, superspreading events. Using similar well-mixed approach Bazant and Bush [27] proposed

to restrict occupancy number and time spent in a room, to mitigate airborne transmission [27]. Schijven et al. estimated risk of infection resulting from sneezing, coughing, speaking, breathing, and singing [28] at different viral loads. Analysis by Bond et al. [29] quantified the importance of confinement in pathogen transport using “effective rebreathed air volume”. Further details on specific aspects of airborne disease transmission: including but not limited to aerosols, flow physics, respiratory droplet size distribution could be found in recent review and opinion articles [30, 31, 32, 33, 34]. Yet, to date mechanistic models describing airborne transmission have not been coupled with real-world distributions and occupancy information towards understanding the large scale features of disease dynamics, for e.g. overdispersion in transmissibility.

The overarching goal of this work is to utilize the mechanistic underpinnings of the airborne disease transmission to explore event-level overdispersion of SARS-Cov-2 spread, utilizing real-world inputs from a large number of social gatherings. The specific goals are:

1. Develop an algorithm based on aerosol dispersion with randomized inputs and available occupancy data to generate distribution of the number of secondary infections per infectious case.
2. Explore if observed patterns of overdispersion in secondary transmissions could be reproduced via simulations using the above algorithm.
3. Derive an analytical function (and not a fit) which can describe the probability density function of number of the secondary infections from the dynamics of the problem.
4. Identify the dominant variables that drive overdispersion and the resulting implications for mitigation measures.

To these ends, an aerosol spread model is solved over a hundred thousand random social-contact settings, utilizing real-world occupancy and area information coupled with realistic input distributions of viral-load and ventilation rates, to obtain the probability distributions of the number of secondary infections per infectious case, in those settings.

2 Method

In this section, first we describe the model to obtain the probability of infection resulting from inhalation of infectious aerosols generated from speaking and breathing, in an indoor, confined, and ventilated micro-environment. Next, we connect this model to an algorithm - a first of its kind to our knowledge, that accepts randomized inputs from distributions of viral load, exhaled aerosol size distribution, ventilation rate, speech and exposure time corresponding to specific inputs of number of people occupying specific indoor areas. The occupancy information is obtained from a large SafeGraph data-set of full-service restaurants from ten major US cities. The restaurants serve as our study setting for time-limited, point source transmission via mostly non-repeating, random contacts, in contrast to households or workplaces where repeated contacts are made with the same individuals and over longer periods of time. In this work, we focus on disease spread by asymptomatic infectious cases (asymptomatic at the time of disease spread therefore including presymptomatics) under the assumption that individuals with symptoms would not be engaged in indoor dining. Hence, we consider only speech and breath as the mechanisms by which respiratory aerosols are ejected into specific micro-environments. The algorithm developed for this work, is shown in Fig. 1.

2.1 Aerosol dispersion from an infectious case

We use the standard, turbulent diffusivity based closure [35] to model the spatio-temporal evolution of ensemble averaged infectious aerosol concentration c_a (number of aerosol particles per unit volume of air associated with emissions from the infector), as shown by Eqn. 1. This approach was used by Drivas et al. [36] to model indoor concentration fields from point sources.

The turbulent diffusivity D_T in Eqn. 1 is a function of the air-change rate a and area of the space given by A . The second term on the R.H.S. is a sink term that accounts for removal of particles by ventilation (a : air change rate) and by deposition, while w_d is the deposition velocity and V being the volume of the indoor space.

$$\frac{\partial c_a}{\partial t} = D_T \left(\frac{\partial^2 c_a}{\partial x^2} + \frac{\partial^2 c_a}{\partial y^2} + \frac{\partial^2 c_a}{\partial z^2} \right) - \left(a + \frac{w_d A}{V} \right) c_a \quad (1)$$

We treat the infectious case as a continuous point source. The initial condition is $c_a(x, y, z, 0) = 0$ with reflection boundary conditions at the six walls. Since the virus remains embedded inside the infectious aerosols and since they are non-volatile, the ratio of the number of virions to the number of aerosol particles in any given volume of air can be assumed to remain constant post-ejection. To retain analytical tractability of the solution, in this work, we consider fast evaporation and a constant, post-evaporation, aerosol volume-averaged w_d ; see [23, 37, 19] for detailed evaporation and deposition considerations of respiratory droplets. Therefore, the ensemble averaged concentration of virus RNA copies per unit volume of air c is proportional to c_a , or

$$c = c_a \mathcal{N}_a \quad (2)$$

The average number of virions within an aerosol particle, \mathcal{N}_a , is the constant of proportionality. Using the identity Eq. 2 we can immediately convert Eqn. 1 into an evolution equation for c , as shown in Eqn. 3:

$$\frac{\partial c}{\partial t} = D_T \left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} + \frac{\partial^2 c}{\partial z^2} \right) - \left(a + \frac{w_d A}{V} \right) c \quad (3)$$

While one can use more complex approaches to model turbulent mixing and aerosol dispersion, the advantage of the relatively simple, yet sufficiently robust and accurate, diffusivity based closure is the analytical tractability and the inexpensive solution it offers. As will be seen later, such an attribute is pivotal to generating the large number of realizations of the c field, utilized in this work. As such, for a continuous point source Q_{x_0, y_0, z_0} at (x_0, y_0, z_0) , the solution of the concentration field $c(x, y, z, t')$ at time t' was given by [36]:

$$c(x, y, z, t') = \int_0^{t'} \frac{Q_{x_0, y_0, z_0} e^{-\left(a + \frac{w_d A}{V}\right)t}}{(4\pi D_T t)^{3/2}} R_x R_y R_z \psi(t) dt \quad (4)$$

with wall reflection terms for $i, j, k \neq 0$

$$R_x = \sum_{i=-\infty}^{\infty} \left[e^{-\frac{(x+2iL-x_0)^2}{4D_T t}} + e^{-\frac{(x+2iL+x_0)^2}{4D_T t}} \right] \quad (5)$$

$$R_y = \sum_{j=-\infty}^{\infty} \left[e^{-\frac{(y+2jW-y_0)^2}{4D_T t}} + e^{-\frac{(y+2jW+y_0)^2}{4D_T t}} \right] \quad (6)$$

$$R_z = \sum_{k=-\infty}^{\infty} \left[e^{-\frac{(z+2kH-z_0)^2}{4D_T t}} + e^{-\frac{(z+2kH+z_0)^2}{4D_T t}} \right] \quad (7)$$

L, W, H are the length, width, and height of the room, respectively. Here, we have introduced a virus survivability function $\psi(t)$ in the solution, as in Chaudhuri et al. [19]

$$\psi(t) = (1/2)^{t/t_{\frac{1}{2}}} \quad (8)$$

$$t_{\frac{1}{2}} = f(T, RH, UV_{index}) \quad (9)$$

Here T is the temperature, RH is the relative-humidity and UV_{index} is the ultra-violet index inside the room of interest. Based on the experiments by Dabisch et al. [20] and calculator from DHS [38], for $T = 21.7^\circ C$, $RH = 0.50$, $UV_{index} = 0$ i.e. typical ASHRAE recommended indoor air conditions, for SARS-CoV-2, half-life $t_{\frac{1}{2}} = 32.07$ minutes. The SARS-CoV-2 virus half-life reduces monotonically with temperature and non-monotonically with relative humidity [39] like other enveloped viruses, as shown by Marr et al. [40].

The above solution given by Eqns. 4-7 (without $\psi(t)$) was validated by Cheng et al. [41] who released CO from a point source and measured its spatio-temporal dispersion characteristics inside typical built environments. Indeed aerosols deposit (accounted for in this paper) unlike CO, but since their motion is predominantly controlled by turbulent diffusion inside a room (turbulent diffusivity $D_T \gg$ molecular diffusivity of CO or effective diffusivity of aerosol particles), the validation exercise is highly relevant. They suggested the following correlation between turbulent diffusivity, area of room, and air change rate:

$$D_T = A \times (0.52 \times ACH + 0.32)/3600 \quad (10)$$

which is used for the present study as well. Recall that A is the area, ACH is the air changes per hour, therefore $a = ACH/3600$.

Denoting τ as the time duration of the event under considerations, the ensemble averaged number of infectious virions, generated from both speech and breath of the infected individual located at x_0, y_0, z_0 , that is inhaled at x, y, z_0 upto time τ is denoted $\mathcal{N}_v(x, y, z_0, \tau)$ and it is given by:

$$\mathcal{N}_v(x, y, z_0, \tau) = \int_0^\tau c(x, y, z_0, t') \dot{V}_b dt' \quad (11)$$

While infectious aerosol emissions from breathing take place over the entire time duration of the event, emissions from talking are assumed to occur only for a time $t_s < \tau$ (see supplementary material, section 3 for details). This is accounted for in our calculation of the corresponding time-varying virus concentration field. \dot{V}_b is the average volume of air inhaled per second. The local probability of infection due to infectious aerosols produced by speaking and breathing $\mathcal{P}_{s+b}(x, y, z_0, \tau)$ is calculated using the dose-response model originally proposed by Haas [42]. The dose-response constant is chosen as $r_v = 1/1440$ based on the estimations by Haas [43] and Schijven et al. [28] for the original SARS-CoV-2 variant. The subscript $s + b$ indicates infection caused by exhaled aerosols due to speech (s) and breath (b) from the infected individual.

$$\mathcal{P}_{s+b}(x, y, z_0, \tau) = 1 - e^{-r_v \mathcal{N}_v(x, y, z_0, \tau)} \quad (12)$$

The number of secondary infections Z within a room is thus given by

$$Z = \int_0^L \int_0^W \rho_p \mathcal{P}_{s+b}(x, y, z_0, \tau) dx dy \quad (13)$$

Here ρ_p is the susceptible population density (assumed to be uniform), estimated as $\rho_p = n/A$ at the given point of interest (POI). POI is a term used in SafeGraph data-set and in Chang et al. [44] referring to a specific business location like a restaurant. A is the indoor area of that POI. In this paper, we place one infectious case at each POI. Therefore, Z is a measure of individual-level infectivity as well. Here n is the number of susceptible individuals present at a given POI i.e. $n = n_p - 1$ where n_p is the total number of people present at that POI. N is the sample space variable corresponding to n .

2.2 Algorithm for generating the pdf of secondary infections

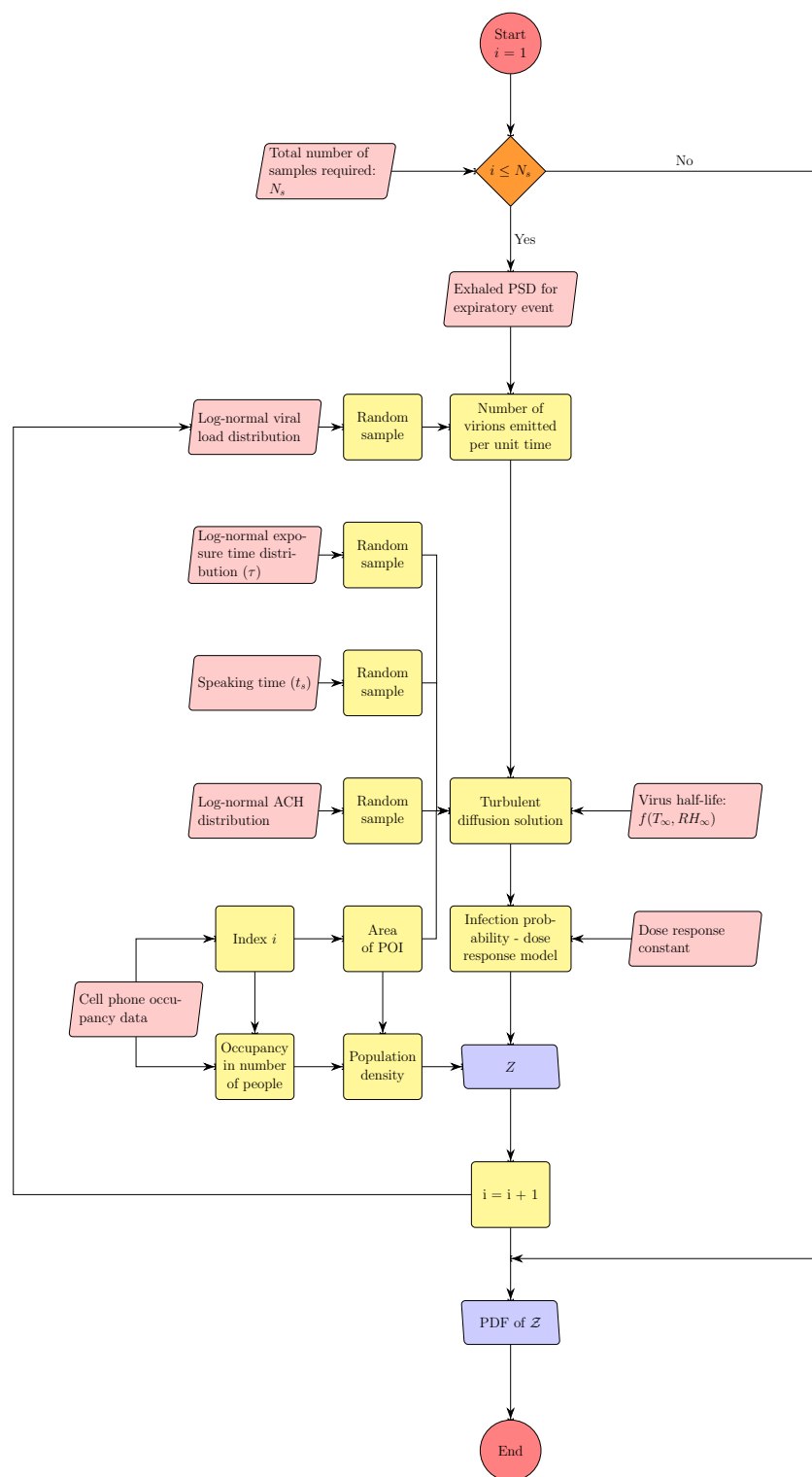


Figure 1: The algorithm for estimating distribution of secondary infections Z .

With a basic semi-analytical framework available to compute the number of secondary infections inside a room of specified dimensions, for a given viral load, ACH , and occupancy, we use the algorithm detailed in Fig. 1 with randomized inputs from specified distributions (e.g., for viral load, air change rate) and available data (e.g., for occupancy, room area) to generate the pdf of \mathcal{Z} .

The algorithm runs over N_s iterations where $N_s = 103679$ corresponding to the total number of full service restaurants in 10 US cities, available from SafeGraph data, used for generating a pdf at the end; i being the iteration index. First, the volume of the mucosalivary fluid ejected per unit time, evaluated from the exhaled respiratory droplet size volume size distribution (vsd) multiplied with a random sample from the log-normal viral load distribution, is fed into the code to generate the number of virions emitted per unit time. This is used in the turbulent diffusion solutions shown in Eqns. 4-7. Random samples of exposure time, speaking time, ACH , and area of POI, along with virus half-life, which is calculated separately, also act as input parameters for the solver. The resulting concentration field is used in the dose response model (with a specified dose response constant) to obtain probability fields. Next, probability fields are multiplied with the POI specific population density data (from SafeGraph) to obtain Z values. Once the index value reaches N_s , the iteration ends and the pdf of \mathcal{Z} is calculated using Z values from every iteration. Note that the random variable - number of secondary infections at specific POIs or individual infectivity is denoted by Z , while the sample-space variable corresponding to Z is denoted by \mathcal{Z} . Pdf of \mathcal{Z} i.e. probability density function, defined as probability per unit distance in the sample space of \mathcal{Z} [35], is a powerful tool to describe overdispersion, irrespective of the specific number of events or samples. Furthermore pdfs are amenable to analytical descriptions. Fundamentally, any long tailed pdf represents an inherently overdispersed random variable, because the long tail represent finite (but could be small) probability of an event where $Z \gg \text{mean of } Z$, while any pdf other than that represented by a delta function represents heterogeneity. In view of these, in this paper, the analysis of overdispersion will be addressed using both simulated and analytical pdfs of \mathcal{Z} .

Three key inputs - particle size distribution, viral load distribution, and occupancy information are described below, details of the rest of the inputs: ventilation rate distribution,

speech time and exposure time distributions can be found in the supplementary material (section 3). A table summarizing the important parameter values, distributions, and sources can also be found there (Table 1).

2.2.1 Particle size distribution of exhaled aerosols

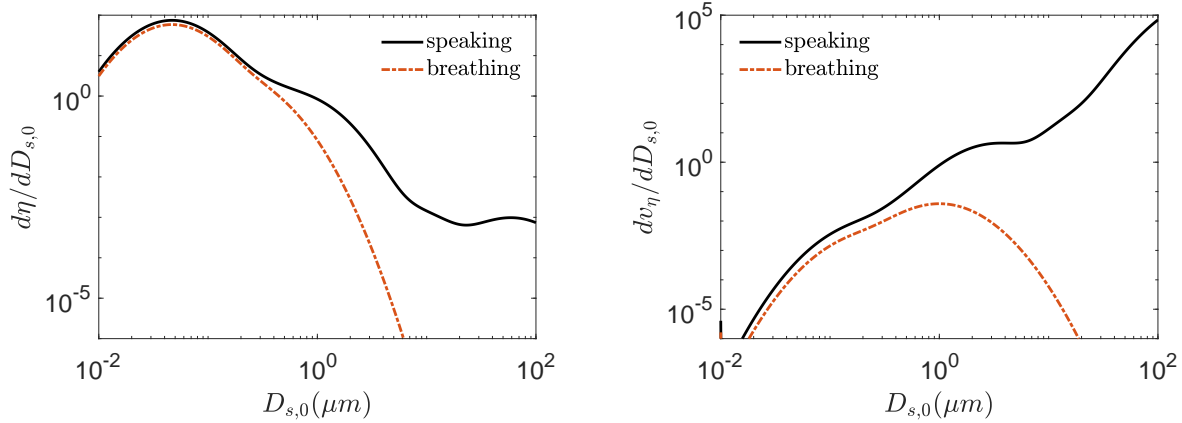


Figure 2: Aerosol number size distribution (nsd) on left, and volume size distribution (vsd) on right, for speaking and breathing as expressed by Eq. 14 and Eq. 4 respectively, from [33] as a function of initial particle diameter sample space variable $D_{s,0}$ i.e. at the moment of exhalation.

The particle size distributions at the source of the expiratory events: speaking and breathing are obtained from the review by Pöhlker et al. [33]. A multimode lognormal fitting has been found to describe the corresponding distribution reasonably well. Number size distribution (nsd) for exhaled aerosol particles for different expiratory events can be described using a single function with event specific constants. The general form for the distribution can be expressed as

$$\frac{d\eta}{dD_{s,0}} = \log_{10}(e) \sum_{i=1}^n \frac{A_i}{D_{s,0}} \exp \left(- \left[\frac{\ln(D_{s,0}/D_i)}{\sigma_i} \right]^2 \right) \quad (14)$$

where η is the number concentration of particles (whereas q_η is the volume concentration of particles), D_0 is the particle diameter at ejection i.e. at time $t = 0$, and $D_{s,0}$ is the corresponding sample space variable. Moreover, A_i and σ_i are constants that depend on the mode and type of expiratory event. For further details the reader is referred to the

supplementary material (section 1). Figure 2 shows the number and volume size distributions that are used in the simulation for the speaking and breathing events. The volume size distribution (vsd) is integrated over $D_{s,0}$ to obtain the volume of the exhaled respiratory liquid per unit volume of air for the given expiratory event. This volume flow rate ratio multiplied with the individual viral load ρ and the volume flow of air per unit time, yields the ejected number of virions (RNA copies) emitted per unit time i.e. the source term Q_{x_0,y_0,z_0} of Eqn. 4. Here we choose particles only with $D_{s,0} < 100\mu m$ as the larger ones will settle in less than 10 s even after accounting for evaporation [19]. A detailed description of the particle deposition velocity calculation is presented in the supplementary material (section 2).

2.2.2 Viral load distribution

Viral load, (ρ) has been associated with infectivity and thus, the number of secondary infections for a given setting [45, 46]. Analyzing respiratory droplet and aerosol laden cough jets Chaudhuri et al. [19] showed that the corresponding number of infected individuals could vary by orders of magnitude due to variation in the viral load. Chen et al. [10] analyzed a large number of SARS-CoV-2 viral load databases and suggested that viral load is an important contributor to heterogeneity in secondary infections. They also showed that the viral load distributions were similar for symptomatic and asymptomatic stages of infection. This point was further amplified by direct measurements of Yang et al. [47] who showed that viral load distributions are nearly identical among hospitalized (symptomatic) and asymptomatic population. In this paper, we utilize measurements of viral load in asymptomatic (including presymptomatic) population from Yang et al. [47] as an input into the algorithm shown in Fig. 1. According to [47] at the time of saliva collection, the infected individual was either asymptomatic or presymptomatic. The pdf of $\ln(\rho_v)$ (ρ_v is the sample space variable of ρ) is shown in Fig. 3. We generate and use samples of ρ from this log-normal distribution in our calculations.

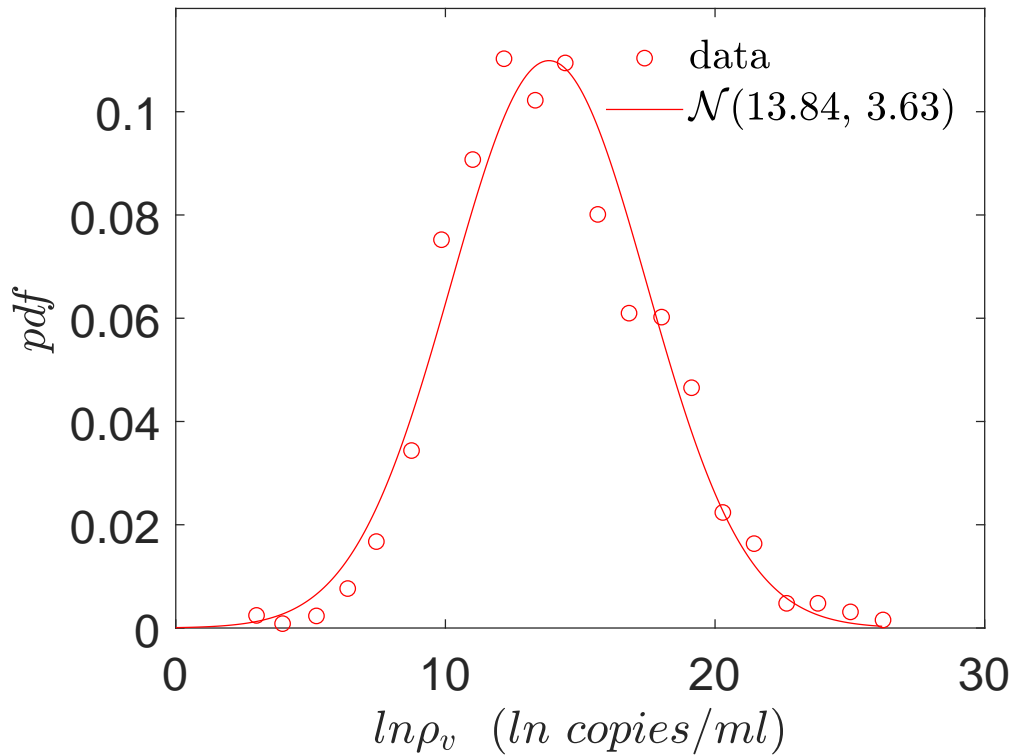


Figure 3: Viral load pdf for asymptomatic (including presymptomatic) population, based on the histogram data from Yang et al. [47]. The red line shows a normal distribution with μ and σ given by 13.84 and 3.63 respectively.

2.2.3 Occupancy and area of different points of interest from SafeGraph data

Two of the important inputs in the simulation are the areas of different points of interests (POI) and the number of people occupying them during each time period of interest. These data were obtained from SafeGraph - a company that collects anonymous data from mobile devices. For our simulation, data that is available from individual confined spaces, rather than that from a collection of several indoor spaces, was felt to be most appropriate. Hence, we used SafeGraph data for areas of full service restaurants (POI) over ten cities in USA, namely Atlanta, Chicago, Dallas, Houston, Los Angeles, Miami, New York City, Philadelphia, San Francisco, and Washington D.C. The SafeGraph-tabulated area for each restaurant is multiplied by 0.5 to convert the total area of a given restaurant to the corresponding sit-

ting area since the dining area is estimated as 50 % of the total restaurant area (based on typical restaurant design guides [48]; change of this factor does not qualitatively change the observations). The occupancy information in these POIs between hours 12:00 - 13:00 and 18:00 - 19:00 over seven days starting from March 1, 2020 is obtained from the data sets created by Chang et al. [44] who developed a mobility network based SEIR model using the SafeGraph data. These two time periods represent typical lunch and dinner times, and hence highest occupancy periods of any day. The pdf of the averaged number of susceptible individuals in restaurants (total occupancy minus one) between hours 12:00 - 13:00 and 18:00 - 19:00 over ten US cities is shown in Fig. 4. Also, shown in the figure is an exponential fit and the corresponding fitting parameter.

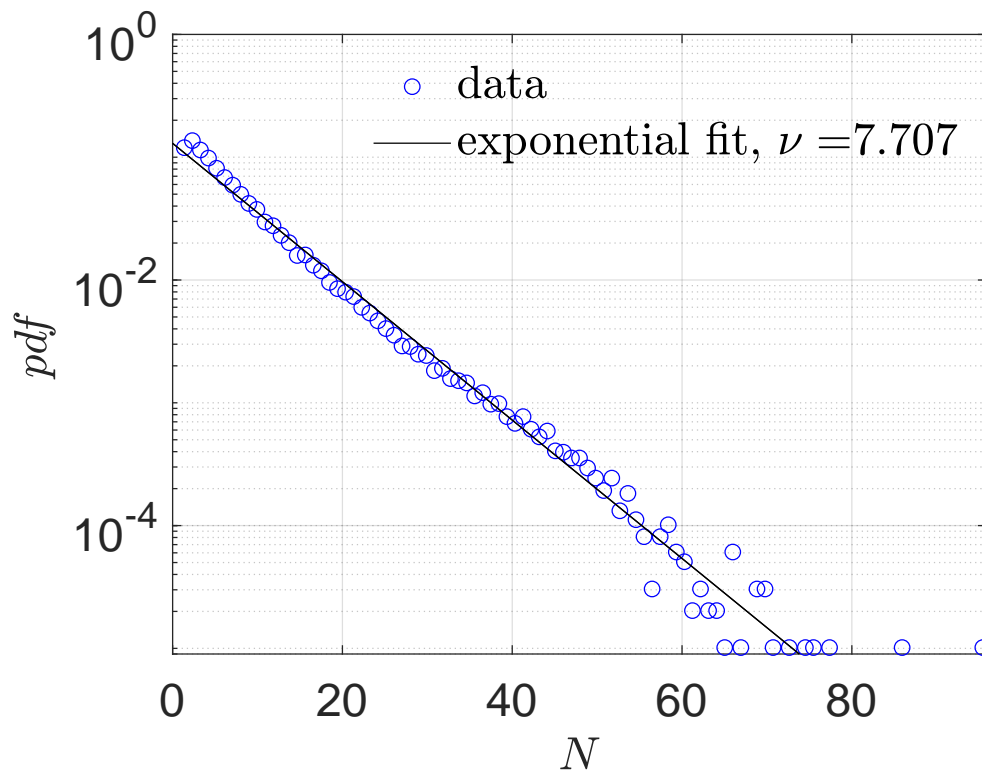


Figure 4: Pdf of number of susceptible individuals in full service restaurants at ten US cities, from SafeGraph data.

3 Results and discussion

3.1 Spatial distributions of particle concentration and infection probability for fixed conditions

First, we present the spatial distribution of the virus concentration (RNA copies/ m^3 of air) with the source at $x = 2.5m$ and $y = 2.5m$ from the origin (at the bottom left corner) of a $10m \times 10m \times 3m$ room after 15 minutes of aerosol exhalation by speech and breath. In Fig. 5 the first row (a-c) presents results with a constant viral load $\rho = 10^9$ copies/ml, but with increasing ACH . Note that the third column represents a case without wall reflections and hence can simulate outdoor conditions. The second row of Fig. 5 (d-f) represents a constant but five times higher viral load of $\rho = 5 \times 10^9$ copies/ml. The results demonstrate strong inhomogeneity of virus concentration and also show that contours scale linearly with ρ for the same ACH due the linear nature of the governing Eqn. 3. Using the dose response model (Eqn. 12) the corresponding contours of probability of infection are shown in Fig. 6. Indeed near unity \mathcal{P}_{s+b} are found near the source, and there is decay with distance from the source. The \mathcal{P}_{s+b} contours do not scale linearly with ρ due to the nonlinear nature of the dose response model. Increasing ACH invariably reduces virus concentration for a given ρ . However, the reduction in the probability of infection may not be proportional to the reduction of the virus concentration due to the non-linearity involved in the dose response. Interestingly, the simulated outdoor conditions (c, f) show much smaller infection probability both near and far from the source w.r.t. the confined cases. This can be attributed to the inherently higher ACH inside the volume of interest, but primarily due to absence of confinement which allows the virus concentration to freely decay with space.

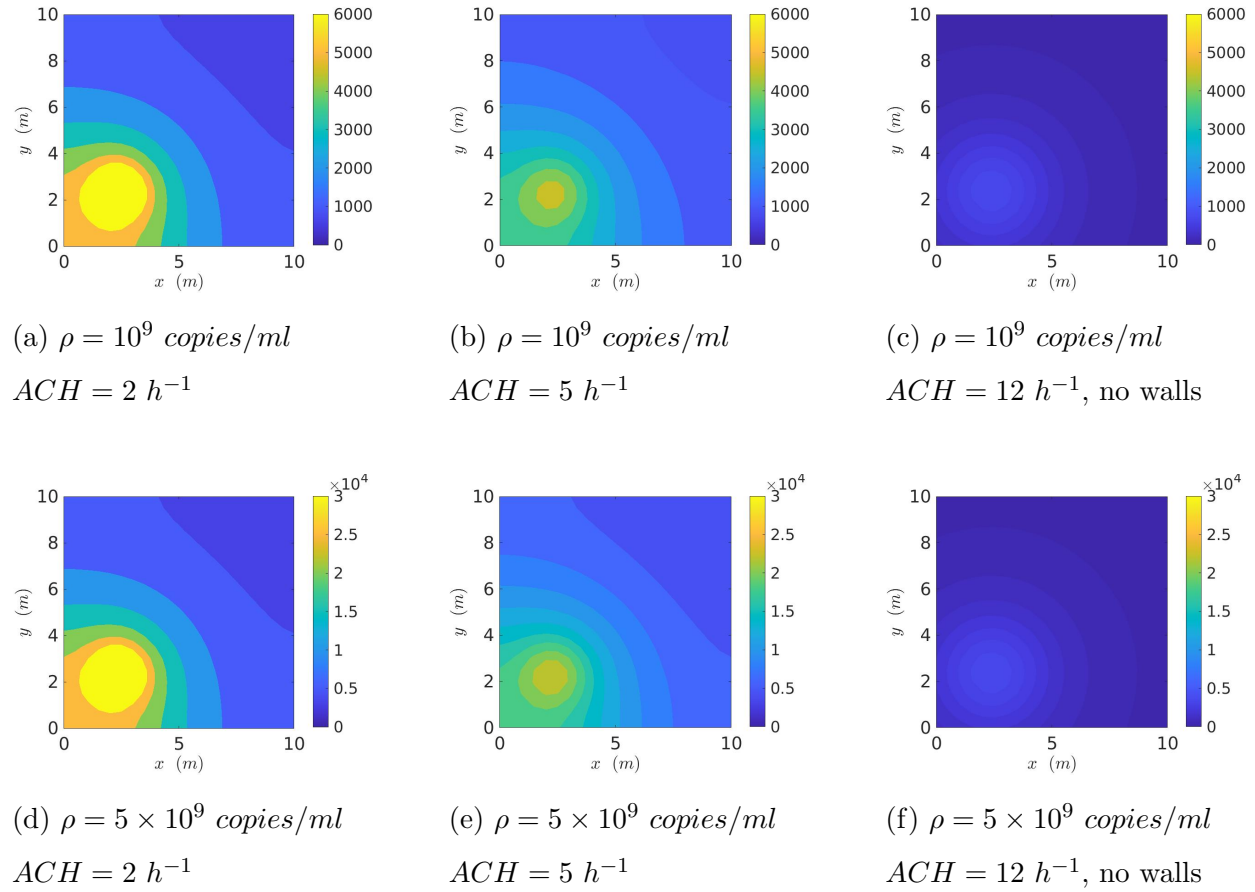


Figure 5: Contour plots of spatially resolved virus concentration (RNA copies/ m^3) at time $t = 15$ minutes from start of expiration event (source located at $x = 2.5m$, $y = 2.5m$).

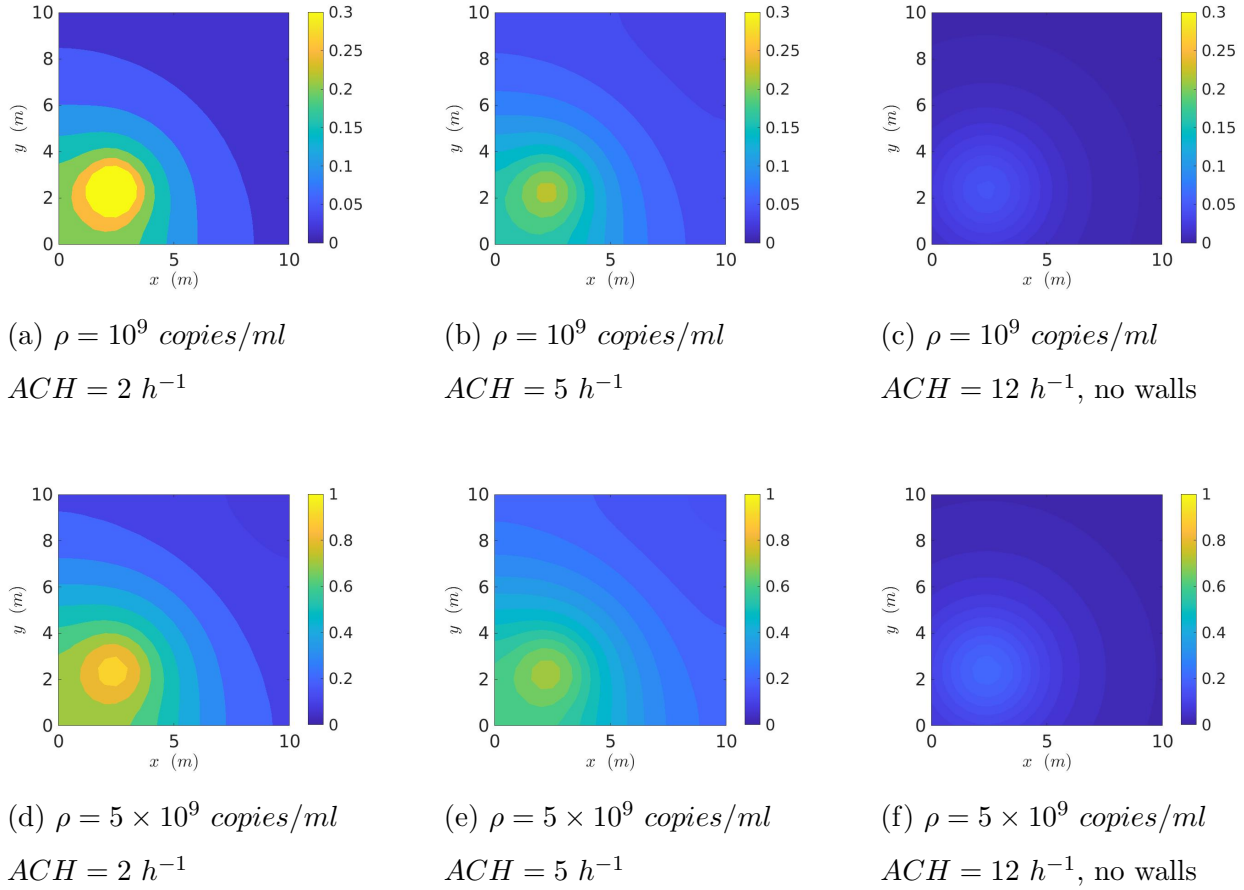


Figure 6: Contour plots of spatially resolved probability of infection $\mathcal{P}_{s+b}(x, y)$ at time $t = 15$ minutes from start of expiration event (source located at $x = 2.5\text{m}$, $y = 2.5\text{m}$).

3.2 Statistical distributions of secondary infections generated from realistic input distributions

A simulation based on the algorithm presented in Fig. 1 is run for each data point available from the predefined SafeGraph dataset, resulting in a sample size of $N_s = 103679$. We place one infected individual at each POI, at random locations within its premises. As such, most inputs, including viral load, ACH , exposure time, speaking time, source location (x_0, y_0) , are randomized. ACH is generated from a log-normal distribution with a median $ACH = 2.16 \text{ hr}^{-1}$, such that $\mu_{ACH} = 0.7701$, $\sigma_{ACH} = 0.7554$ based on measurements by Bohanon et al. [49]. Room height $H = 3 \text{ m}$, height of source (seated) $z_0 = 1 \text{ m}$, indoor conditions ($T = 21.7^\circ\text{C}$, $RH = 0.50$, $UV_{index} = 0$), and dose response constant $r_v = 1/1440$ are held

constant. The resulting distribution of number of secondary infections - pdf of \mathcal{Z} is presented in Fig. 7. An analytical solution $g(\mathcal{Z})$ to be derived in sub-section 3.3 is also shown in the figure.

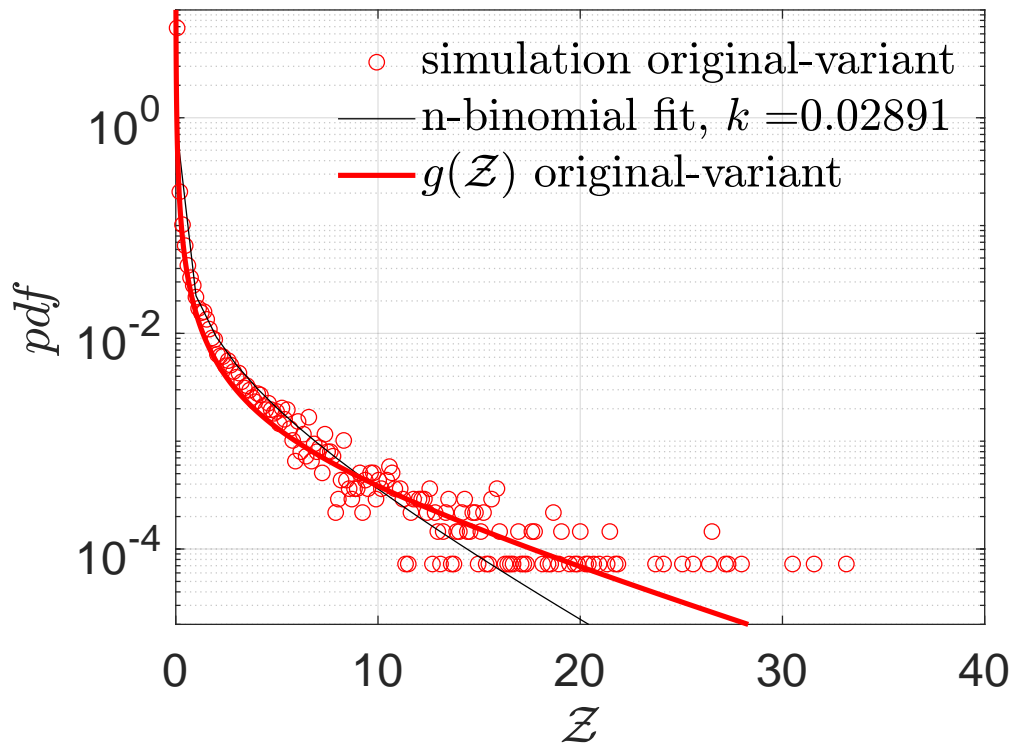


Figure 7: Pdf of the number of secondary infections: \mathcal{Z} and negative binomial fit. The analytical function g is given by Eqn. 22. $g(\mu = 13.84, \sigma = 3.63, \alpha = 1.123 \times 10^{-10}, \nu = 7.71)$.

The stretched tailed nature of the simulated pdf is immediately apparent. This shows that there is small but finite probability of tens of secondary infections per infected individual. For this simulation the $mean(\mathcal{Z})_s = \langle \mathcal{Z} \rangle_s = 0.14$ indicating that over this one hour, on average less than 1 person got infected per infector. The calculated total number of infections is 14482, with many of the infections occurring in large superspreading events. As such, only 3.57 % of the infected individuals infected 80 % of the population over this time. This could also be the reason why it is generally difficult to culture the virus from the air, though that was unequivocally demonstrated by Lednický et al. [50]. High probability of

infection, which as shown in the paper typically occurs at high viral load, could be correlated with high probability of virus detection in the air. Direct virus detection from air could therefore necessitate sampling from a large population of infected individuals. Clearly, the finite probability of $Z \gg \langle Z \rangle$ recovers the inherently overdispersed nature of SARS-CoV-2 transmission dynamics. Fitting a negative binomial probability distribution function to the Z -pdf yields a good fit, with a dispersion parameter $k = 0.029$. While the fit quality worsens at the pdf tails, the dispersion parameter is in the same order as the corresponding values for SARS and measles estimated by Lloyd-Smith et al. [1]. However, it is to be noted that we are considering only infections over a period of about 1 hour on average, as opposed to the entire course of infection, hence $\langle Z \rangle$ should not be interpreted as \mathcal{R}_0 . Similarly, the qualitative k value thus obtained should be interpreted with care.

Figure 8 shows the joint pdf of ρ and Z . The close correlation of the two random

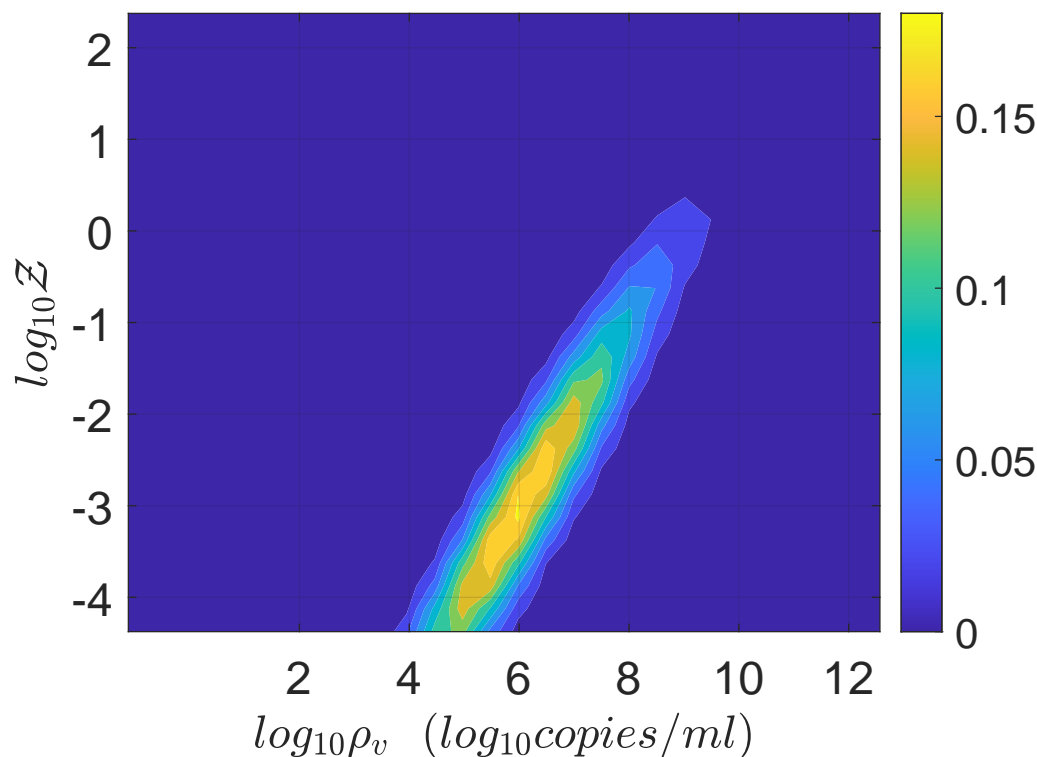


Figure 8: Joint pdf of viral load and number of secondary infections.

variables across nearly six orders of magnitude is immediately apparent from this figure.

While correlation may not imply causality in general, it is reasonable in this case that the extreme variation in ρ is causing a similar variation in Z , with other parameters controlling the slope and strength of the correlation.

We present the jpdf of ACH and Z in Fig. 9. It is apparent that the highest number of infections Z occur at lower air exchange rates, as expected; however, the majority occur at an intermediate air exchange rate of about 1-1.5 ACH , in part because very low and very high air exchange rates are simply less common. Furthermore, it is also clear that dispersion of Z and ACH are indeed negatively correlated. The effect of universal, high ventilation rates, and masks will be taken up later.

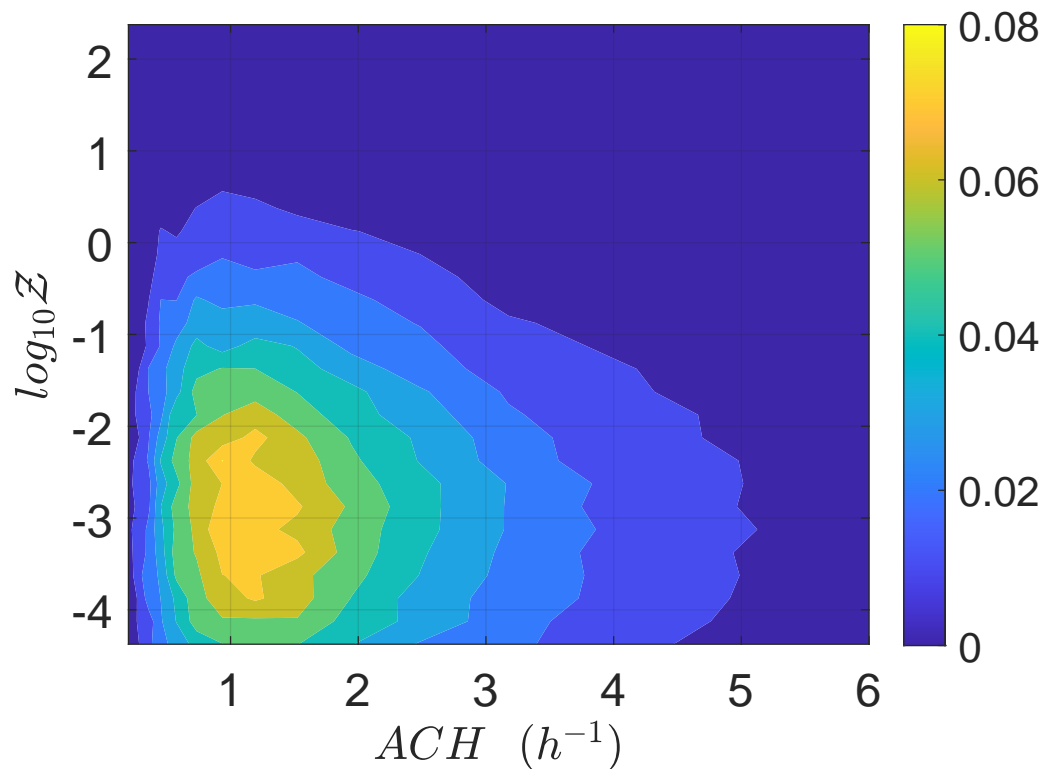


Figure 9: Joint pdf of ACH and Z .

3.3 Derivation of the analytical pdf of \mathcal{Z} : $g(\mathcal{Z})$

The viral load pdf from Yang et al. [47] can be well approximated by a log-normal distribution with parameters μ and σ given by $\mu = 13.84$ and $\sigma = 3.63$. This is shown in Yang et al. and also shown in Fig. 3. As such, we used the following pdf of viral load $f(\rho_v)$ for generating inputs into the simulation

$$f(\rho_v) = \frac{1}{\rho_v \sigma \sqrt{2\pi}} e^{-(\ln(\rho_v) - \mu)^2 / 2\sigma^2} \quad (15)$$

Due to its extremely large (over $\mathcal{O}(12)$) variation, our analysis shows that viral load ρ is one of the most dominant variables in controlling overdispersion of secondary infections \mathcal{Z} , as is apparent from Fig. 8. This result can also be presented in terms of secondary attack rate - a more generalized descriptor, defined as $\tilde{Z} = Z/n$ where n is the number of susceptible individuals present at the given POI over the period of interest τ . \tilde{Z} is the sample space variable corresponding to \tilde{Z} . Variation of \tilde{Z} w.r.t. ρ is shown in Fig. 10. Clearly, this plot reflects the dose-response function (Eqn. 12) on a macro scale given the dominant influence of ρ in controlling probability of infection and eventually secondary attack rates. Therefore, we propose a function similar to Eqn. 12 to model the response of \tilde{Z} to ρ variation. This is shown below:

$$\tilde{Z} = 1 - e^{-\alpha\rho} \quad (16)$$

We can also write

$$\rho = -\frac{1}{\alpha} \ln(1 - \tilde{Z}) \quad (17)$$

The constant α can be estimated as the inverse of the average number of virions inhaled per unit volume of mucosalivary liquid ejected that is required for infection probability of $1 - e^{-1} = 0.63$. For the average room volume $\langle V \rangle = 6.34 \times 10^2 \text{ m}^3$, average speaking time $\langle t_s \rangle = 1469 \text{ s}$, average exposure time $\langle \tau \rangle = 3914 \text{ s}$, average air change rate, $\langle ACH \rangle = 2.87 \text{ hr}^{-1}$, and deposition parameter based on average room area and volume $\beta_0 = 0.002 \text{ s}^{-1}$, α can be estimated as

$$\alpha = \frac{r_v \dot{Q}_l \langle t_s \rangle \dot{V}_b}{\langle V \rangle} \int_0^{\langle \tau \rangle} \psi(t) e^{-(\langle ACH \rangle / 3600 + \beta_0)t} dt \quad (18)$$

$\psi(t)$ is given by Eqn. 9, and $\dot{Q}_l = 2.223 \times 10^{-6} \text{ ml/s}$. Note that all the $\langle \rangle$ quantities mention averages over the distributions used in the present simulation. Equation 18 yields

$\alpha = 1.123 \times 10^{-10}$ ml/copies. The comparison of Eq. 16 and the simulation results are shown in Fig. 10.

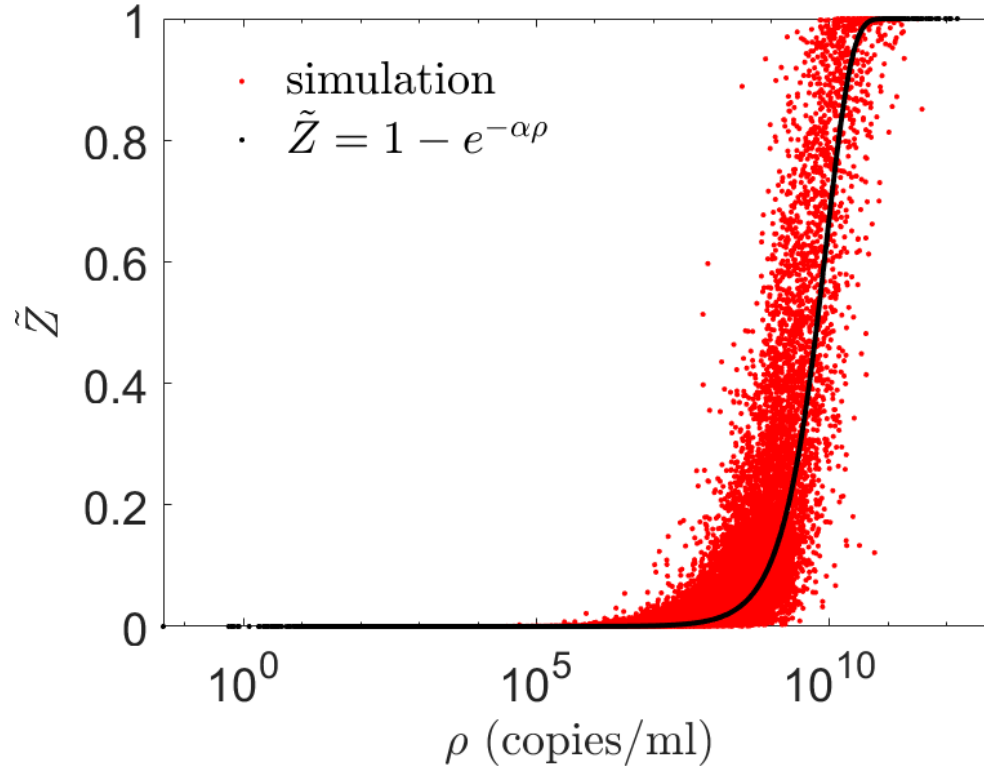


Figure 10: Scatter plot of viral load (ρ) vs secondary attack rate (\tilde{Z}). The black dot curve shows Eqn. 16 with $\alpha = 1.123 \times 10^{-10}$ ml/copies.

With the functional form of the pdf of the viral load known (given by Eqn. 15) we can immediately substitute Eqn. 17 into Eqn. 15 to eventually derive the pdf of \tilde{Z} using the generalized equation below.

$$\phi(\tilde{Z}) = f\left(-\frac{1}{\alpha}\ln(1 - \tilde{Z})\right) \frac{d\left(-\frac{1}{\alpha}\ln(1 - \tilde{Z})\right)}{d\tilde{Z}} \quad (19)$$

Using the log-normal form of f we derive the analytical function below which could be used to model the pdf of secondary attack rate \tilde{Z} : $\phi(\tilde{Z})$. However, the same method should be applicable to other functional forms of f , like a Weibull distribution as in [10] instead of

log-normal.

$$\phi(\tilde{Z}) = \frac{1}{\left(-|1 - \tilde{Z}| \ln(1 - \tilde{Z})\right) \sigma \sqrt{2\pi}} e^{-\left\{\ln\left(-\frac{1}{\alpha} \ln(1 - \tilde{Z})\right) - \mu\right\}^2 / 2\sigma^2} \quad (20)$$

Comparison of Eqn. 20 with the simulation results is shown in Fig. 11.

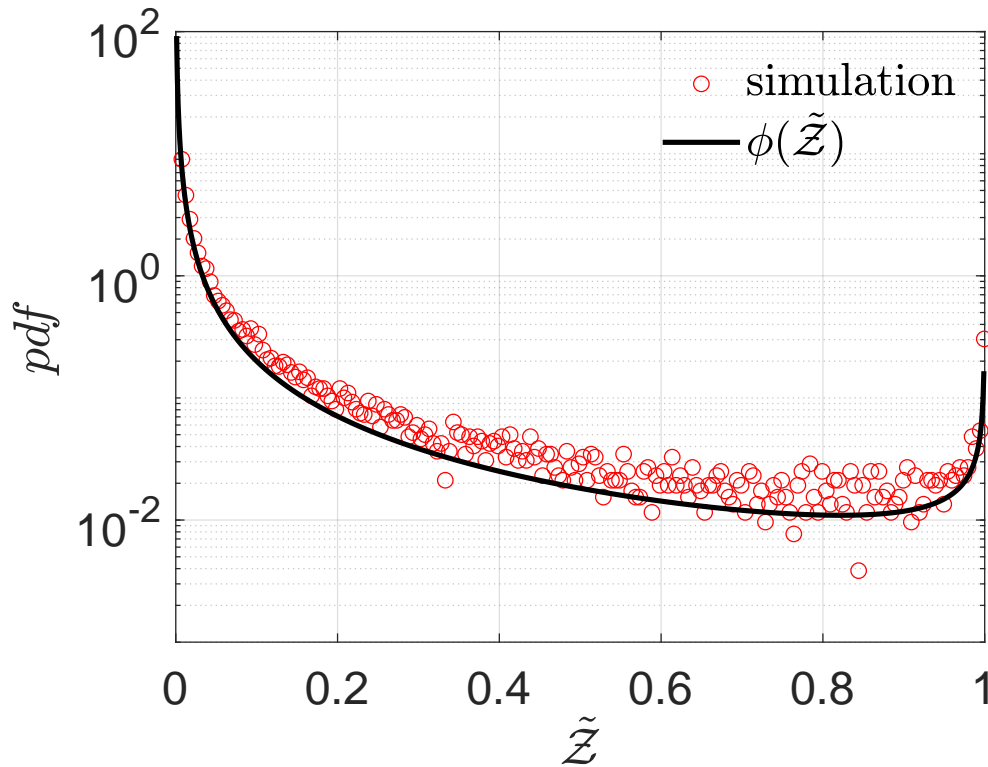


Figure 11: Pdf of \tilde{Z} and its comparison with the analytical result given by Eqn. 20

It is evident that Eqn. 20 describes the simulation data to good quantitative accuracy. It is also remarkable that very important effects of area, ACH , virus kinetics, exposure and speaking times could be encapsulated within one constant α . It is to be recognized that the equation is valid only for $\tilde{Z} < 1$. This is an inherent feature emanating from the derivative of the functional form of the dose response model which yields the $1 - \tilde{Z}$ term in the denominator. Importantly, the equation can describe the range $0 \leq \tilde{Z} < 1$ with high degree of veracity. Now, we can write $Z = N(1 - e^{-\alpha\rho})$ using Eqn. 16. For a fixed α , clearly N and $1 - e^{-\alpha\rho}$ are independent random variables. Therefore, for any given pdf of N given

by $h(N)$, pdf of \mathcal{Z}/N given by $\phi(\mathcal{Z}/N)$ and a known α , and using the general equation that describe the pdf of the product of two independent random variables, we can write the pdf of \mathcal{Z} given by

$$g(\mathcal{Z}) = \int_0^\infty h(N) \phi(\mathcal{Z}/N) \frac{1}{N} dN. \quad (21)$$

Using the exponential distribution $h(N, \nu) = \frac{1}{\nu} e^{-\frac{N}{\nu}}$ shown in Fig. 4 we derive for $\mathcal{Z} < N$

$$g(\mathcal{Z}) = \int_0^\infty \frac{e^{-\frac{N}{\nu}}}{-\nu\sigma\sqrt{2\pi}(N-\mathcal{Z})\ln(1-\mathcal{Z}/N)} e^{-\left\{\ln\left(-\frac{1}{\alpha}\ln(1-\mathcal{Z}/N)\right)-\mu\right\}^2/2\sigma^2} dN \quad (22)$$

It is to be noted that the constants μ, σ are properties of the viral load distribution, ν is the constant of the occupancy distribution, and α encapsulates $\tau, t_s, ACH, A, V, t_{\frac{1}{2}}$ etc. according to Eqn. 18. It is apparent that the pdf $g(\mathcal{Z})$ is stretched to higher (lower) \mathcal{Z} values when μ, σ, ν or α increases (decreases). The remarkable match between this analytical function: g - the analytical pdf of the number of secondary infections and that obtained from the simulation data has been shown in Fig. 7. We revisit it here for further discussion. It is evident that the stretched tail of the pdf of \mathcal{Z} results from the two stretched exponential functionals of \mathcal{Z} and N . This can be verified by noting that when either $\sigma \rightarrow 0$ or $\nu \rightarrow 0$ the overdispersion of Z vanishes. The first stretched exponential arising from the lognormal distribution of viral load ρ_v and the latter from the exponential distribution of number of people at the different POIs. Therefore, it is the joint contribution of overdispersed viral load and overdispersed occupancy that results in overdispersion of secondary infection numbers causing superspreading events. This is shown here with a single equation. The $\langle Z \rangle$ obtained from the analytical pdf, given by $\langle Z \rangle = \int_0^\infty \mathcal{Z} g(\mathcal{Z}) d\mathcal{Z}$. We find mean and standard deviation as $\langle Z \rangle = 0.13$, $std(Z) = 0.99$, respectively, in comparison to $\langle Z \rangle_s = 0.14$ and $std(Z)_s = 0.87$ from the simulations. The analytical pdf $g(\mathcal{Z})$ is expected to be a generalized result and could be applied for any large number of indoor POIs, without much restriction on their type.

3.4 Variants and mitigation measures

Finally, we test whether the derived Eqn. 22 can describe overdispersed transmission associated with a different viral load distribution equally well. To this end, we generate a viral load

distribution with a mean which is 1000 times the mean of the original distribution presented in Fig. 3. Thus, while Fig. 3 represents the pdf of the original variant, the new log-normal distribution characterized $\mu = 20.67$ and $\sigma = 3.63$ might represent the pdf of viral load of the δ -variant infected individuals. Early data from Li et al. [51] suggest that the viral loads of the δ -variant could be as high as “1000 times greater compared to A/B lineage infections during initial epidemic wave in China in early 2020”. The simulation results in terms of \mathcal{Z} -pdf are presented in Fig. 12. The greater transmissibility of the δ -variant under the assumption of 1000-fold higher mean viral load is immediately apparent. In comparison to the $\langle Z \rangle = 0.13$ for the original variant, the $\langle Z \rangle = 2.64$ for δ . Therefore, just based on viral load, according to the calculations and model, δ -variant could be nearly 20 times higher transmissible on average w.r.t. the original variant, over about an hour of contact. Interestingly, with just increased μ , Eqn. 22 can capture the pdf of \mathcal{Z} for the δ -variant, remarkably well, alongside the one for the original variant. However, it is to be recognized that there is some uncertainty on the assumption that mean viral load of the δ -variant is 1000 times higher than the original variant. Data also suggest that viral load and infectiousness potential (using a proxy of culture-positivity for example), while monotonic in nature, may not demonstrate a linear relationship [52]. Furthermore, $\langle Z \rangle \neq \mathcal{R}_0$. Hence, the enhancement factor thus found is valid only within the context of the assumptions and scope of this work.

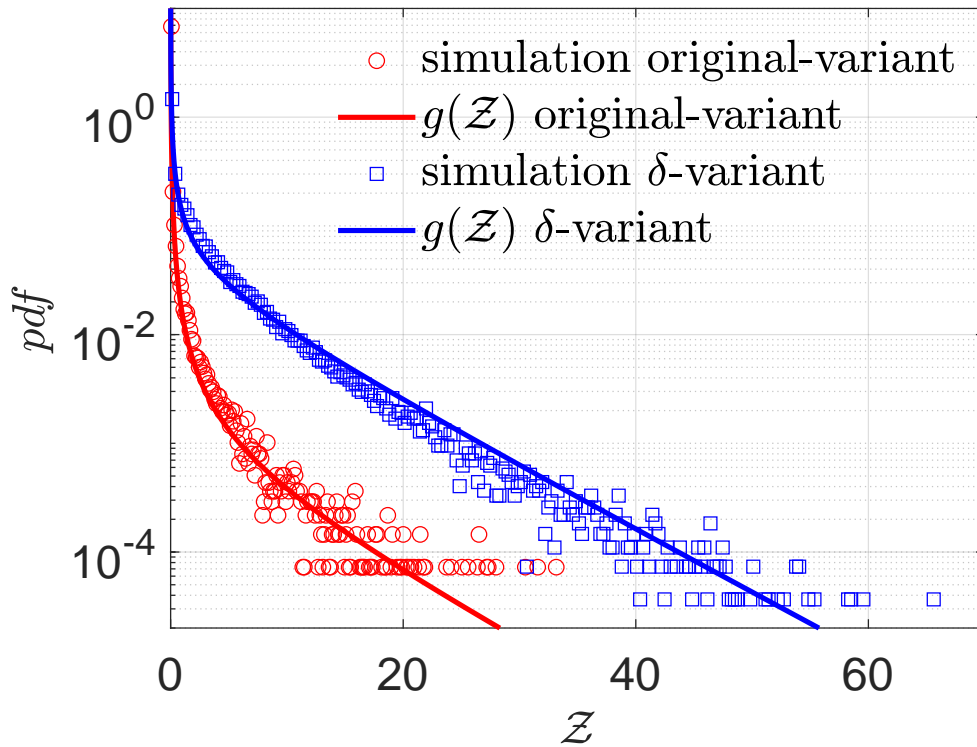


Figure 12: Pdf of Z and its comparison with the analytical result given by Eqn. 22

for original variant (red circle symbols, red bold line)

$g(\mu = 13.84, \sigma = 3.63, \alpha = 1.123 \times 10^{-10}, \nu = 7.71)$ and for δ -variant (blue square symbols, blue bold line) $g(\mu = 20.67, \sigma = 3.63, \alpha = 1.123 \times 10^{-10}, \nu = 7.71)$.

Finally, we explore the effect of adopting uniformly high ventilation rates and masks, on the distribution of secondary infections. To that end, we keep the ventilation rates constant at $ACH = 5 \text{ hr}^{-1}$ and assume that the entire population (including the infectors and susceptibles) are wearing masks that provide 50% reduction by volume of exhaled aerosols and 50% reduction in the correspondingly inhaled aerosols. The results are shown in Fig 13.

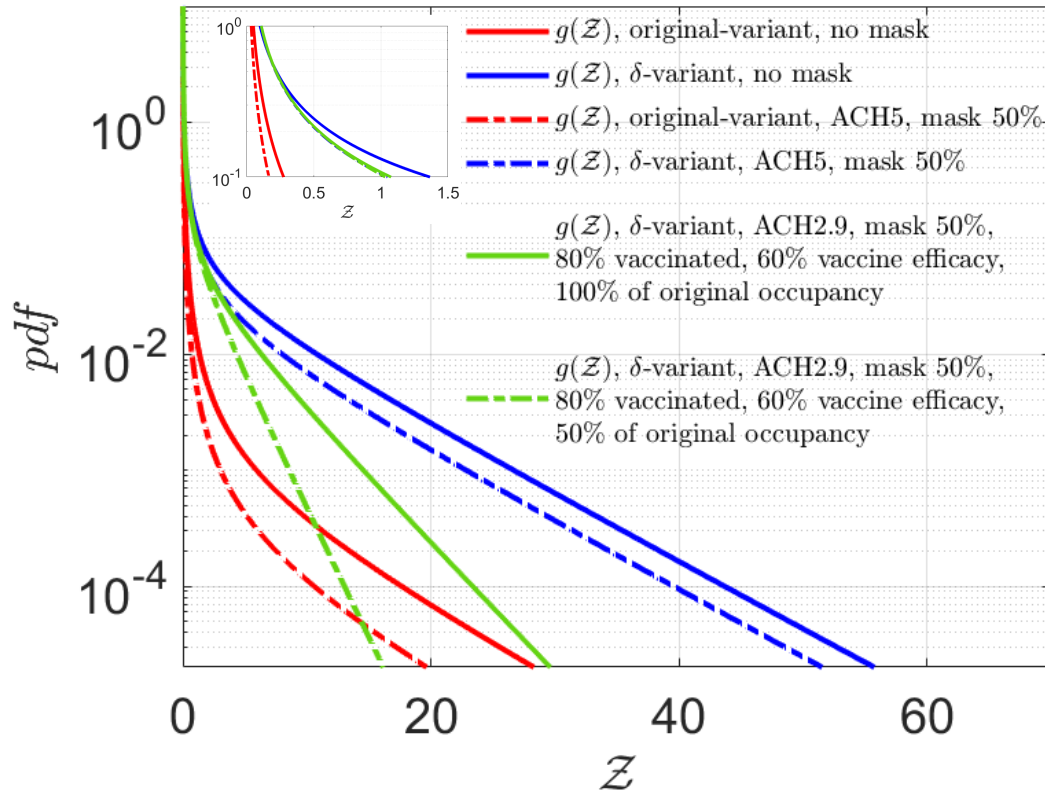


Figure 13: Effect of masking, fixed ventilation rates, vaccines, and reduced occupancy. Solid red and blue lines correspond to the analytical solutions given by Eqn. 22 for the original and δ -variant respectively, for $\langle ACH \rangle = 2.9 \text{ hr}^{-1}$ and without masks, as shown in Fig. 12. For fixed $ACH = 5 \text{ hr}^{-1}$, and with masks blocking 50 % volume of aerosols during inhalation and exhalation, the dashed red line shows the analytical solution for original variant $g(\mu = 13.84, \sigma = 3.63, \alpha = 2.36 \times 10^{-11}, \nu = 7.71)$ while the dashed blue line shows the analytical result for δ -variant $g(\mu = 20.67, \sigma = 3.63, \alpha = 2.36 \times 10^{-11}, \nu = 7.71)$. The solid green line: $g(\mu = 20.67, \sigma = 3.63, \alpha = 2.81 \times 10^{-11}, \nu_v = 4.01)$ shows the effect of 80% vaccination coverage, with 60% vaccine efficacy, with all individuals wearing masks that block 50% of the aerosols during exhalation and inhalation. The dash-dotted green line: $g(\mu = 20.67, \sigma = 3.63, \alpha = 2.81 \times 10^{-11}, \nu_v = 2.00)$ shows the effect of 80% vaccination coverage, with 60% vaccine efficacy, with all individuals wearing masks that block 50% of the aerosols during exhalation and inhalation along with occupancy restriction to 50% of the original occupancy. The top left inset shows the zoomed in view of the left side of the pdfs. Please refer Table 3 in the supplementary material for detailed values.

We observe that such interventions result in significant reduction in transmissibility for the original variant; reducing the mean to $\langle Z \rangle = 0.04$ (from $\langle Z \rangle = 0.13$ obtained without any interventions) with significant reduction in the extension of the tail. Such intervention effect remains substantial, yet less pronounced for the δ -variant where the $\langle Z \rangle = 1.69$ (w.r.t. $\langle Z \rangle = 2.64$ obtained without any interventions) and tail remains sufficiently stretched with some shift in the overall pdf towards lower Z . Once again, we note that these numbers are obtained over nearly an hour of exposure time on average. Note that, in our model, α does not change between the original and the δ variant. Only the μ increases, resulting in an increase in the higher proportion of secondary attack rates close to unity as the virus strain switches from the original variant to the δ -variant. A more detailed analysis of the influence of individual parameters: mean viral load, mean occupancy, and mean ventilation rates on the mean and standard deviation of Z could be found in the supplementary material (section 5).

Within the scope of the present study - social gatherings in restaurants, we ask what kind of spread could be expected for the δ -variant given the period of exposure and available occupancy data, in a population where a large fraction is already vaccinated? This is shown in Fig. 13 by the green curves. Using the realistic *ACH* distribution and with masks that can reduce both emission and inhalation of aerosol volumes by 50%, respectively, we do a basic calculation including the effect of vaccination. Assuming vaccine efficacy $\eta_{vac} = 0.6$ and vaccination coverage efficiency $\eta_{cov} = 0.8$ representing fraction of the population vaccinated, we estimate the new population of susceptible individuals at a given POI as $n_v = (1 - \eta_{vac}\eta_{cov})n$. We do not consider any change in viral load or change in distribution of infectious cases. Fitting an exponential distribution to n_v , the new constant $\nu_v = 4.01$. Clearly, from Fig. 13 we observe a significant drop in the number of secondary infections and superspreading events. The pdf of the number of secondary infections with the δ -variant with partially effective masks and vaccines is much less stretched than the original variant without masks or vaccine. Still the finite risk of superspreading event sustains. However, with 80% vaccination and 50% reduced occupancy, coupled with masks, a significant reduction in overdispersion is attained. This behooves the need for rapidly vaccinating the population alongside physical intervention measures like high-quality masks, reduced occupancy, and

across the board higher ventilation rates.

4 Conclusions

Overdispersion leading to superspreading events is a major driving force of the Covid-19 pandemic. Understanding the factors that lead to overdispersion in individual infectivity are also considered long standing scientific problems. Coupling an aerosol mixing model with real-world inputs: exhaled aerosol size distribution for speech and breath, measured viral load distribution, occupancy information from SafeGraph for full-service restaurants in major US cities, realistic ventilation rate distributions, we explore the overdispersion in the number of secondary infections per infector. The simulated results demonstrate that aerosol transmission route is consistent with the overdispersed individual infectivity with viral load variability being a dominant factor that controls secondary attack rates alongside ventilation rate, exposure time, and speaking time. We derive analytical expressions that can accurately and probably for the first time, describe the simulated pdfs of the secondary attack rates and number of secondary infections, respectively. The function for the secondary infection number pdf elucidates quantitatively, how overdispersed distributions of viral load and occupancy, jointly control the overdispersion in the number of secondary infections per infector. Finally, given the input data, modeling assumptions and the scope of the study, it appears significant reduction in transmissibility (both the average, as well as the dispersion) of the highly transmissible δ -variant necessitates all possible mitigation measures: efficient masks, high ventilation rates, and reduced occupancy, even after a significant fraction of the population has been vaccinated.

5 Acknowledgements

UofT researchers acknowledge support from the Fields Institute for Research in Mathematical Sciences through their project Mathematics for Public Health and Variants of Concern sponsored by the Canadian Institutes of Health Research. SC acknowledges the Heuckroth Distinguished Faculty Award in Aerospace Engineering from UTIAS. SM acknowledges the

Tier 2 Canada Research Chair in Mathematical Modeling and Program Science (grant No. CRC-950-232643).

References

- [1] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz, “Superspreading and the effect of individual variation on disease emergence,” *Nature*, vol. 438, no. 7066, pp. 355–359, 2005.
- [2] L. Hamner, “High sars-cov-2 attack rate following exposure at a choir practice—skagit county, washington, march 2020,” *MMWR. Morbidity and mortality weekly report*, vol. 69, 2020.
- [3] S. L. Miller, W. W. Nazaroff, J. L. Jimenez, A. Boerstra, G. Buonanno, S. J. Dancer, J. Kurnitski, L. C. Marr, L. Morawska, and C. Noakes, “Transmission of sars-cov-2 by inhalation of respiratory aerosol in the skagit valley chorale superspreading event,” *Indoor air*, vol. 31, no. 2, pp. 314–323, 2021.
- [4] J. Lu, J. Gu, K. Li, C. Xu, W. Su, Z. Lai, D. Zhou, C. Yu, B. Xu, and Z. Yang, “Covid-19 outbreak associated with air conditioning in restaurant, guangzhou, china, 2020,” *Emerging infectious diseases*, vol. 26, no. 7, p. 1628, 2020.
- [5] Y. Li, H. Qian, J. Hang, X. Chen, P. Cheng, H. Ling, S. Wang, P. Liang, J. Li, S. Xiao, *et al.*, “Probable airborne transmission of sars-cov-2 in a poorly ventilated restaurant,” *Building and Environment*, p. 107788, 2021.
- [6] S. Y. Park, Y.-M. Kim, S. Yi, S. Lee, B.-J. Na, C. B. Kim, J.-i. Kim, H. S. Kim, Y. B. Kim, Y. Park, *et al.*, “Coronavirus disease outbreak in call center, south korea,” *Emerging infectious diseases*, vol. 26, no. 8, p. 1666, 2020.
- [7] X. Hao, S. Cheng, D. Wu, T. Wu, X. Lin, and C. Wang, “Reconstruction of the full transmission dynamics of covid-19 in wuhan,” *Nature*, vol. 584, no. 7821, pp. 420–424, 2020.

- [8] A. Endo *et al.*, “Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china,” *Wellcome open research*, vol. 5, 2020.
- [9] G. Großmann, M. Backenköhler, and V. Wolf, “Heterogeneity matters: Contact structure and individual variation shape epidemic dynamics,” *Plos one*, vol. 16, no. 7, p. e0250050, 2021.
- [10] P. Z. Chen, N. Bobrovitz, Z. Premji, M. Koopmans, D. N. Fisman, and F. X. Gu, “Heterogeneity in transmissibility and shedding sars-cov-2 via droplets and aerosols,” *Elife*, vol. 10, p. e65774, 2021.
- [11] K. Sneppen, B. F. Nielsen, R. J. Taylor, and L. Simonsen, “Overdispersion in covid-19 increases the effectiveness of limiting nonrepetitive contacts for transmission control,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 14, 2021.
- [12] P. Z. Chen, M. Koopmans, D. N. Fisman, and F. X. Gu, “Understanding why superspreading drives the covid-19 pandemic but not the h1n1 pandemic,” *The Lancet Infectious Diseases*, 2021.
- [13] L. Morawska and D. K. Milton, “It is Time to Address Airborne Transmission of COVID-19,” *Clinical Infectious Diseases*, 07 2020. ciaa939.
- [14] J. G. Allen and L. C. Marr, “Recognizing and controlling airborne transmission of sars-cov-2 in indoor environments,” *Indoor air*, vol. 30, no. 4, p. 557, 2020.
- [15] F. Gregson, J. Robinson, R. Miles, C. Royall, and J. Reid, “Drying kinetics of salt solution droplets: Water evaporation rates and crystallization,” *The Journal of Physical Chemistry B*, vol. 123, no. 1, pp. 266–276, 2018.
- [16] J. W. Tang, W. P. Bahnfleth, P. M. Bluyssen, G. Buonanno, J. L. Jimenez, J. Kurnitski, Y. Li, S. Miller, C. Sekhar, L. Morawska, *et al.*, “Dismantling myths on the airborne transmission of severe acute respiratory syndrome coronavirus (sars-cov-2),” *Journal of Hospital Infection*, 2021.

- [17] T. Greenhalgh, J. L. Jimenez, K. A. Prather, Z. Tufekci, D. Fisman, and R. Schooley, “Ten scientific reasons in support of airborne transmission of sars-cov-2,” *The lancet*, vol. 397, no. 10285, pp. 1603–1605, 2021.
- [18] K. A. Prather, L. C. Marr, R. T. Schooley, M. A. McDiarmid, M. E. Wilson, and D. K. Milton, “Airborne transmission of sars-cov-2,” *Science*, vol. 370, no. 6514, pp. 303–304, 2020.
- [19] S. Chaudhuri, S. Basu, and A. Saha, “Analyzing the dominant sars-cov-2 transmission routes toward an ab initio disease spread model,” *Physics of Fluids*, vol. 32, no. 12, p. 123306, 2020.
- [20] P. Dabisch, M. Schuit, A. Herzog, K. Beck, S. Wood, M. Krause, D. Miller, W. Weaver, D. Freeburger, I. Hooper, *et al.*, “The influence of temperature, humidity, and simulated sunlight on the infectivity of sars-cov-2 in aerosols,” *Aerosol Science and Technology*, vol. 55, no. 2, pp. 142–153, 2021.
- [21] L. Bourouiba, E. Dehandschoewercker, and J. W. Bush, “Violent expiratory events: on coughing and sneezing,” *Journal of Fluid Mechanics*, vol. 745, pp. 537–563, 2014.
- [22] M. Abkarian, S. Mendez, N. Xue, F. Yang, and H. A. Stone, “Speech can produce jet-like transport relevant to asymptomatic spreading of virus,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 41, pp. 25237–25245, 2020.
- [23] W. Chen, N. Zhang, J. Wei, H.-L. Yen, and Y. Li, “Short-range airborne route dominates exposure of respiratory infection during close contact,” *Building and Environment*, p. 106859, 2020.
- [24] E. Riley, G. Murphy, and R. Riley, “Airborne spread of measles in a suburban elementary school,” *American journal of epidemiology*, vol. 107, no. 5, pp. 421–432, 1978.
- [25] S. Rudnick and D. Milton, “Risk of indoor airborne infection transmission estimated from carbon dioxide concentration,” *Indoor air*, vol. 13, no. 3, pp. 237–245, 2003.

- [26] G. Buonanno, L. Morawska, and L. Stabile, “Quantitative assessment of the risk of airborne transmission of sars-cov-2 infection: prospective and retrospective applications,” *Environment international*, vol. 145, p. 106112, 2020.
- [27] M. Z. Bazant and J. W. Bush, “A guideline to limit indoor airborne transmission of covid-19,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 17, 2021.
- [28] J. Schijven, L. C. Vermeulen, A. Swart, A. Meijer, E. Duizer, and A. M. de Roda Husman, “Quantitative microbial risk assessment for airborne transmission of sars-cov-2 via breathing, speaking, singing, coughing, and sneezing,” *Environmental health perspectives*, vol. 129, no. 4, p. 047002, 2021.
- [29] T. C. Bond, A. Bosco-Lauth, D. K. Farmer, P. W. Francisco, J. R. Pierce, K. M. Fedak, J. M. Ham, S. H. Jathar, and S. VandeWoude, “Quantifying proximity, confinement, and interventions in disease outbreaks: a decision support framework for air-transported pathogens,” *Environmental science & technology*, vol. 55, no. 5, pp. 2890–2898, 2021.
- [30] C. C. Wang, K. A. Prather, J. Sznitman, J. L. Jimenez, S. S. Lakdawala, Z. Tufekci, and L. C. Marr, “Airborne transmission of respiratory viruses,” *Science*, vol. 373, no. 6558, p. eabd9149, 2021.
- [31] R. Mittal, R. Ni, and J.-H. Seo, “The flow physics of covid-19,” *Journal of fluid Mechanics*, vol. 894, 2020.
- [32] L. Bourouiba, “The fluid dynamics of disease transmission,” *Annual Review of Fluid Mechanics*, vol. 53, pp. 473–508, 2021.
- [33] M. L. Pöhlker, O. O. Krüger, J.-D. Förster, T. Berkemeier, W. Elbert, J. Fröhlich-Nowoisky, U. Pöschl, C. Pöhlker, G. Bagheri, E. Bodenschatz, *et al.*, “Respiratory aerosols and droplets in the transmission of infectious diseases,” *arXiv preprint arXiv:2103.01188*, 2021.
- [34] S. Chaudhuri, A. Saha, and S. Basu, “An opinion on the multi-scale nature of covid-19 type disease spread,” *Current Opinion in Colloid & Interface Science*, p. 101462, 2021.

- [35] S. B. Pope, *Turbulent Flows*. Cambridge University Press, 2006.
- [36] P. J. Drivas, P. A. Valberg, B. L. Murphy, and R. Wilson, “Modeling indoor air exposure from short-term point source releases,” *Indoor Air*, vol. 6, no. 4, pp. 271–277, 1996.
- [37] S. Chaudhuri, S. Basu, P. Kabi, V. R. Unni, and A. Saha, “Modeling the role of respiratory droplets in covid-19 type pandemics,” *Physics of Fluids*, vol. 32, no. 6, p. 063309, 2020.
- [38] U. S. Department of Homeland Security, “Estimated airborne decay of sars-cov-2; <https://www.dhs.gov/science-and-technology/sars-airborne-calculator>,” tech. rep.
- [39] D. H. Morris, K. C. Yinda, A. Gamble, F. W. Rossine, Q. Huang, T. Bushmaker, R. J. Fischer, M. J. Matson, N. Van Doremalen, P. J. Vikesland, *et al.*, “Mechanistic theory predicts the effects of temperature and humidity on inactivation of sars-cov-2 and other enveloped viruses,” *Elife*, vol. 10, p. e65902, 2021.
- [40] L. C. Marr, J. W. Tang, J. Van Mullekom, and S. S. Lakdawala, “Mechanistic insights into the effect of humidity on airborne influenza virus survival, transmission and incidence,” *Journal of the Royal Society Interface*, vol. 16, no. 150, p. 20180298, 2019.
- [41] K.-C. Cheng, V. Acevedo-Bolton, R.-T. Jiang, N. E. Klepeis, W. R. Ott, O. B. Fringer, and L. M. Hildemann, “Modeling exposure close to air pollution sources in naturally ventilated residences: Association of turbulent diffusion coefficient with air change rate,” *Environmental science & technology*, vol. 45, no. 9, pp. 4016–4022, 2011.
- [42] C. N. HAAS, “Estimation of risk due to low doses of microorganisms: a comparison of alternative methodologies,” *American journal of epidemiology*, vol. 118, no. 4, pp. 573–582, 1983.
- [43] C. N. Haas, “Action levels for sars-cov-2 in air: Preliminary approach,” *Risk Analysis*, vol. 41, no. 5, pp. 705–709, 2021.
- [44] S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky, and J. Leskovec, “Mobility network models of covid-19 explain inequities and inform reopening,” *Nature*, vol. 589, no. 7840, pp. 82–87, 2021.

- [45] H. Kawasuji, Y. Takegoshi, M. Kaneda, A. Ueno, Y. Miyajima, K. Kawago, Y. Fukui, Y. Yoshida, M. Kimura, H. Yamada, *et al.*, “Transmissibility of covid-19 depends on the viral load around onset in adult and symptomatic patients,” *PloS one*, vol. 15, no. 12, p. e0243597, 2020.
- [46] X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, *et al.*, “Temporal dynamics in viral shedding and transmissibility of covid-19,” *Nature medicine*, vol. 26, no. 5, pp. 672–675, 2020.
- [47] Q. Yang, T. K. Saldi, P. K. Gonzales, E. Lasda, C. J. Decker, K. L. Tat, M. R. Fink, C. R. Hager, J. C. Davis, C. D. Ozeroff, *et al.*, “Just 2% of sars-cov-2- positive individuals carry 90% of the virus circulating in communities,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 21, 2021.
- [48] R. McClain, “How much room do i need for a restaurant? <https://bizfluent.com/info-12010139-much-room-need-restaurant.html>,” 2018.
- [49] H. R. Bohanon, J.-J. Piade, M. K. Schorp, and Y. Saint-Jalm, “An international survey of indoor air quality, ventilation, and smoking activity in restaurants: a pilot study,” *Journal of Exposure Science & Environmental Epidemiology*, vol. 13, no. 5, pp. 378–392, 2003.
- [50] J. A. Lednicky, M. Lauzard, Z. H. Fan, A. Jutla, T. B. Tilly, M. Gangwar, M. Usmani, S. N. Shankar, K. Mohamed, A. Eiguren-Fernandez, *et al.*, “Viable sars-cov-2 in the air of a hospital room with covid-19 patients,” *International Journal of Infectious Diseases*, vol. 100, pp. 476–482, 2020.
- [51] B. Li, A. Deng, K. Li, Y. Hu, Z. Li, Q. Xiong, Z. Liu, Q. Guo, L. Zou, H. Zhang, M. Zhang, F. Ouyang, J. Su, W. Su, J. Xu, H. Lin, J. Sun, J. Peng, H. Jiang, P. Zhou, T. Hu, M. Luo, Y. Zhang, H. Zheng, J. Xiao, T. Liu, R. Che, H. Zeng, Z. Zheng, Y. Huang, J. Yu, L. Yi, J. Wu, J. Chen, H. Zhong, X. Deng, M. Kang, O. G. Pybus, M. Hall, K. A. Lythgoe, Y. Li, J. Yuan, J. He, and J. Lu, “Viral infection and transmission in a large, well-traced outbreak caused by the sars-cov-2 delta variant,” *medRxiv*, 2021.

- [52] M. C. Shamier, A. Tostmann, S. Bogers, J. De Wilde, J. IJpelaar, W. A. van der Kleij, H. De Jager, B. L. Haagmans, R. Molenkamp, B. B. O. Munnink, *et al.*, “Virological characteristics of sars-cov-2 vaccine breakthrough infections in health care workers,” *medRxiv*, 2021.