RESEARCH ARTICLE | MAY 31 2022

Analysis of overdispersion in airborne transmission of COVID-19 🅕 FREE

Special Collection: Flow and the Virus

Swetaprovo Chaudhuri ✉ 🆔 ; Prasad Kasibhatla; Arnab Mukherjee; William Pan; Glenn Morrison; Sharmistha Mishra; Vijaya Kumar Murty

Check for updates

View Online   Export Citation   CrossMark

Analysis of overdispersion in airborne transmission of COVID-19 (F)

View Online   Export Citation   CrossMark

Swetaprovo Chaudhuri,[1,a)] (ID) Prasad Kasibhatla,[2] Arnab Mukherjee,[1] William Pan,[3] Glenn Morrison,[4] Sharmistha Mishra,[5] and Vijaya Kumar Murty[6,7]

AFFILIATIONS

[1]Institute for Aerospace Studies, University of Toronto, Toronto, Ontario M3H 5T6, Canada
[2]Nicholas School of Environment, Duke University, Durham, North Carolina 27710, USA
[3]Global Institute of Health, Duke University, Durham, North Carolina 27710, USA
[4]Gilligs School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina 27599, USA
[5]Division of Infectious Diseases, Department of Medicine, University of Toronto, Toronto, Ontario M5S 3H2, Canada
[6]Fields Institute for Research in Mathematical Sciences, Toronto, Ontario M5T 3J1, Canada
[7]Department of Mathematics, University of Toronto, Toronto, Ontario M5S 2E4, Canada

Note: This paper is part of the special topic, Flow and the Virus.
a)Author to whom correspondence should be addressed: swetaprovo.chaudhuri@utoronto.ca

ABSTRACT

Superspreading events and overdispersion are hallmarks of the COVID-19 pandemic. However, the specific roles and influence of established viral and physical factors related to the mechanisms of transmission, on overdispersion, remain unresolved. We, therefore, conducted mechanistic modeling of SARS-CoV-2 point-source transmission by infectious aerosols using real-world occupancy data from more than 100 000 social contact settings in ten US metropolises. We found that 80% of secondary infections are predicted to arise from approximately 4% of index cases, which show up as a stretched tail in the probability density function of secondary infections per infectious case. Individual-level variability in viral load emerges as the dominant driver of overdispersion, followed by occupancy. We then derived an analytical function, which replicates the simulated overdispersion, and with which we demonstrate the effectiveness of potential mitigation strategies. Our analysis, connecting the 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 regard to airborne transmission of SARS-CoV-2 and thereby emerges as a powerful tool toward assessing the probability of outbreaks and the potential impact of mitigation strategies on large scale disease dynamics.

I. 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 leads 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 is an expected feature that emerges from the right-sided tail beyond 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

26 January 2024 00:00:59

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 the SARS-CoV-2 spread. Agent based modeling by Sneppen *et al.*[11] suggested that non-repeating, random contacts, such as those in restaurants and bars, are a dominant contributor to the SARS-CoV-2 spread. Chen *et al.*[12] argued 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 contributes 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 confirm the importance of airborne transmission of SARS-CoV-2 by respiratory aerosols.[13–17] A large number of small respiratory droplets (when size $<100\,\mu m$ at the point of exhalation often referred to 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, a 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 a similar well-mixed approach, Bazant and Bush[27] proposed to restrict the occupancy number and time spent in a room to mitigate airborne transmission.[27] Schijven *et al.*[28] estimated risk of infection resulting from sneezing, coughing, speaking, breathing, and singing at different viral loads. Analysis by Bond *et al.*[29] quantified the importance of confinement in pathogen transport using "effective rebreathed air volume." Dbouk and Drikakis[30–32] studied the transport and evaporation of virus contaminated saliva droplets and looked at how these can be implemented in practical models. Further details on specific aspects of airborne disease transmission, including but not limited to aerosols, flow physics, and respiratory droplet size distribution, could be found in a recent review and opinion articles.[33–37] Yet, to date mechanistic models describing airborne transmission have not been coupled with real-world distributions and occupancy information toward understanding the large scale features of disease dynamics, for example, overdispersion in transmissibility.

The overarching goal of this work is to leverage the mechanistic underpinnings of the airborne disease transmission to explore event-level overdispersion of the SARS-CoV-2 spread using 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 the 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. We focus specifically on 103 679 restaurant settings from ten metropolitan areas in the US, motivated by the finding that these locations represent important centers of transmission in the early phase of the pandemic in the US.[38] Transmission in such random, non-repetitive settings represents point source transmission. Emergence of later variants (delta, omicron) characterized by shorter generation times with respect to the original variant[39,40] highlights the importance of understanding transmission and overdispersion in point source transmission settings. We note that the simulation in this paper addresses overdispersion in the number of secondary cases generated per infectious case over a period of about an hour in such hundred thousand instances, conditional on the presence of one index case at each location. The analytical expressions developed and validated with the simulations can address overdispersion over much wider time scales and contact settings, when coupled with appropriate inputs. In any case, in this paper, we do not directly address the number of secondary cases spawned by an infectious case over their entire course of infection experiencing different types of contact.

II. ALGORITHM

A semi-analytical framework based on turbulent diffusion of infectious aerosols exhaled by an infectious, asymptomatic (or pre-symptomatic) individual is coupled with a dose response model[41] to compute the number of secondary infections inside a room of specified dimensions and ventilation rate. This is described in detail in Sec. III. Using random samples of viral load, volume flow rate of ejected respiratory liquid, air changes per hour (ACH) generated from correspondingly observed distributions, along with real world occupancy and area data from SafeGraph, and exhaled aerosol size distribution, we use the algorithm depicted in Fig. 1 to generate the probability density function (pdf) of the number of secondary infections per infectious case.

The algorithm runs over N_s iterations where $N_s = 103\,679$ corresponding to the number of full service restaurants in ten US cities, available from SafeGraph data, used for generating a pdf at the end; i being the iteration index. The restaurants serve as proxy for non-repeating, random, social contact settings and are referred to as points of interest (POI). First, at each such an indoor micro-environment of

FIG. 1. The algorithm for estimating the distribution of secondary infections Z.

to generate the number of virions emitted per unit time. The reader is requested to refer to the Appendix for details on the truncated normal distribution of the volume flow rate of ejected mucosalivary liquid. This is used in the turbulent diffusion solutions shown in Eqs. (4)–(7) in Sec. III, which establishes the spatio-temporally resolved virion concentration field. Random samples of exposure time, speaking time, *ACH*, and area of the point of interest (POI), along with virus half-life, which is calculated separately, also act as input parameters for the solver. The resulting virus concentration field is used in the dose response model (with a specified dose response constant) to obtain probability of infection fields. Next, these probability fields are multiplied with the POI specific population density data (from SafeGraph) to obtain the number of secondary infections: Z. Once the iteration index 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—the 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 $\mathcal{Z}$,[42] is a convenient mathematical 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 represents finite (but could be small) probability of an event where $Z \gg$ 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 in Sec. III; details of the rest of the inputs: ventilation rate distribution, speech time, and exposure time distributions can be found in the Appendix. A table summarizing the important parameter values, distributions, and sources can also be found there (Table I).

III. MATERIALS AND METHODS

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 the number of people occupying specific indoor areas. The occupancy information is obtained from a large SafeGraph dataset 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.

the POI, the volume of the mucosalivary fluid ejected per unit time due to speech and breath (random sample obtained from a truncated normal distribution whose mean is evaluated from the exhaled respiratory droplet volume size distribution, vsd) multiplied with a random sample from the log-normal viral load distribution is fed into the code

TABLE I. Input parameters.

Symbols	Definitions	Values	References
N_s	Total number of random samples used	103 679	Total number of data points from restaurants over ten US cities
ρ	Viral load—log-normal distribution with mean μ and standard deviation σ	$\mu = 13.839\,4$ and $\sigma = 3.630\,1$ for original variant	Data from Ref. 57
ACH	Air change rate (ACH)—measure of amount of air replaced in a room per unit time per unit volume of the room, log-normal distribution	$\mu_{ACH} = 0.7701$, $\sigma_{ACH} = 0.7554$	Indoor ventilation rate is typically log-normally distributed. Parameters obtained from measured ACH distribution in restaurants from Ref. 59
$\dot{Q}_{ls}$	Volume of respiratory liquid ejected during speaking per unit time	$\langle \dot{Q}_{ls} \rangle = 2.222 \times 10^{-6}\,\mathrm{ml/s}$, $\sigma_{\dot{Q}_{ls}} = 3.078 \times 10^{-6}\,\mathrm{ml/s}$	Data from Refs. 36 and 66–69 were used to determine the normal distribution parameters
N	Susceptible population at a given POI from SafeGraph data, exponential distribution	$\nu = 7.707$	SafeGraph occupancy and area datasets
τ	Exposure time of a susceptible individual to the ejected aerosols, log-normal distribution	$\mu_\tau = 0$, $\sigma_\tau = 0.4$	Distribution chosen such that average exposure time corresponds to the occupancy time over which the seven day averaged SafeGraph data are obtained.
t_s	Duration of speaking of the infected individual, uniform distribution	$0.25\tau \leq t_s \leq 0.5\tau$	Estimated such that $t_s \leq 0.5\tau$ to ensure the speaking time is shared between infectors and susceptibles
$D_{s,0}$	Initial droplet diameter distribution (log-normal)	$10^{-2}\mu\mathrm{m} \leq D_{s,0} \leq 10^2\mu\mathrm{m}$	Range used by Pöhlker et al.[36]
D_i	Mean geometric diameter of a mode of expiratory event (b—breath, s—speech)	$(D_1)_b = 0.07\mu\mathrm{m}, (D_2)_b = 0.3\mu\mathrm{m},$ $(D_1)_s = 0.07\mu\mathrm{m}, (D_2)_s = 0.3\mu\mathrm{m},$ $(D_3)_s = 1.00\mu\mathrm{m}, (D_4)_s = 10\mu\mathrm{m},$ $(D_5)_s = 96\mu\mathrm{m}$	Input parameter in the log-normal fit function for respiration PSDs provided by Pöhlker et al.[36] (Table VI)
A_i	Number concentration at D_i	$(A_1)_b = 7.7\mathrm{cm}^{-3}, (A_2)_b = 1.1\mathrm{cm}^{-3},$ $(A_1)_s = 9.8\mathrm{cm}^{-3}, (A_2)_s = 1.4\mathrm{cm}^{-3},$ $(A_3)_s = 1.7\mathrm{cm}^{-3}, (A_4)_s = 0.03\mathrm{cm}^{-3},$ $(A_5)_s = 0.17\mathrm{cm}^{-3}$	Input parameter in the log-normal fit function for respiration PSDs provided by Pöhlker et al.[36] (Table VI)
σ_i	Modal geometric standard deviation	$(\sigma_1)_b = 0.9, (\sigma_2)_b = 0.9, (\sigma_1)_s = 0.9,$ $(\sigma_2)_s = 0.9, (\sigma_3)_s = 0.9, (\sigma_4)_s = 0.98,$ $(\sigma_5)_s = 0.97$	Input parameter in the log-normal fit function for respiration PSDs provided by Pöhlker et al.[36] (Table VI)
$\dot{V}$	Volume of air exhaled in an expiratory event (b—breath, s—speech)	$\dot{V}_s = 194.4\mathrm{cm}^3\mathrm{s}^{-1}$, $\dot{V}_b = 100\mathrm{cm}^3\mathrm{s}^{-1}$	Characteristic values used for calculations in Pöhlker et al.[36] (Table I)
$\dot{V}_b$	Inhaled volume of air	$100\mathrm{cm}^3\mathrm{s}^{-1}$	Inhalation and exhalation volume during one breath cycle considered to be equal. Value for exhalation volume obtained from Pöhlker et al.[36] (Table I)
H	Room height	$3\,\mathrm{m}$	
u^*	Friction velocity	$0.03\mathrm{ms}^{-1}$	Chosen from the range of values used in Ref. 64. Figure 4[64] shows that all values in the range have overlapping deposition velocity curves in the higher particle size range, which determines the average deposition velocity value
T	Ambient temperature	$294.7\mathrm{K}$	Recommended typical indoor condition by ASHRAE
RH	Relative humidity	0.5	Recommended typical indoor condition by ASHRAE
UV_{index}	Strength of ultraviolet radiation	0	Recommended typical indoor condition by ASHRAE
$t_{1/2}$	Half-life of airborne virus within aerosols	$32.07\mathrm{min}$	Based on above parameters at standard indoor conditions, from Refs. 20 and 51 estimated using the DHS calculator
r_v	Dose response constant	1/1440 copies	Obtained from Ref. 28

A. Aerosol dispersion from an infectious case

We use the standard, turbulent diffusivity based closure[42] to model the spatio-temporal evolution of ensemble averaged infectious aerosol concentration c_a (the number of aerosol particles per unit volume of air associated with emissions from the infector), as shown in Eq. (1). This approach was used by Drivas et al.[43] to model indoor concentration fields from point sources. Venkatram and Weil[44] have also shown that gradient transport based turbulence models work reasonably well in indoor environments.

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

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

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 six walls. Since the virus remains embedded inside the infectious aerosols and 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 Refs. 19, 23, and 45 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

$$c = c_a \mathcal{N}_a. \tag{2}$$

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

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

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. Indeed, direct numerical simulation or large eddy simulation based computational fluid dynamics (CFD) could produce high fidelity solutions for one specific micro-environment at high computational cost similar to the studies by Pendar and Páscoa[46] who looked at the fluid dynamics and transport characteristics of a sneeze within an indoor location, and Zheng et al.[47] who studied the transmission of aerosolized viruses within a densely populated urban area. Other viable techniques like the Lagrangian particle model coupled with a continuous random walk model used by Wang et al.[48] to simulate the droplet dynamics of expelled droplets also exist as alternatives. However, it is evident that for the 100 000 micro-environments investigated here, such techniques

cannot be used. Investigating such a large number of cases behooves application of a robust reduced order model for aerosol dispersion. The robustness of the present reduced order model used is proven by the comparison with experimental results as in Ref. 49 as well as the comparison with experimental and CFD results given by Hathway et al.[50] These will be discussed later in further details. 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[43]

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

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

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

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

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

L, W, and 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 Ref. 19

$$\psi(t) = (1/2)^{t/t_{\frac{1}{2}}},$$
$$t_{\frac{1}{2}} = f(T, RH, UV_{index}). \tag{8}$$

Based on the experiments by Dabisch et al.[20] and calculations from DHS,[51] for $T = 21.7\,°$C, $RH = 0.50$, $UV_{index} = 0$, i.e., typical ASHRAE recommended indoor air conditions, for SARS-CoV-2, half-life comes out as $t_{\frac{1}{2}} = 32.07$ min. Here, T is the temperature, RH is the relative-humidity, and UV_{index} is the ultra-violet index inside the room of interest. The SARS-CoV-2 virus half-life reduces monotonically with temperature and non-monotonically with relative humidity[52] like other enveloped viruses, as shown by Marr et al.[53]

The above solution given in Eqs. (4)–(7) [without $\psi(t)$] was validated by Cheng et al.[49] who released CO from a point source and measured its spatiotemporal 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 among turbulent diffusivity, area of room, and air change rate:

$$D_T = L_0^2 \times (0.52 \times ACH + 0.32)/3600, \tag{9}$$

which is used for the present study as well. Here, L_0 is given as $(A \times H)^{1/3}$, ACH is the air changes per hour, and therefore, $a = ACH/3600$. The model described through Eqs. (1)–(9) was further compared with the experimental and computational results provided by Hathway et al.[50] for central point source bioaerosol dispersion within a ventilated room. The model used shows qualitative and quantitative match within the uncertainty limits. The detailed comparison is provided in the Appendix.

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_o, that is inhaled at x, y, z_0 up to time τ is denoted $\mathcal{N}_v(x, y, z_0, \tau)$, and it is given by

$$\mathcal{N}_v(x, y, z_0, \tau) = \int_0^\tau c(x, y, z_0, t') \dot{V}_b dt'. \quad (10)$$

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 the Appendix 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.[41] The dose-response constant is chosen as $r_v = 1/1440$ based on the estimations by Haas[54] and Schijven et al.[28] for the original SARS-CoV-2 variant

$$\mathcal{P}_{s+b}(x, y, z_0, \tau) = 1 - e^{-r_v \mathcal{N}_v(x, y, z_0, \tau)}. \quad (11)$$

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

$$Z = \int_0^L \int_0^W \rho_p \mathcal{P}_{s+b}(x, y, z_0, \tau) dx dy. \quad (12)$$

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 dataset and in Ref. 38 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.

B. Particle size distribution of exhaled aerosols

The particle size distributions at the source of the expiratory events, speaking and breathing, are obtained from the review by Pöhlker et al.[36] A multimode log-normal fitting has been found to describe the corresponding distribution reasonably well. The 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

$$\frac{d\eta}{dD_{s,0}} = \log_{10}(e) \sum_{i=1}^{n} \frac{A_i}{D_{s,0}} \exp\left(-\left[\frac{\ln(D_{s,0}/D_i)}{\sigma_i}\right]^2\right), \quad (13)$$

where η is the number 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 the expiratory event. For further details, the reader is referred to the Appendix. 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 term when multiplied with the volume flow of air per unit time gives us the mean volume flow rate of ejected respiratory liquid. (A truncated normal distribution is assumed for it, and more details can be found in the Appendix.) A random sample of this volume flow rate multiplied with the individual viral load ρ 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 Eq. (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 Appendix.

C. Viral load distribution

Viral load (ρ) has been associated with infectivity and, thus, the number of secondary infections for a given setting.[55,56] Analyzing respiratory droplets and aerosol laden cough jets, Chaudhuri et al.[19]
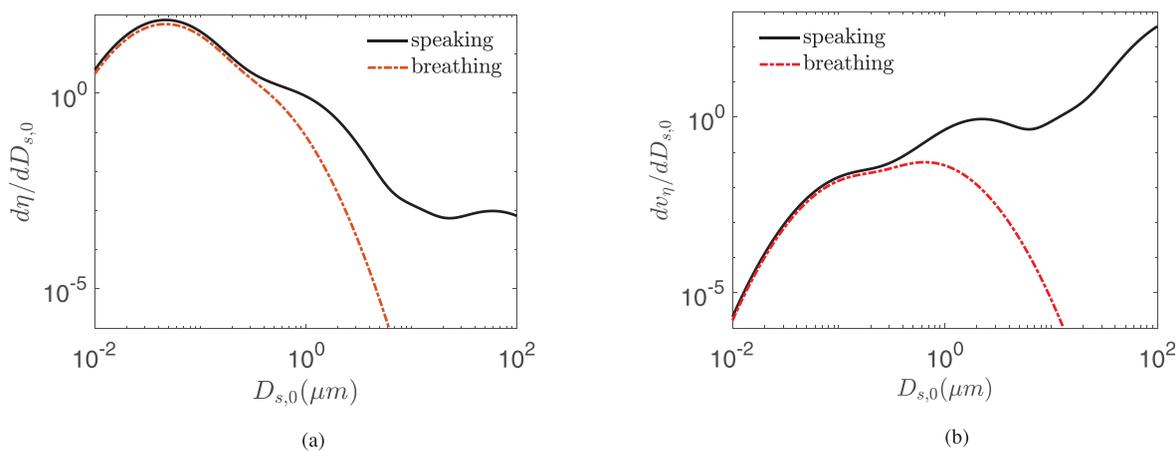
FIG. 2. Aerosol number size distribution (NSD) on top and volume size distribution (VSD) on bottom for speaking and breathing as expressed in Eqs. (B3) and (B5), respectively, from Ref. 36 as a function of the initial particle diameter sample space variable $D_{s,0}$, i.e., at the moment of exhalation.

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 the 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.*,[57] 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 Ref. 57 as an input into the algorithm shown in Fig. 1. According to Ref. 57 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(a). We generate and use samples of ρ from this log-normal distribution in our calculations.

D. 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 interest (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 are available from individual confined spaces, rather than that from a collection of several indoor spaces, were 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 sitting area since the dining area is estimated as 50% of the total restaurant area. (Based on typical restaurant design guides,[58] change in 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 datasets created by Chang *et al.*[38] 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. 3(b). Also, shown in the figure is an exponential fit and the corresponding fitting parameter.

IV. RESULTS AND DISCUSSION
A. Spatial distributions of the particle concentration for fixed conditions

First, we present the spatial distribution of the airborne virus concentration (RNA copies/m³ of air, virions are encapsulated within aerosols of initial size 0.01–100 μm) with the source at $x = 2.5$m and $y = 2.5$m from the origin (at the bottom left corner) for a specific micro-environment: a 10 m × 10 m × 3 m room, after 15 min of aerosol exhalation by speech and breath. In Fig. 4, the first column (a, c, and e) presents results with a constant viral load $\rho = 10^9$ copies/ml, but with increasing *ACH*. Note that the third row represents a case without wall reflections and, hence, can simulate outdoor conditions. The second column of Figs. 4(b), 4(d), and 4(f) represents a constant but five times higher viral load of $\rho = 5 \times 10^9$ copies/ml. The results demonstrate strong inhomogeneity of the virus concentration and also show that the contours scale linearly with ρ for the same *ACH* due the linear nature of the governing equation (3). Using the dose response model [Eq. (11)], the corresponding contours of probability of infection are shown in Fig. 5. It could be found that indeed near unity probability of infection (from speech and breath) $\mathcal{P}_{s+b}$ are found near the source, and there is decay with distance from the source. The subscript $s + b$ indicates infection caused by exhaled aerosols due to speech (s) and breath (b) from the infected individual. 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 in the virus concentration due to the non-linearity involved in the dose response. Interestingly, the simulated outdoor conditions shown in Figs. 4(e) and 4(f) show much smaller virus concentration both near and far from the source with respect to the confined cases.
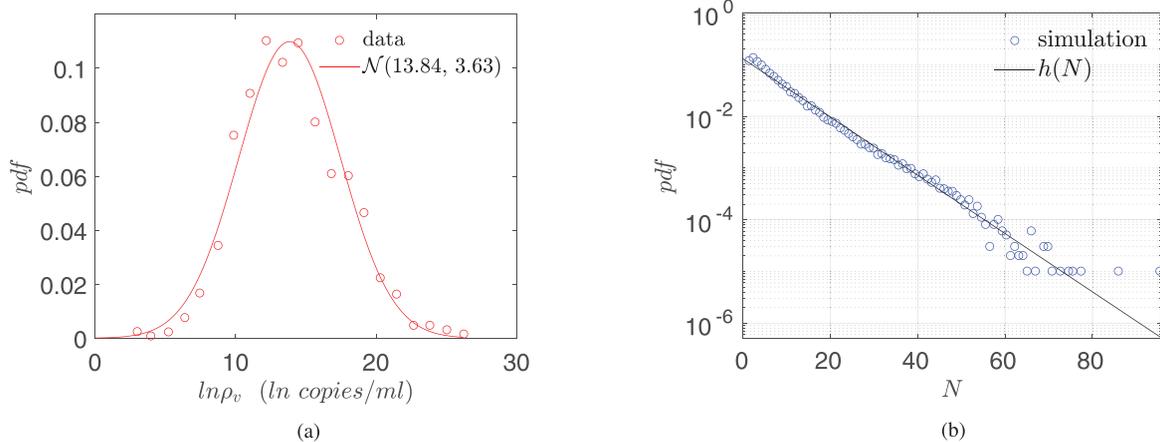
FIG. 3. (a) Viral load pdf for asymptomatic (including presymptomatic) population based on the histogram data from Ref. 57. The red line shows a normal distribution with μ and σ given by 13.84 and 3.63, respectively, and (b) pdf of the number of susceptible individuals in full service restaurants at ten US cities from SafeGraph data.
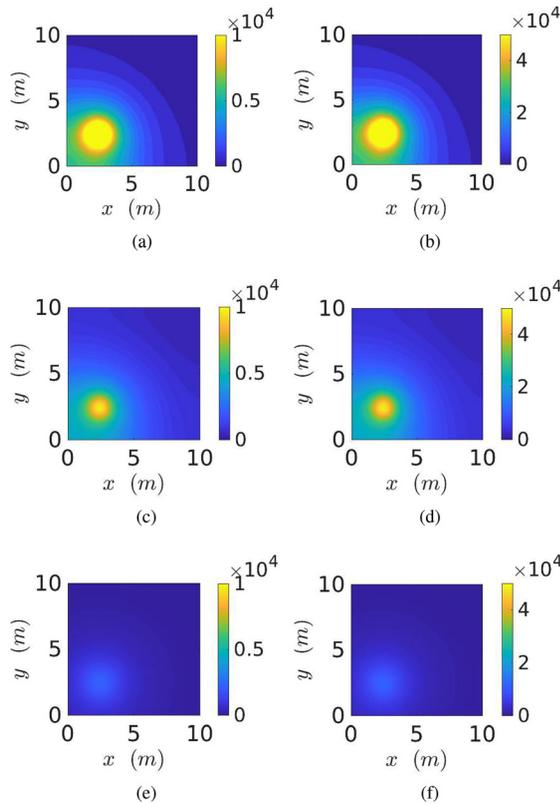
FIG. 4. Contour plots of the spatially resolved virus concentration (RNA copies/m^3) at time $t = 15$ min from start of the expiration event (the source located at $x = 2.5$m, $y = 2.5$m) (a) $\rho = 10^9$ copies/ml, $ACH = 2\,\mathrm{h}^{-1}$, (b) $\rho = 5 \times 10^9$ copies/ml, $ACH = 2\,\mathrm{h}^{-1}$, (c) $\rho = 10^9$ copies/ml, $ACH = 5\,\mathrm{h}^{-1}$, (d) $\rho = 5 \times 10^9$ copies/ml, $ACH = 5\,\mathrm{h}^{-1}$, (e) $\rho = 10^9$ copies/ml, $ACH = 12\,\mathrm{h}^{-1}$, and (f) $\rho = 5 \times 10^9$ copies/ml, $ACH = 12\,\mathrm{h}^{-1}$.

FIG. 5. Contour plots of the spatially resolved probability of infection $\mathcal{P}_{s+b}(x,y)$ at time $t = 15$ min from start of the expiration event (the source located at $x = 2.5$m, $y = 2.5$m) (a) $\rho = 10^9$ copies/ml, $ACH = 2\,\mathrm{h}^{-1}$, (b) $\rho = 5 \times 10^9$ copies/ml, $ACH = 2\,\mathrm{h}^{-1}$, (c) $\rho = 10^9$ copies/ml, $ACH = 5\,\mathrm{h}^{-1}$, (d) $\rho = 5 \times 10^9$ copies/ml, $ACH = 5\,\mathrm{h}^{-1}$, (e) $\rho = 10^9$ copies/ml, $ACH = 12\,\mathrm{h}^{-1}$, and (f) $\rho = 5 \times 10^9$ copies/ml, $ACH = 12\,\mathrm{h}^{-1}$.

This can be attributed to the inherently higher ACH inside the volume of interest but primarily due to the absence of confinement, which allows the virus concentration to freely decay with space.

B. 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 $N_s = 103\,679$. 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, volume of ejected respiratory liquid per unit time, and source location (x_0, y_0), are randomized. ACH is generated from a log-normal distribution with a median $ACH = 2.16\,\mathrm{h}^{-1}$, such that $\mu_{ACH} = 0.7701$, $\sigma_{ACH} = 0.7554$ based on measurements by Bohanon et al.[59] The room height $H = 3$ m, the height of the source (seated) $z_0 = 1$ m, indoor conditions ($T = 21.7^\circ$C, $RH = 0.50$, $UV_{index} = 0$), and dose response constant $r_v = 1/1440$ [see Eq. (11)] are held constant. The resulting distribution of the number of secondary infections—pdf of $\mathcal{Z}$ is presented in Fig. 6. An analytical solution $g(\mathcal{Z})$ to be derived in Sec. IV C is also shown in the figure.

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 $mean(Z)_s = \langle Z \rangle_s = 0.14$ (where the subscript s represents the simulated result as opposed to an analytical result) obtained from the simulated pdf, indicates that over an exposure time of nearly an hour, on average less than one person got infected, per infector. The calculated total number of infections is 12 648 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 Lednicky et al.[60] 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 $\mathcal{Z}$-pdf yields a good fit with dispersion parameter $k = 0.03$. While the fit quality worsens at the pdf tails, the dispersion parameter is in the same
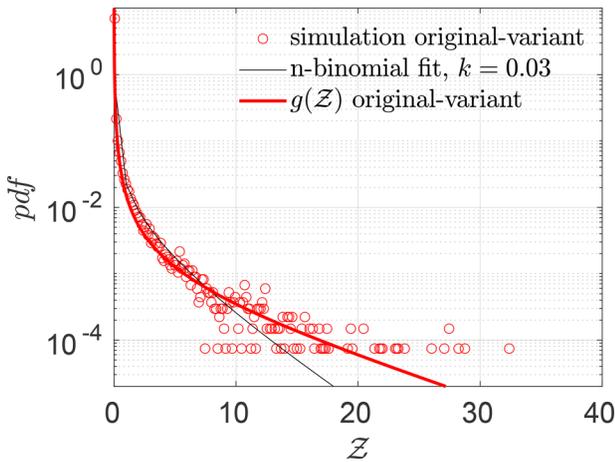
FIG. 6. Pdf of the number of secondary infections: $\mathcal{Z}$ and negative binomial fit. The analytical function g is given in Eq. (21). $g($ median viral load $\mu = 13.84$, $\sigma = 3.63, \alpha = 1.113 \times 10^{-10}$, Average occupancy $\nu = 7.71$). See Eqs. (14) and (17) for σ and α, respectively. Average exposure time $\langle\tau\rangle = 3914$ s, average speaking time $\langle t_s \rangle = 1469$ s, and average air change rate, $\langle ACH \rangle = 2.87$ h^{-1}.

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 h 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 7(a) shows the joint probability density function (jpdf) of ρ and $\mathcal{Z}$.

The close correlation of the two random 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 viral load is causing a similar variation in the number of secondary infections with other parameters controlling the slope and strength of the correlation.

We present the joint probability density function (jpdf) of ACH and Z in Fig. 7(b). It is apparent that the highest number of infections Z occurs 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.

C. Analytical pdf of the number of secondary infections $\mathcal{Z}$: $g(\mathcal{Z})$

What leads to the stretched tail pdf of the number of secondary infections? Can we model it from first principles? These questions are taken up in this subsection. First, we note that the viral load pdf from Ref. 57, obtained for asymptomatic individuals (including presymptomatics), 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 by Yang et al. and also shown in Fig. 3(a). As such, we used the following pdf of viral load $f(\rho_v)$ for generating inputs into the simulation:

$$f(\rho_v) = \frac{1}{\rho_v \sigma \sqrt{2\pi}} e^{-(ln(\rho_v)-\mu)^2/2\sigma^2}. \quad (14)$$

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 Z, as apparent from Fig. 7(a). 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. $\tilde{\mathcal{Z}}$ is the sample space variable corresponding to $\tilde{Z}$. Variation of $\tilde{Z}$ with respect to ρ is shown in Fig. 8(a). Clearly, this plot reflects the dose-response function [Eq. (11)] on a macro scale given the dominant influence of ρ in controlling the probability of infection and eventually secondary attack rates. Therefore, we propose a function similar to the dose-response function to model the response of $\tilde{Z}$ to ρ variation. This is shown below

$$\tilde{Z} = 1 - e^{-\alpha\rho}. \quad (15)$$

We can also write

$$\rho = -\frac{1}{\alpha} ln(1 - \tilde{Z}). \quad (16)$$
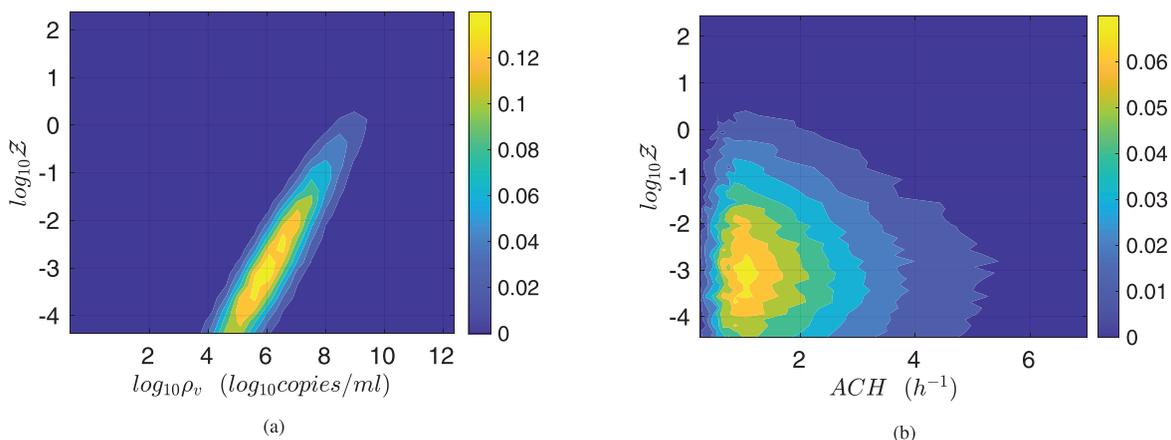
FIG. 7. Joint pdf of (a) the viral load and number of secondary infections and (b) the ACH and number of secondary infections.

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 an infection probability of $1 - e^{-1} = 0.63$. Utilizing mean values of individual input distributions: average room volume $\langle V \rangle = 6.34 \times 10^2$ m^3, average speaking time $\langle t_s \rangle = 1469$ s, average exposure time $\langle \tau \rangle = 3914$ s, average air change rate, $\langle ACH \rangle = 2.87$ h^{-1}, and deposition parameter based on the average room area and volume $\beta_0 = 0.002$ s^{-1}, α can be estimated as

$$\alpha = \frac{r_v \langle \dot{Q}_l \rangle \langle t_s \rangle \dot{V}_b}{\langle V \rangle} \int_0^{\langle \tau \rangle} \psi(t) e^{-(\langle ACH \rangle / 3600 + \beta_0)t} dt. \tag{17}$$

The virus survivability function $\psi(t)$ is given in Eq. (8), and the average volume flow of the exhaled liquid per unit time is $\langle \dot{Q}_l \rangle = 2.222 \times 10^{-6}$ ml/s. Note that all the $\langle \rangle$ quantities mention averages over the distributions used in the present simulation. Equation (17) yields $\alpha = 1.113 \times 10^{-10}$ ml/copies. The comparison of Eq. (15) and the simulation results are shown in Fig. 8(a).

With the functional form of the pdf of the viral load known [given in Eq. (14)], we can immediately substitute Eq. (16) into Eq. (14) to eventually derive the pdf of $\tilde{Z}$ using the generalized equation below

$$\phi(\tilde{Z}) = f\left(-\frac{1}{\alpha} ln(1 - \tilde{Z})\right) \frac{d\left(-\frac{1}{\alpha} ln(1 - \tilde{Z})\right)}{d\tilde{Z}}. \tag{18}$$

Using the log-normal form of f, we derive the analytical function below, which could be used to model the pdf of the 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 Ref. 10 instead of log-normal

$$\phi(\tilde{Z}) = \frac{1}{(-|1 - \tilde{Z}|ln(1 - \tilde{Z}))\sigma\sqrt{2\pi}} e^{-\left\{ln\left(-\frac{1}{\alpha}ln(1-\tilde{Z})\right)-\mu\right\}^2 / 2\sigma^2}. \tag{19}$$

Comparison of Eq. (19) with the simulation results is shown in Fig. 8(b).

It is evident that Eq. (19) describes the simulation data to good quantitative accuracy. It is also remarkable that very important effects of area, ACH, virus kinetics, and 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 $0 \le \tilde{Z} < 1$ with high degree of veracity. Now, we can write $Z = N(1 - e^{-\alpha\rho})$ using Eq. (15), where N is the sample space variable corresponding to n. 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 Z/N given by $\phi(Z/N)$ 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 Z given by

$$g(Z) = \int_0^\infty h(N)\phi(Z/N)\frac{1}{N} dN. \tag{20}$$

Using the exponential distribution for occupancy at different points of interest: POIs (restaurants in our case) $h(N, \nu) = \frac{1}{\nu} e^{-\frac{N}{\nu}}$ shown in Fig. 3(b), we derive for $Z < N$

$$g(Z) = \int_0^\infty \frac{e^{-\frac{N}{\nu} - \frac{\left\{ln\left(-\frac{1}{\alpha}ln(1-Z/N)\right)-\mu\right\}^2}{2\sigma^2}}}{-\nu\sigma\sqrt{2\pi}(N-Z)ln(1-Z/N)} dN. \tag{21}$$

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 Eq. (17). It is apparent that the pdf $g(Z)$ is stretched to higher (lower) 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
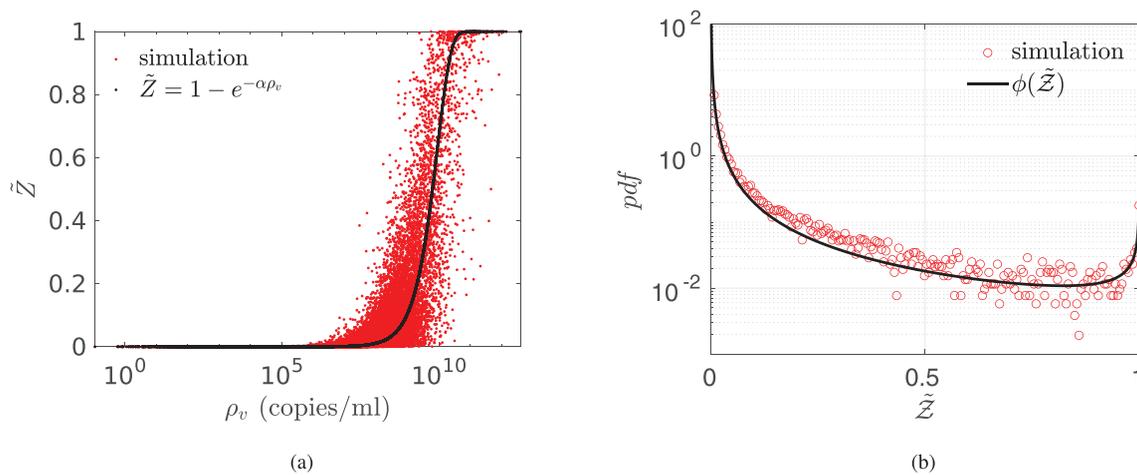
FIG. 8. (a) Scatter plot of the viral load (ρ) vs secondary attack rate ($\tilde{Z}$). The black dot curve shows Eq. (15) with $\alpha = 1.113 \times 10^{-10}$ ml/copies and (b) pdf of $\tilde{Z}$ and its comparison with the analytical result given in Eq. (19).
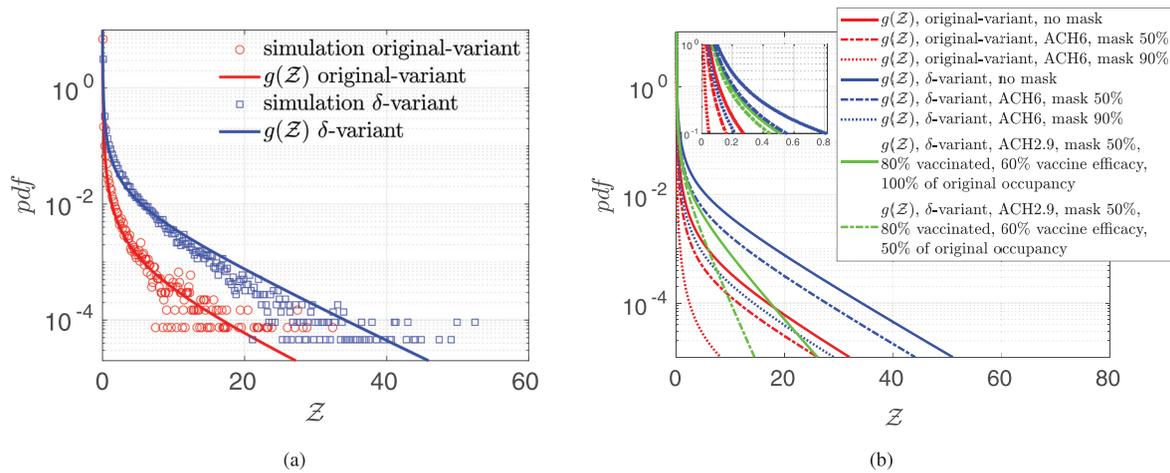
FIG. 9. (a) Pdf of $\mathcal{Z}$ and its comparison with the analytical result given in Eq. (21) for original variant (red circle symbols, red bold line) $g(\mu = 13.84, \sigma = 3.63, \alpha = 1.113 \times 10^{-10}, \nu = 7.71)$ and for δ−variant (blue square symbols, blue bold line) $g(\mu = 17.97, \sigma = 3.63, \alpha = 1.113 \times 10^{-10}, \nu = 7.71)$ and (b) effect of masking, fixed ventilation rates, vaccines, and reduced occupancy. Solid red and blue lines correspond to the analytical solutions given in Eq. (21) for the original and δ−variant, respectively, for $\langle ACH \rangle = 2.9\,h^{-1}$ and without masks, as shown in Fig. 9(a). For fixed $ACH = 6\,h^{-1}$, and with masks blocking 50% volume of aerosols during inhalation and exhalation, the dashed red line shows the analytical solution for original variant $g(\mu = 13.84, \sigma = 3.63, \alpha = 2.187 \times 10^{-11}, \nu = 7.71)$ while the dashed blue line shows the analytical result for δ−variant $g(\mu = 17.97, \sigma = 3.63, \alpha = 2.187 \times 10^{-11}, \nu = 7.71)$. For fixed $ACH = 6\,h^{-1}$, and with masks blocking 90 % volume of aerosols during inhalation and exhalation, the dotted red line shows the analytical solution for original variant $g(\mu = 13.84, \sigma = 3.63, \alpha = 8.747 \times 10^{-13}, \nu = 7.71)$ while the dotted blue line shows the analytical result for δ−variant $g(\mu = 17.97, \sigma = 3.63, \alpha = 8.747 \times 10^{-13}, \nu = 7.71)$. The solid green line: $g(\mu = 17.97, \sigma = 3.63, \alpha = 2.78 \times 10^{-11}, \nu_v = 4.01)$ 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 dashed-dotted green line $g(\mu = 17.97, \sigma = 3.63, \alpha = 2.78 \times 10^{-11}, \nu_v = 2.00)$ 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 II in the Appendix for detailed values.

has been shown in Fig. 6. 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 functional of $\mathcal{Z}$ and N. This can be verified by noting that when either $\sigma \to 0$ or $\nu \to 0$, the overdispersion of the number of secondary infections vanishes. The first stretched exponential arising from the log-normal distribution of viral load ρ_v and the latter from the exponential distribution of the number of people at the different POIs. A more rigorous way to show this is to replace exponential distribution describing $h(N)$ in Eq. (20) with a Dirac delta function $\delta(N - N_0)$ and using its sifting property to get

$$g_D(\mathcal{Z}, N_0) = \frac{e^{-\left\{ ln\left(-\frac{1}{\alpha}ln(1-\mathcal{Z}/N_0)\right)-\mu\right\}^2 / 2\sigma^2}}{-\sigma\sqrt{2\pi}(N_0 - \mathcal{Z})ln(1 - \mathcal{Z}/N_0)}, \quad (22)$$

where unless $N_0 \gg \nu$, the overdispersion in Z is greatly diminished with respect to the exponential distribution of N. While the near homogeneous population distribution could be representative of classrooms in elementary schools in a city, indoor occupancy of various social settings is expected to be overdispersed as in the case of restaurants. Therefore, in general, 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 is given by $\langle Z \rangle = \int_0^\infty \mathcal{Z} g(\mathcal{Z}) d\mathcal{Z}$. We find mean and standard deviation as $\langle Z \rangle = 0.12$, $std(Z) = 0.94$, respectively, in comparison to $\langle Z \rangle_s = 0.14$ and $std(Z)_s = 0.76$ 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 social-contact settings without much restriction on their type.

D. Variants and mitigation measures

Finally, we test whether the derived Eq. (21) can describe overdispersed transmission associated with a different viral load distribution equally well. To this end, we generate a viral load distribution for the δ−variant. Jüni et al.[61] have provided us with a representative curve for δ−variant viral load vs time based on the data from Ref. 62. Using this curve, the time-averaged median of the viral load log-normal distribution was obtained, and hence the parameter μ corresponding to the distribution. Thus, while Fig. 3(a) represents the pdf of the original variant, the new log-normal distribution characterized by $\mu = 17.97$ and $\sigma = 3.63$ might represent the pdf of viral load of the δ−variant infected individuals. The simulation results in terms of $\mathcal{Z}$-pdf are presented in Fig. 9(a). The greater transmissibility of the δ−variant due to the higher mean viral load is immediately apparent. In comparison to the $\langle Z \rangle = 0.12$ for the original variant, the $\langle Z \rangle = 0.98$ for δ (from simulations $\langle Z \rangle_s = 0.95$ for δ). Therefore, just based on viral load, according to the calculations and model, the δ−variant could be nearly $8.2\times$ higher transmissible on average with respect to the original variant, over about an hour of contact. Interestingly, with just increased μ, Eq. (21) 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 viral load and infectiousness potential (using a proxy of

culture-positivity, for example), while monotonic in nature, may not demonstrate a linear relationship.[63] Furthermore, $\langle Z \rangle \neq \mathcal{R}_0$. Hence, the enhancement factor, thus, found is valid only within the context of the assumptions.

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 $ACH = 6\,\mathrm{h}^{-1}$ and assume that the entire population (including the infectors and susceptibles) is 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. 9(b).

We observe that such interventions result in significant reduction in transmissibility for the original variant; reducing the mean to $\langle Z \rangle = 0.05$ (from $\langle Z \rangle = 0.12$ obtained without any interventions) when the masks are blocking 50% of aerosols during both inhalation and exhalation events with significant reduction in the extension of the tail. Such intervention effects remain substantial for the δ-variant too, where $\langle Z \rangle = 0.47$ (with respect to $\langle Z \rangle = 0.98$ obtained without any interventions) though tail remains sufficiently stretched with some shift in the overall pdf toward lower $\mathcal{Z}$. An even severe reduction in transmissibility is observed when the masks are set to block 90% of aerosols traveling through it, giving us $\langle Z \rangle = 0.004$ for the original variant and $\langle Z \rangle = 0.078$ for the δ-variant. 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 (median viral load) 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 Appendix.

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. 9(b) 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 the viral load or change in distribution of infectious cases. Fitting an exponential distribution to n_v, the new constant: mean occupancy of susceptibles $\nu_v = 4.01$. Clearly, from Fig. 9(b), 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.

V. CONCLUSIONS

Understanding the mechanistic factors that lead to overdispersion in secondary cases has been considered long-standing scientific problems. We coupled an aerosol mixing model with real-world epidemiological and viral-biology inputs: exhaled aerosol size distribution for speech and breath, measured viral load distribution, and realistic ventilation rate distributions with occupancy information from more than hundred thousand social contact settings in major US cities to explore overdispersion in the number of secondary infections per infector. The simulated results demonstrate that the aerosol transmission route is consistent with the overdispersed individual infectivity with viral load variability being the dominant factor that controls secondary attack rates alongside the ventilation rate, exposure time, and speaking time. We also derived, for the first time, analytical expressions that accurately described the simulated pdfs of the secondary attack rates and the number of secondary infections, respectively, and further elucidated how the quantitative relationship between variability in individual-level viral load and event-level occupancy jointly control the overdispersion. Our findings suggest that even in the context of airborne transmission, approximately 4% of index cases in indoor settings would account for 80% of secondary cases, thus highlighting the importance of understanding and focusing mitigation efforts on drivers of superspreading events. The findings highlight the importance of measures to decrease exposures during periods of high viral shedding (such as via isolation by timely testing to detect periods of high viral shedding), as well as improving ventilation, and the increased probability of outbreaks with variants of concern associated with higher viral loads. Finally, the analytical function further offers an opportunity to estimate the spatially defined probability of outbreaks and outbreak size from point-source transmission events given viral load and occupancy across indoor settings.

AUTHOR DECLARATIONS
Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: COMPARISON WITH EXPERIMENTAL AND COMPUTATIONAL RESULTS

Hathway et al.[50] carried out experiments and CFD simulations in a $3.35 \times 4.26 \times 2.26$ m^3 room with ventilation. Though the model is validated with spatial CO measurements[12] to add another

layer of validation, we compared our model with results from Ref. 50 for ACH values of 6 and $12\,h^{-1}$ with a bioaerosol ejector placed at the center of the room. Results are compared along three lines at a height of 1.15 m. Our model was normalized using the average value over all three lines so that they can be compared with the normalized experimental and computational results provided in the paper. Hathway *et al.*[50] have attributed the large variance in certain experimental measurements as shown in Fig. 10 (points with large error bars) to loss of viability during nebulization and sampling. In spite of that, comparisons in Fig. 10 show qualitative and quantitative agreement at most experimental sampling points, and for the points with large variance, our results fall within the given range. As for comparison with the CFD results, the very high peaks at the center of the $Z = 1.67$ m line do not appear in our model, which is purely an artifact of the difference in the turbulence diffusivity model involved in the two studies.

APPENDIX B: PARTICLE SIZE DISTRIBUTION

From Ref. 36, we have the number size distribution as

$$\psi = \sum_{i=1}^{n} A_i \exp\left(-\left[\frac{\ln\left(D_{s,0}/D_i\right)}{\sigma_i}\right]^2\right), \tag{B1}$$

where they have defined $\psi = d\eta/dlog_{10}D_{s,0}$ (η is the number concentration of particles). D_i is the mode mean geometric diameter, σ_i is the modal geometric standard deviation, and A_i is the number

concentration at D_i. These constants are mode specific parameters, which have prescribed values for each of the ith modes (refer to Table VI in Ref. 36 for the values of the constants). For our purpose, we require the function $\psi_{D,\eta} = d\eta/dD_{s,0}$

$$\psi = \frac{d\eta}{dlog_{10}D_{s,0}} = \sum_{i=1}^{n} A_i \exp\left(-\left[\frac{\ln\left(D_{s,0}/D_i\right)}{\sigma_i}\right]^2\right), \tag{B2}$$

$$\psi_{D,\eta}(D_{s,0}) = log_{10}(e) \sum_{i=1}^{n} \frac{A_i}{D_{s,0}} \exp\left(-\left[\frac{\ln\left(D_{s,0}/D_i\right)}{\sigma_i}\right]^2\right). \tag{B3}$$

It is to be noted that A_i is a number concentration parameter, which means ψ would correspond to number of particles per unit volume. Hence, to obtain the distribution for the total number of particles ejected per unit time, we have to multiply Eq. (B3) by the total volume exhaled per unit time $\dot{V}$. The $\dot{V}$ values are taken as $360\,Lh^{-1}$ and $700\,Lh^{-1}$ for breathing and speaking, respectively, both of which fall within the typical range for $\dot{V}$ as mentioned in Table I of Ref. 36. The number size distribution can also be converted to the volume size distribution through the relation

$$\psi_{D,V} = \frac{dv_\eta}{dD_{s,0}} = \frac{\pi}{6}D_{s,0}{}^3\psi_{D,\eta}(D_{s,0}), \tag{B4}$$

$$\psi_{D,V} = log_{10}(e)\frac{\pi}{6}\sum_{i=1}^{n} A_i D_{s,0}{}^2 \exp\left(-\left[\frac{\ln\left(D_{s,0}/D_i\right)}{\sigma_i}\right]^2\right), \tag{B5}$$

where v_η is the volume concentration of particles.
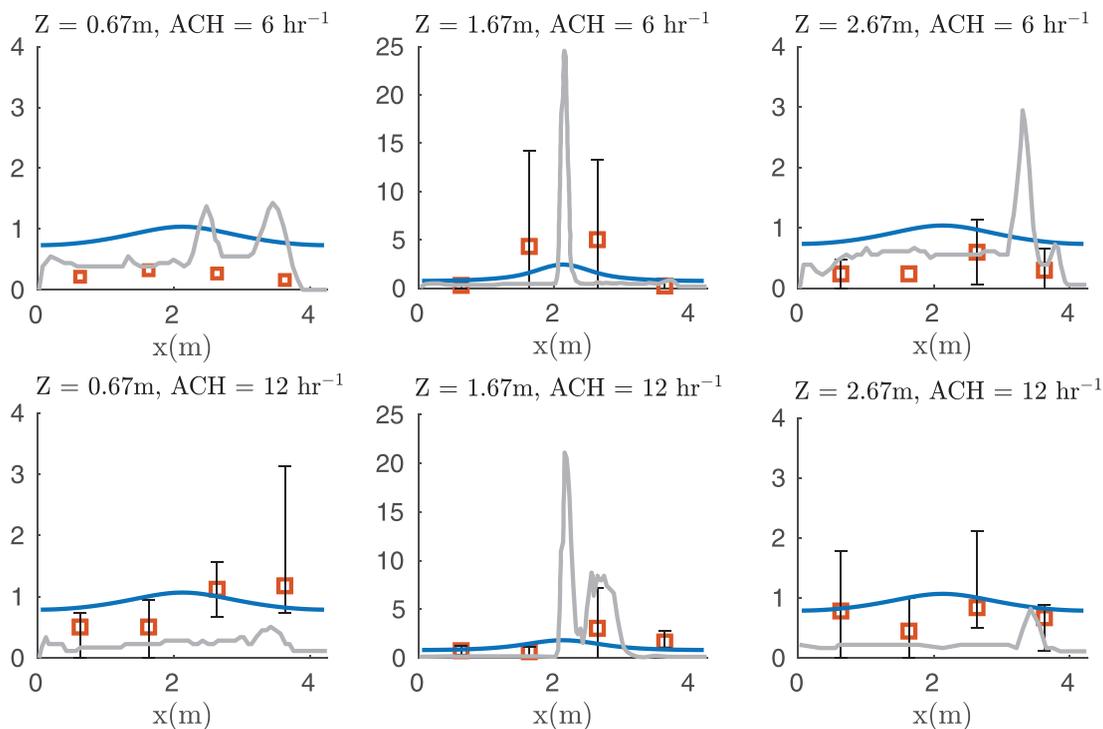
FIG. 10. Normalized concentrations from simulations along three different lines obtained with the model used in this paper (shown using blue solid line) compared to experimental data from Ref. 50 (shown using square markers along with error bars) and CFD results from the same paper[50] (shown using gray solid line) for ACH values of $6\,h^{-1}$ (top row) and $12\,h^{-1}$ (bottom row) using a central continuous point source of aerosols.

APPENDIX C: VOLUME FLOW RATE OF RESPIRATORY LIQUID

$\dot{Q}_l$ is obtained from integrating the volume size distribution over the entire range of particle size diameters and multiplying it with the volume flow rate of air per unit time. It is reasonable to expect this quantity to vary from one infector to another as we move from one POI to another. To account for this change, we assume a truncated normal distribution for the speaking $\dot{Q}_{ls}$ with the mean value obtained from VSD of Ref. 36 as $\langle \dot{Q}_{ls} \rangle = 2.222 \times 10^{-6}$ ml/s. The standard deviation value was calculated as $\sigma_{\dot{Q}_{ls}} = 3.078 \times 10^{-6}$ ml/s using similar particle size distribution plots provided by Johnson et al.[66] using their BLO model for their own data along with the data from Refs. 67–69. It is to be noted that if we leave the normal distribution with its limits extending indefinitely on both sides of the mean then it would encompass extremely large and small values for $\dot{Q}_{ls}$, which would be unrealisitic. Hence, we created a truncated normal distribution with the value of Johnson et al.[66] ($\dot{Q}_{ls} = 1.292 \times 10^{-8}$ ml/s) put as the lower limit whereas an equal distance (as between Johnson et al.'s lower limit and Pöhlker et al.'s mean value) is taken toward the upper limit ($\dot{Q}_{ls} = 4.433 \times 10^{-6}$ ml/s) from the mean. It is ensured that the Duguid et al.[67] and Loudon et al.[68,69] $\dot{Q}_{ls}$ values fall within that range. The contribution of breathing toward $\dot{Q}_{ls}$ is several orders of magnitude lower than that of the lower limit of speaking, and as such no distribution is considered for breathing and instead its contribution $\dot{Q}_{lb}$ is added to the randomly sampled $\dot{Q}_{ls}$ value.

APPENDIX D: CHARACTERISTIC DEPOSITION VELOCITY

To obtain a unique characteristic deposition velocity over a surface for the entire range of particle size, we first require a functional form for the deposition velocity. Lai and Nazaroff[64] developed a model for said deposition velocity on indoor surfaces, which took into account the effects of Brownian diffusion, turbulent diffusion, and gravitational settling. They provided a formulation for a quantity β (wall deposition coefficient) as a function of the particle diameter. This function can then be put through an averaging operation to obtain a value for the average characteristic deposition velocity w_d as

$$\beta_{avg} = \frac{w_d A}{V} = \frac{(v_{dv})_{avg} A_v + (v_{du})_{avg} A_u + (v_{dd})_{avg} A_d}{V}. \quad \text{(D1)}$$

Here, A_v, A_u, and A_d are the surface areas of the vertical surfaces, upward-facing horizontal surfaces, and downward-facing horizontal surfaces, respectively. The corresponding characteristic deposition velocities are given by v_{dv}, v_{du}, and v_{dd}. Lai and Nazaroff[64] wrote the wall deposition coefficient β as a function of the deposition velocities and the room geometry as

$$\beta = \frac{v_{dv} A_v + v_{du} A_u + v_{dd} A_d}{V}. \quad \text{(D2)}$$

The deposition velocities can be expressed in terms of the friction velocity u^* and gravitational settling velocity u_s as

$$v_{dv} = \frac{u^*}{I}, \quad v_{du} = \frac{u_s}{1 - \exp\left(-\dfrac{u_s I}{u^*}\right)}, \quad v_{dd} = \frac{u_s}{\exp\left(\dfrac{u_s I}{u^*}\right) - 1}. \quad \text{(D3)}$$

I is an integrated quantity (refer to Ref. 64 for detailed formulation) that depends on the Schmidt number $Sc = \nu/D$, where ν is the kinematic viscosity and D is the Brownian diffusivity. Brownian diffusivity can be directly obtained[65] from the particle diameter at equilibrium $D_s = D_{s,0}/5$, as

$$D = \frac{k_B T C_c}{3\pi\mu D_s}. \quad \text{(D4)}$$

Here, k_B is the Boltzmann constant, T is the absolute temperature in kelvin, μ is the dynamic viscosity of air, and C_c is a slip correction factor for small particles. For the remaining two unknowns, the friction velocity and the gravitational settling velocity, the former is an input variable while the later can be expressed as[65]

$$u_s = \frac{\rho D_s^2 g C_c}{18\mu}, \quad \text{(D5)}$$

where ρ is the particle density. From Eqs. (D2)–(D5), an average wall deposition velocity on ith surface (where $i = v, u, d$), $(v_{di})_{avg}$ can be obtained by treating this quantity as the expected value of a function of a random variable (D_s)

$$(v_{di})_{avg} = \int_{-\infty}^{\infty} v_{di}(D_s)\psi_{D,V}(D_s)dD_s. \quad \text{(D6)}$$

It is to be noted that this is a volume averaging operation due to $\psi_{D,V}$ being defined as a volume size distribution similar to Eq. (B5) but by using the equilibrium diameter D_s instead. This gives us a relation for β_{avg} that can in turn be related to the characteristics deposition velocity w_d as

$$\beta_{avg} = \frac{w_d A}{V} = \frac{(v_{dv})_{avg} A_v + (v_{du})_{avg} A_u + (v_{dd})_{avg} A_d}{V}. \quad \text{(D7)}$$

APPENDIX E: EXPOSURE TIME, SPEAKING TIME, AND ACH

Exposure time τ variation was modeled with a log-normal distribution such that median $\tau = 1$ h, ($\mu_\tau = 0$, $\sigma_\tau = 0.40$), and speech time t_s was modeled using a uniform distribution. In particular, $0.25\tau \le t_s \le 0.5\tau$. These yield the following averages: $\langle \tau \rangle = 1.0875$ h and $\langle t_s \rangle = 0.409$ h. From the literature, it is found that ventilation rates in indoor built environments are typically log-normally distributed. In particular, Bohanon et al.[59] measured the ventilation rate in 33 restaurants and found them to be log-normally distributed with a median 1.8 l/sm^2 and standard deviation of 2.1 l/sm^2. Following this measured distribution, for $H = 3$ m, we found median $ACH = 2.16$ h^{-1} and used $\mu_{ACH} = 0.7701$, $\sigma_{ACH} = 0.7554$ for the ACH pdf.

APPENDIX F: SENSITIVITY OF MEAN AND STANDARD DEVIATION OF THE NUMBER OF SECONDARY INFECTIONS

In this sub-section, we explore the sensitivity of the most important parameters: $\langle ACH \rangle$, $\langle \rho \rangle$, and $\langle N \rangle$. To that end, we explore the effect of variation one parameter at a time on $\langle Z \rangle$ and standard deviation (Z), keeping the other two constant. The results are shown in Table II and Figs. 11(a)–11(c). All three parameters affect $\langle Z \rangle$ and standard deviation (Z). However, the effect of variation of $\langle \rho \rangle$ is strongest on both the mean and standard deviation of Z, followed by that of $\langle N \rangle$, and $\langle ACH \rangle$.

TABLE II. Sensitivity of secondary infection number statistics.

$\langle ACH \rangle$	α	μ	Σ	$\langle \rho \rangle$	$\langle N \rangle$	$\langle Z \rangle$	Std (Z)
0.1	1.467×10^{-10}	13.839	3.643	7.796×10^{8}	7.707	0.139	1.022
1	1.329×10^{-10}	13.839	3.643	7.796×10^{8}	7.707	0.131	0.987
5	9.391×10^{-11}	13.839	3.643	7.796×10^{8}	7.707	0.106	0.881
10	6.864×10^{-11}	13.839	3.643	7.796×10^{8}	7.707	0.089	0.808
50	2.158×10^{-11}	13.839	3.643	7.796×10^{8}	7.707	0.046	0.589
100	1.149×10^{-11}	13.839	3.643	7.796×10^{8}	7.707	0.033	0.508
2	1.205×10^{-10}	12.672	3.643	2.427×10^{8}	7.707	0.065	0.697
2	1.205×10^{-10}	14.672	3.643	1.793×10^{9}	7.707	0.199	1.255
2	1.205×10^{-10}	16.672	3.643	1.325×10^{10}	7.707	0.571	2.268
2	1.205×10^{-10}	18.672	3.643	9.789×10^{10}	7.707	1.310	3.546
2	1.205×10^{-10}	20.672	3.643	7.234×10^{11}	7.707	2.599	5.027
2	1.205×10^{-10}	21.672	3.643	1.966×10^{12}	7.707	3.289	5.594
2	1.205×10^{-10}	13.839	3.643	7.796×10^{8}	1.707	0.029	0.226
2	1.205×10^{-10}	13.839	3.643	7.796×10^{8}	2.707	0.045	0.349
2	1.205×10^{-10}	13.839	3.643	7.796×10^{8}	4.707	0.076	0.589
2	1.205×10^{-10}	13.839	3.643	7.796×10^{8}	6.707	0.108	0.839
2	1.205×10^{-10}	13.839	3.643	7.796×10^{8}	8.707	0.139	1.077
2	1.205×10^{-10}	13.839	3.643	7.796×10^{8}	10.707	0.171	1.324
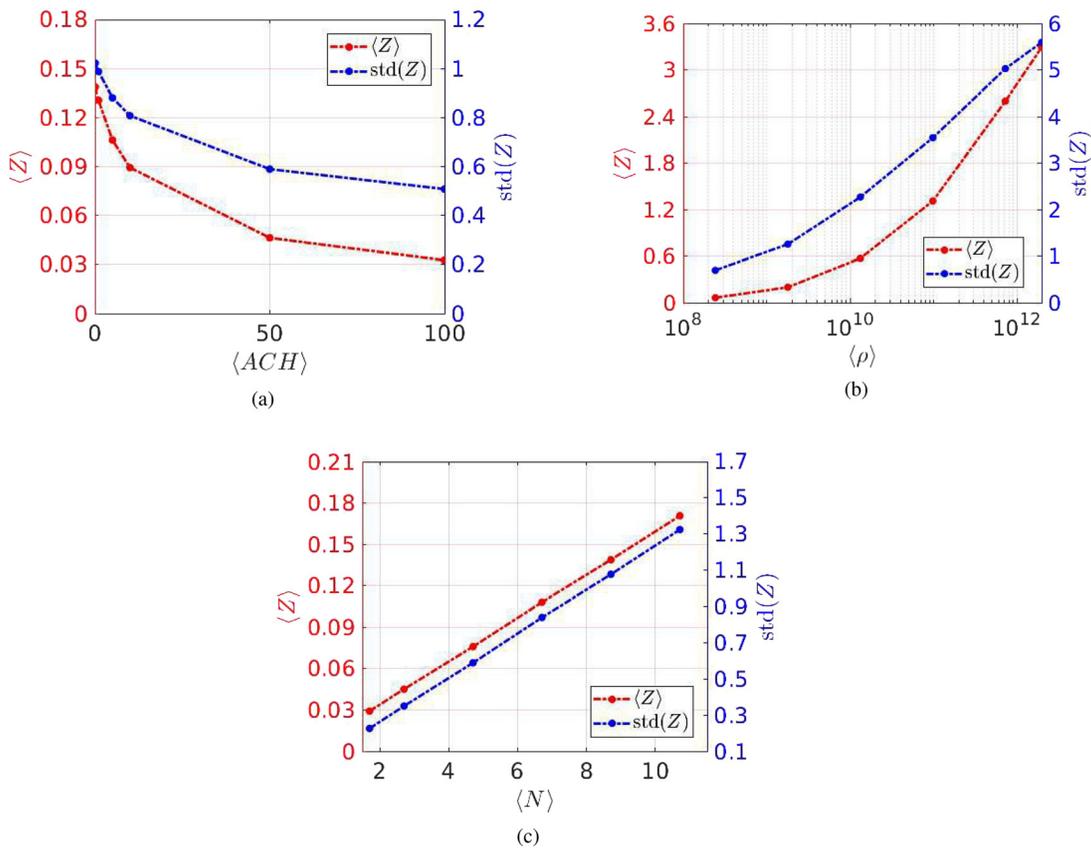
FIG. 11. Variation of the mean $\langle Z \rangle$ and standard deviation *std* (Z) with (a) mean air-change rate, (b) mean viral load, and (c) mean occupancy.

26 January 2024 00:00:59

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 **438**, 355–359 (2005).

[2]L. Hamner, "High SARS-CoV-2 attack rate following exposure at a choir practice–Skagit county, Washington, March 2020," MMWR Morb. Mortal. Wkly. Rep. **69**, 606 (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 **31**, 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," Emerg. Infect. Dis. **26**, 1628 (2020).

[5]Y. Li, H. Qian, J. Hang, X. Chen, P. Cheng, H. Ling, S. Wang, P. Liang, J. Li, S. Xiao, J. Wei, L. Liu, B. J. Cowling, and M. Kang, "Probable airborne transmission of SARS-CoV-2 in a poorly ventilated restaurant," Build. Environ. **196**, 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," Emerg. Infect. Dis. **26**, 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 **584**, 420–424 (2020).

[8]A. Endo et al., "Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China," Wellcome Open Res. **5**, 67 (2020).

[9]G. Großmann, M. Backenköhler, and V. Wolf, "Heterogeneity matters: Contact structure and individual variation shape epidemic dynamics," PLoS One **16**, 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 **10**, 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," Proc. Natl. Acad. Sci. U. S. A. **118**, e2016623118 (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," Lancet Infect. Dis. **21**, 1203 (2021).

[13]L. Morawska and D. K. Milton, "It is time to address airborne transmission of COVID-19," Clin. Infect. Dis. **71**, 2311–2313 (2020).

[14]J. G. Allen and L. C. Marr, "Recognizing and controlling airborne transmission of SARS-CoV-2 in indoor environments," Indoor Air **30**, 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," J. Phys. Chem. B **123**, 266–276 (2019).

[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)," J. Hosp. Infect. **110**, 89 (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," Lancet **397**, 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 **370**, 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," Phys. Fluids **32**, 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 Sci. Technol. **55**, 142–153 (2021).

[21]L. Bourouiba, E. Dehandschoewercker, and J. W. Bush, "Violent expiratory events: On coughing and sneezing," J. Fluid Mech. **745**, 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," Proc. Natl. Acad. Sci. U. S. A. **117**, 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," Build. Environ. **176**, 106859 (2020).

[24]E. Riley, G. Murphy, and R. Riley, "Airborne spread of measles in a suburban elementary school," Am. J. Epidemiol. **107**, 421–432 (1978).

[25]S. Rudnick and D. Milton, "Risk of indoor airborne infection transmission estimated from carbon dioxide concentration," Indoor Air **13**, 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," Environ. Int. **145**, 106112 (2020).

[27]M. Z. Bazant and J. W. Bush, "A guideline to limit indoor airborne transmission of COVID-19," Proc. Natl. Acad. Sci. U. S. A. **118**, e2018995118 (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," Environ. Health Perspect. **129**, 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," Environ. Sci. Technol. **55**, 2890–2898 (2021).

[30]T. Dbouk and D. Drikakis, "On coughing and airborne droplet transmission to humans," Phys. Fluids **32**, 053310 (2020).

[31]T. Dbouk and D. Drikakis, "Fluid dynamics and epidemiology: Seasonality and transmission dynamics," Phys. Fluids **33**, 021901 (2021).

[32]T. Dbouk and D. Drikakis, "Weather impact on airborne coronavirus survival," Phys. Fluids **32**, 093312 (2020).

[33]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 **373**, eabd9149 (2021).

[34]R. Mittal, R. Ni, and J.-H. Seo, "The flow physics of COVID-19," J. Fluid Mech. **894**, 2 (2020).

[35]L. Bourouiba, "The fluid dynamics of disease transmission," Annu. Rev. Fluid Mech. **53**, 473–508 (2021).

[36]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," preprint arXiv:2103.01188 (2021).

[37]S. Chaudhuri, A. Saha, and S. Basu, "An opinion on the multi-scale nature of COVID-19 type disease spread," Curr. Opin. Colloid Interface Sci. **54**, 101462 (2021).

[38]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 **589**, 82–87 (2021).

[39]J. A. Backer, D. Eggink, S. P. Andeweg, I. K. Veldhuijzen, N. van Maarseveen, K. Vermaas, B. Vlaemynck, R. Schepers, S. van den Hof, C. Reusken et al., "Shorter serial intervals in SARS-CoV-2 cases with omicron variant compared to delta variant in the Netherlands, 13–19 December 2021," medRxiv (2022).

[40]W. S. Hart, E. Miller, N. J. Andrews, P. Waight, P. K. Maini, S. Funk, and R. N. Thompson, "Generation time of the alpha and delta SARS-CoV-2 variants," medRxiv (2021).

[41]C. N. HAAS, "Estimation of risk due to low doses of microorganisms: A comparison of alternative methodologies," Am. J. Epidemiol. **118**, 573–582 (1983).

[42]S. B. Pope, Turbulent Flows (Cambridge University Press, 2006).

[43]P. J. Drivas, P. A. Valberg, B. L. Murphy, and R. Wilson, "Modeling indoor air exposure from short-term point source releases," Indoor Air **6**, 271–277 (1996).

[44]A. Venkatram and J. Weil, "Modeling turbulent transport of aerosols inside rooms using eddy diffusivity," Indoor Air **31**, 1886 (2021).

[45]S. Chaudhuri, S. Basu, P. Kabi, V. R. Unni, and A. Saha, "Modeling the role of respiratory droplets in COVID-19 type pandemics," Phys. Fluids **32**, 063309 (2020).

[46]M.-R. Pendar and J. C. Páscoa, "Numerical modeling of the distribution of virus carrying saliva droplets during sneeze and cough," Phys. Fluids **32**, 083305 (2020).

[47]J. Zheng, J. Wu, F. Fang, J. Li, Z. Wang, H. Xiao, J. Zhu, C. Pain, P. Linden, and B. Xiang, "Numerical study of COVID-19 spatial-temporal spreading in London," Phys. Fluids **33**, 046605 (2021).

26 January 2024 00:00:59

[48]B. Wang, H. Wu, and X.-F. Wan, "Transport and fate of human expiratory droplets-a modeling approach," Phys. Fluids **32**, 083307 (2020).

[49]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," Environ. Sci. Technol. **45**, 4016–4022 (2011).

[50]E. Hathway, C. Noakes, P. Sleigh, and L. Fletcher, "CFD simulation of airborne pathogen transport due to human activities," Build. Environ. **46**, 2500–2511 (2011).

[51]U. S. Department of Homeland Security, see https://www.dhs.gov/science-and-technology/sars-airborne-calculator for Estimated Airborne Decay of SARS-CoV-2. The link provides an online calculator where UV index, relative humidity and temperature can be provided as input and the duration for which the SARS-CoV-2 virus will remain stable under said conditions will be provided as output.

[52]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 **10**, e65902 (2021).

[53]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," J. R. Soc. Interface **16**, 20180298 (2019).

[54]C. N. Haas, "Action levels for SARS-CoV-2 in air: Preliminary approach," Risk Anal. **41**, 705–709 (2021).

[55]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 **15**, e0243597 (2020).

[56]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," Nat. Med. **26**, 672–675 (2020).

[57]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," Proc. Natl. Acad. Sci. U. S. A. **118**, e2104547118 (2021).

[58]R. McClain, https://bizfluent.com/info-12010139-much-room-need-restaurant.html for "How Much Room Do I Need For a Restaurant?" (2018).

[59]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," J. Expo. Sci. Environ. Epidemiol. **13**, 378–392 (2003).

[60]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," Int. J. Infect. Dis. **100**, 476–482 (2020).

[61]P. Jüni, S. Baert, P. Bobos et al., "Rapid antigen tests for voluntary screen testing," Sci. Briefs **2**, 1–21 (2021).

[62]M. Kang, H. Xin, J. Yuan et al., "Transmission dynamics and epidemiological characteristics of delta variant infections in China," medRxiv (2021).

[63]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).

[64]A. C. K. Lai and W. W. Nazaroff, "Modeling indoor particle deposition from turbulent flow onto smooth surfaces," J. Aerosol Sci. **31**, 463–476 (2000).

[65]R. Kohli and K. Mittal, Developments in Surface Contamination and Cleaning-Particle Deposition, Control and Removal (Elsevier Inc., 2010).

[66]G. Johnson, L. Morawska, Z. Ristovski, M. Hargreaves, K. Mengersen, C. Chao, M. Wan, Y. Li, X. Xie, D. Katoshevski, and S. Corbett, "Modality of human expired aerosol size distributions," J. Aerosol Sci. **42**, 839–851 (2011).

[67]J. P. Duguid, "The size and the duration of air-carriage of respiratory droplets and droplet-nuclei," Epidemiol. Infect. **44**, 471–479 (1946).

[68]R. G. Loudon and R. M. Roberts, "Relation between the airborne diameters of respiratory droplets and the diameter of the stains left after recovery," Nature **213**, 95–96 (1967).

[69]R. G. Loudon and R. M. Roberts, "Droplet expulsion from the respiratory tract," Am. Rev. Respir. Dis. **95**, 435–442 (1967).

26 January 2024 00:00:59