Analysing the distribution of SARS-CoV-2 infections in schools: integrating model predictions with real world observations

Arnab Mukherjee[1], Sharmistha Mishra[2], Vijaya Kumar Murty[3,4], Swetaprovo Chaudhuri[1]

1 Institute for Aerospace Studies, University of Toronto, Toronto, Ontario, Canada
2 Department of Medicine, Division of Infectious Diseases, University of Toronto, Toronto, Ontario, Canada
3 The Fields Institute for Research in Mathematical Sciences, Toronto, Ontario, Canada
4 Department of Mathematics, University of Toronto, Toronto, Ontario, Canada

* arnab.mukherjee@utoronto.ca

Abstract

School closures were used as strategies to mitigate transmission in the COVID-19 pandemic. Understanding the nature of SARS-CoV-2 outbreaks and the distribution of infections in classrooms could help inform targeted or 'precision' preventive measures and outbreak management in schools, in response to future pandemics. In this work, we derive an analytical model of Probability Density Function (PDF) of SARS-CoV-2 secondary infections and compare the model with infection data from all public schools in Ontario, Canada between September-December, 2021. The model accounts for major sources of variability in airborne transmission like viral load and dose-response (i.e., the human body's response to pathogen exposure), air change rate, room dimension, and classroom occupancy. Comparisons between reported cases and the modeled PDF demonstrated the intrinsic overdispersed nature of the real-world and modeled distributions, but uncovered deviations stemming from an assumption of homogeneous spread within a classroom. The inclusion of near-field transmission effects resolved the discrepancy with improved quantitative agreement between the data and modeled distributions. This study provides a practical tool for predicting the size of outbreaks from one index infection, in closed spaces such as schools, and could be applied to inform more focused mitigation measures.

Author summary

At the start of the COVID-19 pandemic, there was huge uncertainty around the risks of SARS-CoV-2 spread in classrooms. In the absence of early predictions surrounding classroom risks, many jurisdictions across countries closed in-person education. There is great interest in adopting a more 'precision' approach to better inform future interventions in the context of airborne virus risks. For this purpose, we need tools that can predict the probability of the size of outbreaks within classrooms along with the impact of interventions including masks, better ventilation, and physical distancing by limiting the number of students per classroom. To this end, we have developed a robust but practical model that yields the probability of secondary infections stemming from index cases occurring within schools on a given day. During model development, the

major underlying physical and biological factors that dictate the disease transmission process, both at long-range and close-range, have been accounted for. This enables our model to modify its predictions for different scenarios - and possibly allows its use beyond schools. Finally, the model's predictive capability has been verified by comparing its outputs with publicly available data on SARS-CoV-2 diagnoses in Ontario public schools. To our knowledge, this is the first time an analytical model derived from mostly first principles describes real-world infection distributions, satisfactorily. The quantitative match between the theoretical prediction and real-world data offers the proposed model as a possible powerful tool for better-informed precision pandemic mitigation strategies in indoor environments like schools.

Introduction   1

The public health response to SARS-CoV-2 in 2019 demonstrated a critical and growing   2
need to anticipate the probability and size of outbreaks and the potential impact of   3
mitigation measures. Models that can, with reasonable fidelity, predict the secondary   4
infection distribution of cases in different scenarios, serve this critical need in pandemic   5
and outbreak preparedness. Such predictive models would need to capture   6
'superspreading', a fundamental property of respiratory virus transmission – particularly   7
in the context of aerosolized or airborne transmission. Superspreading events occur   8
when the presence of one index case leads to infection of several individuals in a very   9
short span of time. As a mean value description of outbreak size is incapable of   10
capturing such events, Lloyd-Smith et al. [1] demonstrated the effectiveness of   11
empirically fitted outbreak size distributions at modeling heterogeneity in disease spread   12
dynamics. But, to retain predictive capabilities, models cannot rely only on infection   13
data and instead would have to account for all the mechanistic parameters that   14
uniquely define each case of an outbreak. These input parameters and their effects are   15
rooted in physical, biological, and behavioral factors.   16

Recent studies detail the different aspects of SARS-CoV-2 transmission that need to   17
be accounted for. The need for modeling airborne transmission is emphasized by   18
Morawska et al. [2] and Allen et al. [3]. The review by Prather et al. [4] highlights the   19
ability of aerosols to remain airborne for an extended period of time. Chaudhuri et   20
al. [5,6] showed that the viral load is a major contributor in determining the large   21
variations in the number of secondary infections. These were complemented by Chen et   22
al's [7] observations of viral load heterogeneity being strongly tied to the overdispersed   23
nature of SARS-CoV-2 infections, along with Bhavnani et al's [8] findings of a direct   24
correlation between higher index case viral loads and rising secondary infection counts.   25
The alleviating effect of increased ventilation, incorporated through the air change rate   26
parameter, on infection spread was thoroughly detailed by Thornton et al. [9], while   27
Ricolfi et al. [10] studied this effect in schools and found that the inclusion of   28
mechanical ventilation as opposed to natural ventilation drastically reduces the   29
likelihood of infection. The importance of occupancy in infection spread was pointed   30
out by Bazant and Bush [11]. Pioneering works by Haas [12–14] put forward the   31
dose-response model which connects the virus properties to the human physiology and   32
its response to the infectious dose. Schijven et al. [15] utilized the dose-response model   33
to derive a risk assessment model for various expiratory events. Consolidating these   34
findings, Chaudhuri et al. [5] proposed a model for the Probability Density Function   35
(PDF) of secondary infections due to long-range airborne transmission, which relied on   36
observations based on certain numerical results generated from cell-phone based   37
occupancy data [16]. In this study, we systematically develop a model for the PDF of   38
secondary infections from the equation governing the virion concentration at an indoor   39
location, and couple its solution to the dose-response model governing pathogen-host   40

interactions. The formulation is extended to an ensemble of locations to obtain a PDF for secondary infections due to long-range airborne transmission, ensuring an exclusively theoretical foundation. Additionally, this model is coupled with the contributions of short-range (or near-field) virus transmission to ensure that all the major sources of virus spread are captured. The final result is intended to be an analytical model capable of predicting outbreak sizes in real-world indoor locations while being able to adapt its solution to applied mitigation measures owing to its underlying theoretical foundation that captures the influence of such strategies.

One mitigation strategy applied during the pandemic has been widespread school closures, wherein teaching was moved to virtual classrooms. Emerging data suggest the move from in-person to virtual may have had a negative impact on students' academic performance [17]. Perhaps, opportunities exist for more focused mitigation measures that include structural factors like enabling physical distancing with fewer students per classroom, ventilation, and more proximal measures like access and use of masks to offset the need to close schools in the event of a SARS-CoV-2 case in the classroom, or prior to detection of a case even when community transmission is high. In other words, with the ability to forecast the likelihood and magnitude of outbreaks in classrooms ahead of time, along with understanding the impact of various mitigation measures, making decisions about the advantages and disadvantages of school closures could be more well-informed. This opens an avenue for us to apply our model to judge its predictive capabilities in preparation for future outbreaks.

The primary objective of the present study is to derive an analytical model capable of predicting the real-world distribution of the number of secondary infections generated due to the spread of an airborne virus in schools and test its veracity through a comparison with real-world data. For this purpose, we use publicly available data on diagnosed SARS-CoV-2 cases from the public school system in Ontario, Canada, over different epochs in time. The modeling parameters chosen are based on existing literature, while also being guided by school data and mitigation strategies.

Methods

The model is designed to predict the probability density function of the number of secondary airborne infections inside schools. The transmission process can be classified into two broad regimes: the virus traveling from the active index case to a susceptible individual; and the interaction between the virus and the human body. The former is mostly determined by physics-based factors while the latter depends on biological factors, primarily the dose-response of the human body.

First, we briefly review the dose-response model followed by the evolution equation of virus concentration in well-mixed situations i.e., virus concentration is homogeneous in space. Next, a discussion on the contribution of near-field virus concentration in assessing the overall outbreak sizes is provided. Coupling these with the distribution of the variables with the largest relative dispersions, the PDF of the number of secondary infections is derived with and without near-field effects.

For the convenience of the reader, a list of all symbols and their associated definitions are listed in Table 1 and Table 2 in S1 Text.

Dose-response model

Quantitative risk assessment of infections is primarily performed based on two methods [18]. First among these is the well-known Wells-Riley [19,20] approach that provided a relation connecting a quantum of infection to the probability of being infected. The simplicity of this relation allowed for application in various scenarios but

it is unable to capture the underlying mechanisms governing the transmission. On the other hand, the second approach of dose-response models builds upon an analogous framework, while not being restricted by a hypothetical description for infectious dose i.e., quanta, therefore allowing these models to describe the infectious dose in terms of various biological and physical parameters that influence it. These models were first proposed by Haas [12] in the context of waterborne disease spread, but were also later applied in the case of SARS-CoV-2 [13, 21] as

$$\mathcal{P} = 1 - e^{-r\mathcal{N}_v} \tag{1}$$

where the probability of infection is described by $\mathcal{P}$, and $\mathcal{N}_v$ is the number of virions inhaled by a susceptible. The dose-response constant r was explained by Haas [13] as the inverse of the probability of a single virus surviving till it can trigger illness. The probability of infection $\mathcal{P}$ has two interpretations - (1) if $\mathcal{N}_v$ is the virion quantity inhaled by a person at a given location and time, then $\mathcal{P}$ corresponds to that individual's probability of getting infected from the inhaled virions; (2) if $\mathcal{N}_v$ is the average infectious dose inhaled by a group of susceptibles, then $\mathcal{P}$ represents the proportion of susceptibles that will become infected. The latter interpretation will be applied in the present model.

Secondary infection count from the dose-response model

For the probability of infection $\mathcal{P}$, we need to compute $\mathcal{N}_v$ in terms of known or computable quantities that can be obtained from existing school data and index case data. To that end, consider a classroom with one index case and multiple susceptible individuals. At any point in time, the susceptible individuals would be exposed to an airborne virion concentration field $c(\vec{x}, t)$. Here, a well-mixed room is assumed to invoke $c(\vec{x}, t) = c(t)$ i.e., the virion concentration is homogeneous in space. If the index case possessing a viral load ρ continuously ejects mucosalivary fluid at a rate of $Q(t)$ within a room of volume V, then the virion concentration c is governed by

$$\frac{dc(t)}{dt} = \frac{\rho Q(t)}{V} - ac(t) \tag{2}$$

where the effects of air change rate ACH, wall deposition parameter β_0, and virus half-life $t_{1/2}$ appear through the loss parameter $a = ACH/3600 + \beta_0 + \ln(2)/t_{1/2}$ [5]. The ventilation effect is assumed to be uniform here, though it is recognized that practically the aerosol flow can be anisotropic with a degree of directionality. The jet-puff model introduced later will partially address this. The volume flow rate $Q(t)$ is assumed to have a constant value Q_l for the duration of the ejection event $\mathcal{T}$, followed by a zero value for the remaining duration. The mucosalivary fluid ejection event under consideration is speaking, while breathing is neglected due to its comparatively (several orders of magnitude) lower contribution of ejected volume [22]. The solution of Eq. 2 takes the form

$$c(t) = \begin{cases} \rho \frac{Q_l}{Va} \left[1 - \exp(-at) \right] & \text{if } t \leq \mathcal{T} \\\\ \rho \frac{Q_l}{Va} \left[\exp\{-a(t - \mathcal{T})\} - \exp(-at) \right] & \text{if } t > \mathcal{T} \end{cases} \tag{3}$$

For this piecewise description of concentration $c(t)$, the regime bounded by $t \leq \mathcal{T}$ marks the speaking duration of the index case. Beyond that, when $t > \mathcal{T}$, the virion concentration decays. If the susceptible individuals are inhaling these virions at an

average rate of $\dot{V}_b$ for a mean exposure duration of τ, the average number of virions inhaled by a single individual $\mathcal{N}_v$ can be written up to the leading order as

$$\mathcal{N}_v(\tau) = \int_0^\tau c(t)\dot{V}_b dt \approx \frac{\rho Q_l \dot{V}_b}{Va}\mathcal{T} \tag{4}$$

This leading order form was obtained through a series expansion of the result of the integral in Eq. 4 along with the application of a long exposure duration condition ($a\tau \gg 1$), which is easily satisfied in classroom scenarios. At this stage, Eq. 4 leaves us with a computable form for $\mathcal{N}_v$, which can now be substituted back in the dose-response model in Eq. 1 to obtain the probability of infection as

$$\mathcal{P} = 1 - \exp(-w\alpha) \tag{5}$$

where

$$w = \rho r \tag{6}$$

captures the primary sources of variability in $\mathcal{P}$, while α is a constant that incorporates all other factors that influence the transmission process, and takes the form

$$\alpha = \frac{Q_l \dot{V}_b \mathcal{T}}{Va} \tag{7}$$

Observe that α admits mostly physics-based parameters within its description, whereas w is composed exclusively of biological parameters. This means that if the aerosol transmission physics changes it would be reflected through the constant α.

In Eq. 5, $\mathcal{P}$ represents the proportion of secondary infections in a classroom with occupancy n and area A, from which the total number of secondary infections Z occurring there, under well-mixed conditions, is

$$Z = \int_A \chi \mathcal{P} dA \tag{8}$$

where $\chi = n/A$ is the population density in the classroom. The number of secondary infections Z can be expressed through a more generalized term which is independent of occupancy – the secondary attack rate $\tilde{Z}$, defined as the ratio of the number of secondary infections to the number of susceptibles at a location, such that $\tilde{Z} = Z/n$. Noting that an individual's probability of infection is independent of their spatial location due to the well-mixed room assumption, using Eq. 8, $\tilde{Z}$ takes the form

$$\tilde{Z} = \frac{1}{n}\int_A \chi \mathcal{P} dA = \frac{1}{n}\mathcal{P}\int_A \chi dA = \mathcal{P} = 1 - \exp(-w\alpha) \tag{9}$$

The invariance with occupancy makes the secondary attack rate a measure for the proportion of people infected, analogous to the probability $\mathcal{P}$. The number of secondary infections due to one index case in a classroom can now be represented in terms of known or computable quantities as

$$Z = n\tilde{Z} = n\left\{1 - \exp\left(-w\alpha\right)\right\} = n\left\{1 - \exp\left(-\rho r\frac{Q_l \dot{V}_b \mathcal{T}}{Va}\right)\right\} \tag{10}$$

If the index case and the susceptible individuals are wearing masks with filtration efficiency F_{mask}, it would modify the ejection and inhalation terms in Eq 10, now represented by $F_{mask}Q_l$ and $F_{mask}\dot{V}_b$ respectively.

Note that this entire derivation for long-ranged transmission has been performed for a classroom. But, the results we finally intend to obtain are for a school. How this long-range analysis for classrooms can be extended to schools will be explained in the Results section based on certain observations from real-world infection data.

Modified secondary infection count through the inclusion of near-field transmission effects

Up to this point, the model developed only captures long-range airborne transmission within a classroom by invoking the well-mixed assumption. But in reality, there will be instances where the susceptible individuals will come into contact with the highly concentrated virion-laden cloud ejected by the index case before it diffuses. Even in scenarios where there are almost no infections through long-range transmission by virtue of the index case having a very low viral load, this cloud carrying a much higher virion concentration compared to the well-mixed case might still be capable of infecting those who come in contact with it. The concentration is higher since the ejected virions are localized in a smaller volume. To capture this effect, an equation for a modified secondary infection count Z_m is written

$$Z_m \approx \underbrace{n\left\{1 - \exp(-w\alpha)\right\}}_{\text{Long-range transmission}(Z)} + \underbrace{n_j\left\{1 - \exp(-w\alpha_j)\right\}}_{\text{Near-field transmission}(Z_j)} = Z + Z_j \qquad (11)$$

where n is the occupancy of the classroom that hosts the index case, and n_j is the total number of people in the school exposed to near-field transmission. Here, the total number of secondary infections has been modified simply through the addition of a near-field contribution Z_j which mathematically varies from the existing Z by having a modified occupancy n_j and a modified constant α_j that reflects the difference in the underlying physics of near-field transmission compared to the long-range route. Note that susceptibles counted in n that undergo long-range transmission may also get counted again in n_j, which would appear to occasionally overestimate the total infection count. A detailed mathematical proof showcasing why such an overestimation will not occur is provided in the S2 Text.

The question remains about how to obtain the modified constant α_j that describes the near-field transmission physics. This is introduced through a simple jet-puff model for the ejected cloud that provides us with the cloud volume at any time t. An analogous assumption to that of the long-range model is applied i.e., the virion concentration within the cloud is spatially homogeneous. This enables us to use the same approach as the long-range model and solve a concentration equation for virions within the cloud up to its transition to a diffusion-dominated flow, assuming negligible concentration losses due to the transitory nature of the cloud, and subsequently obtain

$$Z_j = n_j\left\{1 - \exp(-w\alpha_j)\right\} \qquad (12)$$

where the modified transmission physics is admitted through the constant α_j as

$$\alpha_j = \frac{Q_l \dot{V}_b t_{tr}}{V_j}\tau_j \qquad (13)$$

In Eq. 13, the transition time to diffusion dominated flow t_{tr}, the volume of the cloud V_j, and the average exposure duration to the cloud τ_j are new parameters unique to the near-field model. Here, t_{tr} and τ_j are input quantities whose values are discussed later, whereas the cloud volume V_j is calculated using a simple jet/puff model in S3 Text.

Analytical PDF of total secondary infection count

At this stage, all the governing equations (Eq. 10, 11, and 12) required to describe the transmission process within a single location have been derived, and hence we have a
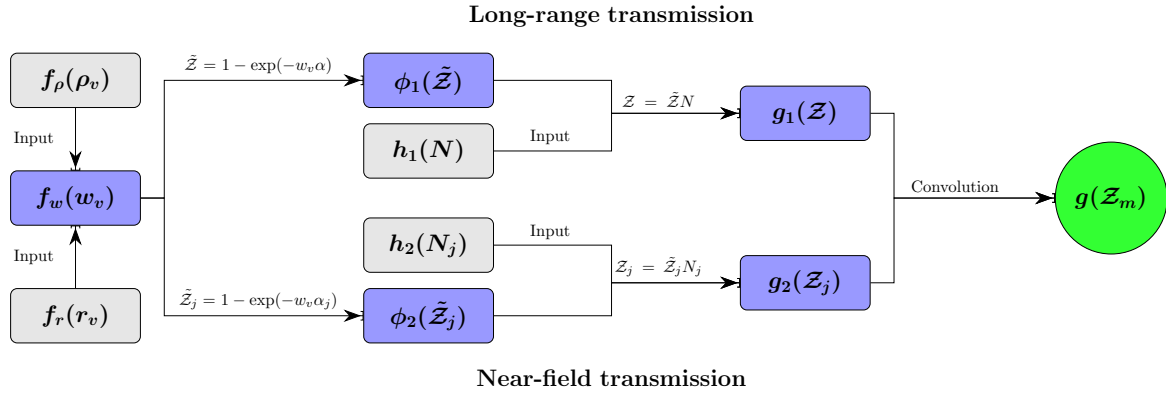
Long-range transmission

Near-field transmission

Fig 1. Algorithm for deriving the analytical PDF of total secondary infection count. The input distributions for viral load ρ_v and dose-response constant r_v are used to compute the PDF of $w_v : f_w(w_v)$. This is applied in both halves of the transmission process – long-range and near-field, to find the PDF of the corresponding secondary attack rates, $\phi_1(\tilde{\mathcal{Z}})$ and $\phi_2(\tilde{\mathcal{Z}}_j)$. These coupled with their respective input occupancy distributions, $h_1(N)$ and $h_2(N_j)$, gets us the PDF of secondary infections for long-range transmission $g_1(\mathcal{Z})$ and near-field transmission $g_2(\mathcal{Z}_j)$. A final convolution operation on these distributions outputs the total secondary infection count PDF, $g(\mathcal{Z}_m)$

closed system of equations. However, the goal is to model the variability of secondary infection count across different locations. This can be captured through a probability density function (PDF). The algorithm to obtain said PDF is shown in the flowchart provided in Fig. 1. Before proceeding, let us define the following pairs of quantities and their corresponding sample space variables: (ρ, ρ_v), (r, r_v), (w, w_v), (n, N), (n_j, N_j) $(\tilde{Z}, \tilde{\mathcal{Z}})$, $(\tilde{Z}_j, \tilde{\mathcal{Z}}_j)$, $(Z, \mathcal{Z})$, $(Z_j, \mathcal{Z}_j)$, and $(Z_m, \mathcal{Z}_m)$.

As shown in Fig. 1, we start from the input distributions for the most dispersive quantities – the viral load and dose-response constant, given by $f_\rho(\rho_v)$ and $f_r(r_v)$. These are used to compute the PDF of w_v, given by $f_w(w_v)$, which attains a lognormal form as derived later in the section detailing the inputs to our model.

$$f_w(w_v) = \frac{1}{w_v \sigma_w \sqrt{2\pi}} e^{-(\ln(w_v) - \mu_w)^2 / 2\sigma_w^2} \tag{14}$$

Here, μ_w and σ_w are the mean and standard deviation of $\ln(w_v)$, respectively. This distribution is utilized in both long-range and near-field transmission routes without change because as stated before, the transmission physics is embedded in the constants α and α_j. Mathematically, the PDF formulations for both routes are analogous, and hence, in the remainder of the section, we will only describe the long-range PDF in detail with the understanding that the same process can be followed for near-field transmission.

First, for long-range transmission, the secondary attack rate described by Eq. 9 is re-written as

$$\tilde{\mathcal{Z}} = 1 - \exp(-w_v \alpha) \tag{15}$$

where α is a constant, and w_v has a distribution $f_w(w_v)$. The justification behind assigning α an ensemble-averaged constant value instead of a distribution is discussed later. In Eq. 15, we observe a functional dependency of $\tilde{\mathcal{Z}}$ on w_v. Therefore, the PDF of $\tilde{\mathcal{Z}}$, given by $\phi_1(\tilde{\mathcal{Z}})$, depends on $f_w(w_v)$ as

$$\phi_1(\tilde{\mathcal{Z}}) = f_w(w_v) \frac{dw_v}{d\tilde{\mathcal{Z}}} \tag{16}$$

Rewriting $w_v = -\frac{1}{\alpha}\ln(1 - \tilde{\mathcal{Z}})$ based on a re-arrangement of Eq. 15, we get ₂₂₁

$$\phi_1(\tilde{\mathcal{Z}}) = f_w\left\{-\frac{1}{\alpha}\ln(1 - \tilde{\mathcal{Z}})\right\}\frac{d\{-\frac{1}{\alpha}\ln(1 - \tilde{\mathcal{Z}})\}}{d\tilde{\mathcal{Z}}} \tag{17}$$

Substituting the lognormal form of $f_w(w_v)$ in Eq. 17, $\phi_1(\tilde{\mathcal{Z}})$ can be expressed as ₂₂₂

$$\phi_1(\tilde{\mathcal{Z}}) = \frac{e^{-\left[\ln\left\{-\frac{1}{\alpha}\ln(1-\tilde{\mathcal{Z}})\right\}-\mu_w\right]^2/2\sigma_w^2}}{\left\{-|1-\tilde{\mathcal{Z}}|\ln(1-\tilde{\mathcal{Z}})\right\}\sigma_w\sqrt{2\pi}} \tag{18}$$

Now, to find the PDF of long-range secondary infection count $g_1(\mathcal{Z})$, we need to ₂₂₃ combine $\phi_1(\tilde{\mathcal{Z}})$ with the input PDF of classroom occupancy $h_1(N)$. This is achieved by ₂₂₄ employing the relation for finding the PDF of the product of two independent random ₂₂₅ variables ($\mathcal{Z} = N\tilde{\mathcal{Z}}$) as ₂₂₆

$$g_1(\mathcal{Z}) = \int_0^\infty h_1(N)\phi_1(\mathcal{Z}/N)\frac{1}{N}dN = \int_0^\infty \frac{h_1(N)e^{-\left[\ln\left\{-\frac{1}{\alpha}\ln(1-\mathcal{Z}/N)\right\}-\mu_w\right]^2/2\sigma_w^2}}{-\sigma_w\sqrt{2\pi}(N - \mathcal{Z})\ln(1 - \mathcal{Z}/N)}dN \tag{19}$$

The exact procedure can now be repeated for near-field transmission to get an ₂₂₇ analogous PDF of near-field secondary infection count $g_2(\mathcal{Z}_j)$. ₂₂₈

Finally, a convolution operation is performed to obtain the PDF of the sum of two ₂₂₉ random variables, $\mathcal{Z}_m = \mathcal{Z} + \mathcal{Z}_j$; the PDF of net secondary infection count $\mathcal{Z}_m$ is ₂₃₀ therefore computed as ₂₃₁

$$g(\mathcal{Z}_m) = \int_0^\infty g_1(\mathcal{Z})g_2(\mathcal{Z}_m - \mathcal{Z})d\mathcal{Z} \tag{20}$$

Inputs to the analytical model ₂₃₂

With the disease transmission model being completely defined through Eq 10, 12 and ₂₃₃ 20, we turn our focus towards the inputs that are required to solve the model and ₂₃₄ obtain the PDF of secondary infections. As observed before, the Z and Z_j equations ₂₃₅ (Eq. 10 and Eq. 12) require inputs to define three types of quantities: w, α (and α_j), n ₂₃₆ (and n_j). The primary sources of variability are embedded within $w = \rho r$ and hence, ρ ₂₃₇ and r will both admit individual distributions to account for this. On the contrary, α ₂₃₈ and α_j are composed of quantities that have comparatively negligible variation, and ₂₃₉ therefore, these quantities will all be assigned ensemble averaged values. Finally, n and ₂₄₀ n_j bring in the effect of occupancy through distributions of their own and is a highly ₂₄₁ relevant parameter since secondary infection count is linearly proportional to occupancy. ₂₄₂

In the subsequent sections, we describe the input distributions for ρ, r, and w, ₂₄₃ followed by a section on all the ensemble-averaged parameters that generate α and α_j, ₂₄₄ and end with a section describing occupancy distributions. ₂₄₅

Probability density function for the viral load ₂₄₆

Viral load ρ is a dominant factor that decides the shape of the $\mathcal{Z}$ distribution, and ₂₄₇ hence its variability needs to be captured when providing it as an input parameter [5, 7]. ₂₄₈ The study by Yang et al. [23] collected viral load data for both asymptomatic and ₂₄₉ symptomatic individuals corresponding to the original variant of SARS-CoV-2, to ₂₅₀ generate a viral load PDF which showed a lognormal distribution of the form ₂₅₁ $f_\rho(\rho_v; \mu_y = 13.83, \sigma_y = 3.8)$ [5, 23], presented in Fig. 2(a) by a solid blue line. ₂₅₂

For the current study, we focus on a period between Sep-Dec 2021 when δ-variant ₂₅₃ was most dominant. Therefore, while retaining a similar lognormal form $f_\rho(\rho_v; \mu_\rho, \sigma_\rho)$, ₂₅₄

the distribution's parameters need to be modified for the δ-variant. The δ-variant viral load vs. time plot from Jüni et al. [24] was used to calculate a time-averaged median viral load whose exponential gives $\mu_\rho = 17.97$, and σ_ρ was assumed to be same as that obtained from Yang et al. [23]. The final δ-variant viral load distribution is also shown in Fig.2(a) by a solid red line.

Probability density function for the dose-response constant

The dose-response constant r, which has the units (RNA) 'copies^{-1}', can be decomposed as

$$r = \frac{1}{\psi k} \tag{21}$$

where k is an analogous dose-response constant parameter similar to r, and can be computed from literature, but generally presented in 'Focus-forming units (FFU)'. To account for the difference in units between r and k, a quantity ψ is introduced whose inverse is often referred to as infectivity. In other words, ψ is a measure of the percentage of viral matter in droplets and aerosols that is capable of initiating infections. To find the PDF of r (corresponding sample space variable r_v), the quantities k and ψ need to be determined.

For k, we refer to Killingley et al. [25], where they found that an infectious dose of 55 FFU of SARS-CoV-2 among volunteers showed 53% infection rate. Substituting these values in the dose-response model of Eq. 1 gives us $k = 73$ FFU. The literature regarding k values for SARS-CoV-2 are few and far between, and insufficient to obtain a distribution. However, assigning a constant k suffices because the primary variability in the dose-response constant appears from ψ which varies over several orders of magnitude [26–28].

For the PDF of ψ, we refer to a study by Lin et al. [28] that found ψ and therefore the dose-response constant to have a lognormal distribution, but not for the δ-variant. Therefore, using the dataset for δ-variant ψ from the study by Despres et al. [27], the dose-response constant's mean and standard deviation were computed to assign the parameters of its lognormal distribution $f_r(r_v; \mu_r, \sigma_r)$ as

$$\mu_r = \mu_{ds} = \ln(m_{ds}) - \frac{\sigma_{ds}^2}{2} = -14.32 \quad \text{and} \quad \sigma_r = \sigma_{ds} = \sqrt{\ln\left(\frac{s_{ds}^2}{m_{ds}^2} + 1\right)} = 1.06 \tag{22}$$

Here m is the mean of the data, and s is the corresponding standard deviation. The subscript 'ds' refers to the values being obtained using the data from Despres et al. [27]. The final PDF for r_v is displayed in Fig. 2(b) as a dark red line, compared to an approximate numerical distribution from Despres at al. [27] shown through blue circles. The pre-δ-variant distribution from Lin et al. [28] is also showcased in pink circles.

Probability density function for the product of viral load and dose-response constant

Since, ρ_v and r_v are both lognormally distributed, their product $w_v = \rho_v r_v$ also has a lognormal distribution with parameters $\mu_w = \mu_r + \mu_\rho = 3.64$ and $\sigma_w = \sqrt{\sigma_r^2 + \sigma_\rho^2} = 3.94$. The PDF of w_v is shown in Fig. 2(c).

Description of the ensemble-averaged input parameters

All the ensemble-averaged parameters are embedded within α and α_j, given by Eq. 7 and Eq. 13, and need to be properly defined as they introduce flow physics into the transmission model. Table 1 describes every parameter that is used to generate α and
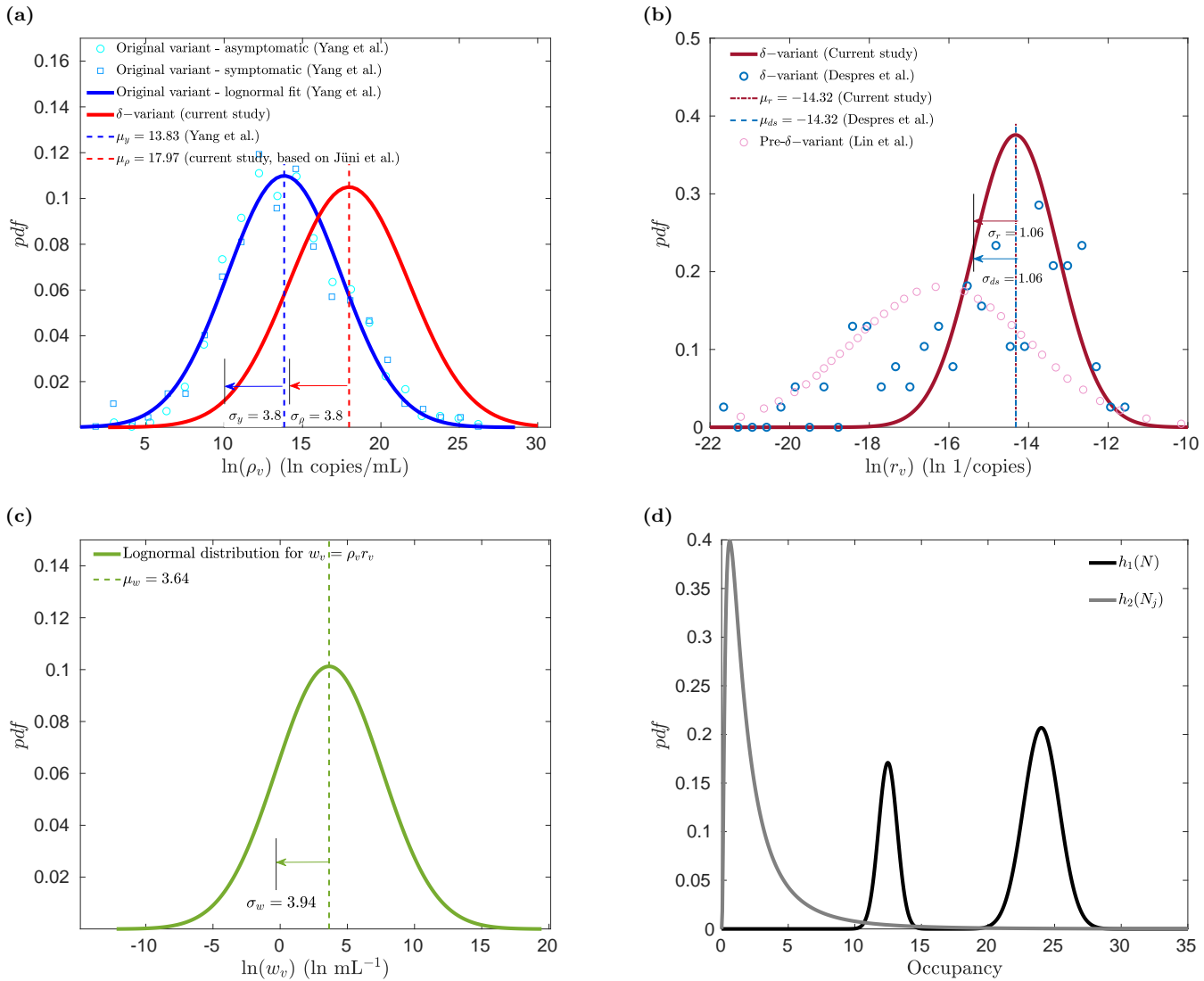
Fig 2. Probability density functions of input parameters. (a) Viral load ρ_v PDF for asymptomatic (circular markers) and symptomatic (square markers) population corresponding to the original SARS-CoV-2 variant [23] along with a lognormal fit ($\mu_y = 13.83$, $\sigma_y = 3.8$) (solid blue line) for the data, compared to the δ- variant viral load lognormal distribution ($\mu_\rho = 17.97$ [24], $\sigma_\rho = 3.8$ [23]) (solid red line), (b) Lognormal PDF of dose-response constant r_v for SARS-CoV-2 (pre-δ-variant) [28] (pink circles), compared to the PDF computed from δ-variant dataset [27] (blue circles), and a lognormal fit (dark red line) to this dataset ensuring equal mean and standard deviation ($\mu_r = \mu_{ds} = -14.32$, $\sigma_r = \sigma_{ds} = 1.06$), (c) Lognormal distribution for product of viral load and dose-response constant, w_v ($\mu_w = 3.64$, $\sigma_w = 3.94$), (d) PDF of classroom occupancy $h_1(N)$ (black line) and PDF of the number of people exposed to near-field transmission, $h_2(N_j)$ (grey line).

α_j, along with their corresponding values, and how the values were computed. The question remains as to how we justify the parameters with α and α_j having ensemble-averaged values, as opposed to distributions like ρ and r. This can be answered if the dispersion for all these parameters across locations is calculated and compared. The quartile coefficient of dispersion for the ith parameter

$$Q_i = \frac{(Q_3 - Q_1)}{(Q_1 + Q_3)} \tag{23}$$

Table 1. Input Parameters for α and α_j

Symbol	Definition	Value	Reference
V	Classroom volume	209 m³	Based on classroom size recommendation by a 2010 expert panel report [29] along with the assumption of classroom height $h = 3$ m
$\dot{V_b}$	Inhalation rate	105 cm³s⁻¹	Inhalation rate is given as $\dot{V_b} = T_v W_a/\Delta$; $T_v = 7$ mL/kg is the average tidal volume per unit weight for children [30]; $W_a = 30$ kg is the assumed average weight of the susceptible individuals; $\Delta = 2$ s is the duration of inhalation based on the average breathing rate of $12 - 15$ times per minute [31]
Q_l	Ejected volume flow rate	2.22×10^{-6} mL/s	The exhalation rate of air for the index case is taken as 700 Lh⁻¹ or 194.4 cm³s⁻¹ [22]. This is combined with the aerosol size distribution for speaking [22] to obtain Q_l as showcased by Chaudhuri et al. [5]
F_{mask}	Mask filtration efficiency	0.5	Morais et al. [32] reported that commonly used homemade masks block 20% ($F_{mask} = 0.8$) to 60% ($F_{mask} = 0.4$) of all incoming aerosols, while N95 masks have much higher efficiency. We take an average value of 0.5 for our case
τ	Room-wide homogeneous virion field exposure duration	6 hrs	Based on the total duration of classes in a school day [33]
ACH	Air change rate	2 hr⁻¹	Reported by Swyers and King, CBC news article, 2021 [34]
β_0	Wall-deposition parameter	0.002 s⁻¹	Refer to the calculations of Chaudhuri et al. [5] based on the study by Lai and Nazaroff [35]
$t_{1/2}$	Virus half-life	32.07 min	Based on the experiments of Dabisch et al. [36] at ASHRAE recommended indoor air conditions, using the DHS calculator [37]
t_{tr}	Transition time to diffusion regime	39 s	Defined as the time when cloud velocity is comparable to the ambient velocity fluctuations [38]. The final secondary infection PDF was found to be much less sensitive to t_{tr} compared to most other parameters, and hence, it was approximated as the time when the cloud velocity becomes 1% of the cloud ejection velocity
τ_j	Near-field cloud exposure duration	60 s	Assumed to be a short duration on average

is employed for this purpose. Here, Q_1 is the first quartile and Q_3 is the third quartile of a given data set i.e., Q_i essentially provides a normalized measure for the spread of the data, while not being too sensitive to outliers in the dataset.

The quartile coefficient of dispersion Q_i for the different parameters in w, α, and α_j were calculated. For the quantities in $w = \rho r$, $Q_\rho = 0.98$ and $Q_r = 0.92$ i.e., $Q_i \approx 1$ for both, suggesting very high dispersion. In comparison, we have $Q_{Q_l} = 0.47$ [5, 22, 39–42] and $Q_{ACH} = 0.47$ [5, 43], both much smaller than 1. Available values for the remaining quantities such as wall-deposition parameter [35], volume inhalation rates [22], virus half-life, classroom volumes, and others suggest significantly less variation. Therefore, ρ and r having much higher Q_i than every other quantity justifies assigning ensemble averaged values to every parameter in α and α_j. Note that some parameters like ejection event duration, exposure duration to the ejected cloud, etc., do not have sufficient data in the literature to calculate dispersion. Hence, in the Results section, we have provided a range within which the PDF of secondary infections could lie to account for a certain degree of variation in such quantities.

Probabiliy density function of occupancy ₃₁₆

The final step is to find an occupancy PDF that describes classroom populations. The ₃₁₇
present study has been carried out for the entire Ontario public education system, ₃₁₈
which as of $2020 - 21$ data has a total of $2,025,265$ students enrolled in $4,833$ schools, ₃₁₉
split into $1,394,040$ elementary students and $631,225$ secondary students [44]. This ₃₂₀
results in the average number of students per school being 419. Based on available ₃₂₁
data [45], the mean of the maximum classroom size is $\mu_1 = 24$ students with a standard ₃₂₂
deviation of $\sigma_1 = 1.35$. A Gaussian distribution was fit based on these parameters as a ₃₂₃
first estimate for $h_1(N)$. But, this description neglects the effect of vaccination on the ₃₂₄
susceptible population distribution. We assume that elementary school students have ₃₂₅
not been vaccinated during the period of study. In contrast, secondary school students ₃₂₆
have a vaccination coverage of $\eta_{cov} = 0.8$ and a vaccination efficacy of $\eta_{vac} = 0.6$ [46]. ₃₂₇
The susceptible occupancy n is simply modified to $n(1 - \eta_{vac}\eta_{cov})$, and the modified ₃₂₈
Gaussian distribution has a mean of $\mu_2 = 12.48$ and a standard deviation of $\sigma_2 = 0.7$. ₃₂₉
To incorporate both cases in a single model, a bimodal distribution $h_1(N)$ is chosen for ₃₃₀
occupancy, with a weighting factor of $p = 0.3$ (for secondary school students) based on ₃₃₁
the percentage of students in elementary and secondary schools. This results in a $h_1(N)$ ₃₃₂
distribution of the form ₃₃₃

$$h_1(N) = (1 - p)\frac{1}{\sigma_1\sqrt{2\pi}}e^{-(N-\mu_1)^2/2{\sigma_1}^2} + p\frac{1}{\sigma_2\sqrt{2\pi}}e^{-(N-\mu_2)^2/2{\sigma_2}^2} \qquad (24)$$

shown in Fig. 2(d) with a black line. ₃₃₄
    For the distribution of susceptibles exposed to near-field transmission $h_2(N_j)$, we ₃₃₅
assume a long-tailed distribution to account for the likely large variation in the number ₃₃₆
of people interacting with the index case at close range. The simplest long-tailed ₃₃₇
distribution is chosen i.e., the lognormal. Its parameters were taken as $\mu_j = 0.5$ and ₃₃₈
$\sigma_j = 1$ to ensure that the average number of near-field transmissions is low, which is ₃₃₉
expected because close-range interactions would be limited during a pandemic. Fig. ₃₄₀
2(d) also showcases $h_2(N_j)$ through a grey line. ₃₄₁

Results ₃₄₂

The results presented correspond to two sources: real-life school infection data and ₃₄₃
modeled equations. Each source was used to obtain a probability density function for ₃₄₄
the total number of secondary infections occurring at a school within one school day. ₃₄₅

Analysis of reported school infection data ₃₄₆

To generate the secondary infection PDF from actual cases, we draw on all reported ₃₄₇
SARS-CoV-2 diagnoses in 4,833 schools within the Ontario public school system. Ten ₃₄₈
distinct dates between 14 September 2021 and 13 December 2021 are chosen. The ₃₄₉
Ontario government official website [47] hosts the infection data reported by each school ₃₅₀
under its jurisdiction almost daily, which involves both student cases, and the grossly ₃₅₁
outnumbered staff cases that we therefore neglected. This data is sufficient to plot the ₃₅₂
PDF for secondary infections in schools within the province of Ontario for any date. ₃₅₃
However, the generated PDF will be based on certain data points that correspond to ₃₅₄
scenarios where a school had no index cases and thus reported no new cases. Recall that ₃₅₅
our modeled PDF is only for scenarios where an index case is present. To exclude ₃₅₆
scenarios with no active index cases in a school, the following operations are performed. ₃₅₇
    Based on existing literature [24], we assume that an average of four days separate a ₃₅₈
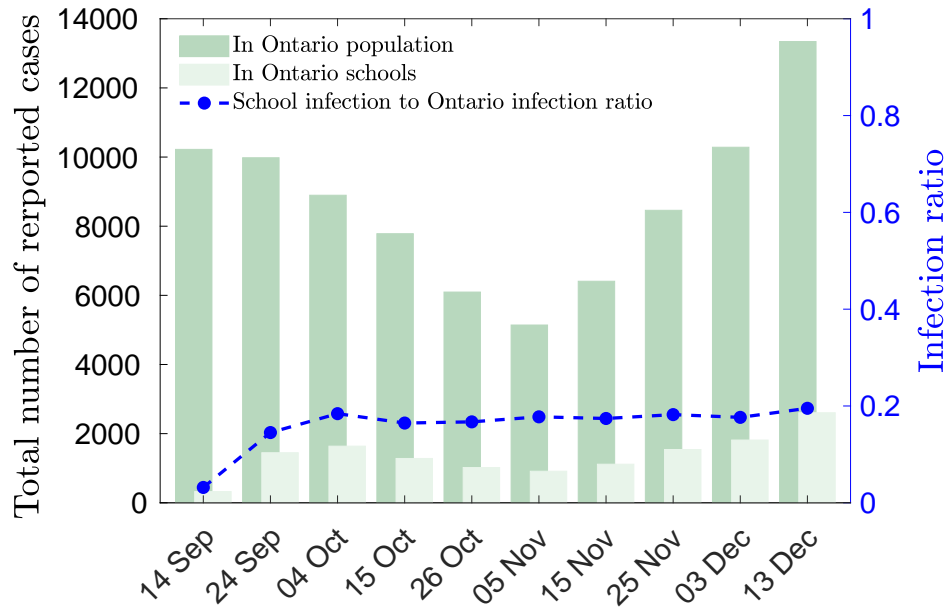susceptible getting infected in a school, and the school testing and reporting said ₃₅₉

Fig 3. Reported SARS-CoV-2 cases in Ontario province and Ontario public schools. Bar charts for cumulative cases reported in Ontario in the two weeks before infection and cases reported in Ontario public schools for different dates are displayed. The blue curve corresponds to the ratio between cases reported in Ontario public schools to those reported in the entire province. Note that the total number of reported cases in Ontario corresponds to two weeks before the date of infection (assumed to be 4 days before infection was caught through testing) and not the date of reporting that is shown in the plot. All dates are for the year 2021.

infection. Fig. 3 shows the cumulative number of reported cases in Ontario in the two weeks before the day of infection [48], compared to those reported in Ontario public schools on the same dates. An immediate observation is that the number of school cases and Ontario cases are directly proportional, shown through the school infection to Ontario infection ratio in Fig. 3 being nearly constant over the three months under consideration. This allows us to assume that on the day of infection, the active index case proportion in the school population is equal to that in the Ontario population, given by ϕ_i. Therefore, the average number of index cases per school I_i equals

$$I_i = \phi_i S \tag{25}$$

where S is the average student count in schools. Based on the data processed in this study, we note that all values of $I_i \leq 1$. Therefore, assuming that index cases are spread evenly across N_s schools in Ontario, the total number of schools with at least one index case N_i on the day of infection is

$$N_i = I_i N_s \tag{26}$$

Next, we focus on the reported school infection data and note that they would comprise infections acquired in schools and those from outside schools. Our model does not account for the later contribution and hence it must be subtracted from the data. Thus, if ϕ_j is the active index case proportion in Ontario (and its schools) on the day the schools test and report for SARS-CoV-2, then the total number of outside contributions per school I_j is

$$I_j = \phi_j S \tag{27}$$

Therefore, if Z_{data} is the total infection count reported by a school, then the corrected

360
361
362
363
364
365
366
367

368
369
370
371

372
373
374
375
376
377

378

infection count Z'_{data} accounting for only in-school infections becomes

$$Z'_{\text{data}} = Z_{\text{data}} - I_j \tag{28}$$

A histogram is then generated through proper binning of the Z'_{data} data set and then converted to a PDF using the relation

$$\text{PDF} = \frac{N_{bin}}{W_{bin} \Sigma N_{bin}} \tag{29}$$

where N_{bin} corresponds to the number of elements in a bin i.e., the number of schools corresponding to a particular infection count, and $W_{bin} = 1$ is the bin width of the histogram. Note that the total bin element count ΣN_{bin} is not equal to the total number of schools N_s, but instead equal to the total number of schools with at least one active index case on the day of infection, N_i.

We further clarify the data analysis with the following example. Consider the data reported in Ontario public schools on November 25, 2021. Based on usual SARS-CoV-2 incubation periods, the expected date of infection is taken as November 21, 2021. There were $8,460$ cases reported in the two weeks prior to this date. With Ontario's population of 14 million, the index case percentage is $\phi_i = 0.06\%$. Superimposing this percentage on the $N_s = 4,833$ Ontario public schools with $S = 419$ average students, we get $I_i = 0.2532$ index cases per school. This number would suggest that, on average, there are schools with no index cases i.e., 1 index case in approximately 4 schools. The modified school count would then become $N_i = 0.2532 \times 4,833 = 1,224$ schools. The dates used for data analysis in this study are at least $8 - 10$ days apart to ensure that the infection events are independent and there is no repeat counting of the same infections.

Application of the analytical model in schools

As noted previously, our analytical model is derived for a single classroom scenario. But, in this study, we instead apply it to model SARS-CoV-2 spread in an entire school. We can justify this approach through the observation that, on average, a single classroom with one index case coupled with their near-field transmission contributions is sufficient to represent SARS-CoV-2 spread within a school. The chain of logic used to make this justification is shown in Fig. 4 and is further discussed below.

Referring back to the example of November 25, 2021, in the previous section, we focus on the population in Ontario – index cases and susceptibles, that go to school. For this date, we observe that on average 1 in every 4 school has an active index case i.e., 3 of 4 such schools do not host an index case and are neglected from our analysis. Next, we turn our attention to every such school that hosts an active index case. Here, the index case spends their time inside a classroom or outside it. As a large portion of their school day is spent within a classroom, this is where they become primarily a source for long-range airborne transmission. Their outside-classroom time, however, would involve small-duration interactions with others that are too short for long-range transmission, but instead contribute to the SARS-CoV-2 spread within that school through near-field transmissions. All other classrooms, which do not host active index cases, are non-contributors to the disease spread dynamics within that school.

These sets of observations showcase how a classroom with an active index case, coupled with that same index case's near-field virus transmission contributions is sufficient to represent the average virus spread scenario in an entire school. Hence, our analytical model, which is able to account for these processes, can now be applied to schools.
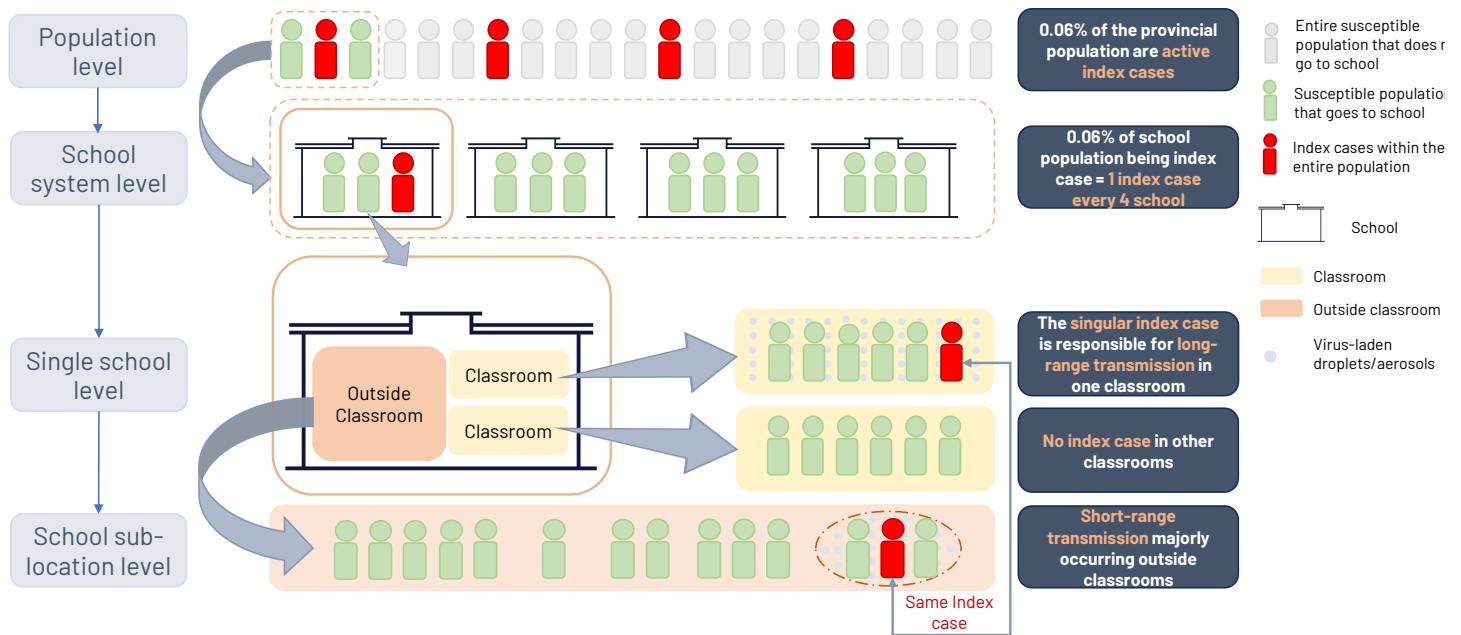
Fig 4. Representation of SARS-CoV-2 spread in a school through a single classroom coupled with outside-classroom effects. The schematic was built based on the example of SARS-CoV-2 infections in school on November 25, 2021. At the population level, individuals are classified into – active index cases who may or may not go to school, susceptibles that go to school, and susceptibles that do not go to school. Moving one level below, we have the entire school system consisting of the susceptible and index case population that goes to school. Here, data tells us that on average only one index case is active across multiple schools. Moving down another level, our focus is shifted to the school hosting that one active index case. Here, the index case divides their time inside a classroom or outside it. At the lowest level, we look at these individual sub-locations (classrooms, outside classrooms) where in one classroom the index case is the cause of long-range and near-field SARS-CoV-2 transmission, whereas every other classroom is devoid of active index cases and thus does not contribute to the disease spread within that school. Additionally, the index case spends time outside classrooms for a shorter duration in closer proximity to some individuals and thus spreads the virus through near-field transmission. Hence, the one classroom with the index case coupled with the index case's outside classroom interactions can approximately represent SARS-CoV-2 spread within an entire school.

Comparison between modeled results and reported school data

The PDFs for the number of secondary infections Z (sample space variable $\mathcal{Z}$) at a school within a single school day, generated from the reported data for the 10 dates chosen, are shown in Fig. 5 using red symbols. All PDFs peak near zero, followed by a strong gradient that smoothens out into a long tail that highlights the overdispersion in SARS-CoV-2 transmission. This common behavior suggests that even though low infection counts are the norm in the presence of an active index case within a school, the existence of the occasional superspreading events described by the data points at large $\mathcal{Z}$, is a constant threat and a reminder for proactive mitigation measures. Most of these PDFs attain small peaks near $\mathcal{Z} \approx 10$ and $\mathcal{Z} \approx 20$, possibly pointing towards susceptible student populations being localized around that mark. The clustering of all the PDFs reveals a shared underlying trend likely described by the properties of a common virus strain (δ-variant) responsible for the infections during Sep-Dec 2021.

In the remainder of this section, we will aim to test whether our current model can replicate such real-world data. To that end, we first employed the initial iteration of our model that solves only for the analytical PDF of total number of secondary infections, $g_1(\mathcal{Z})$ (Eq. 19), due to long-range transmission from an active index case. The input
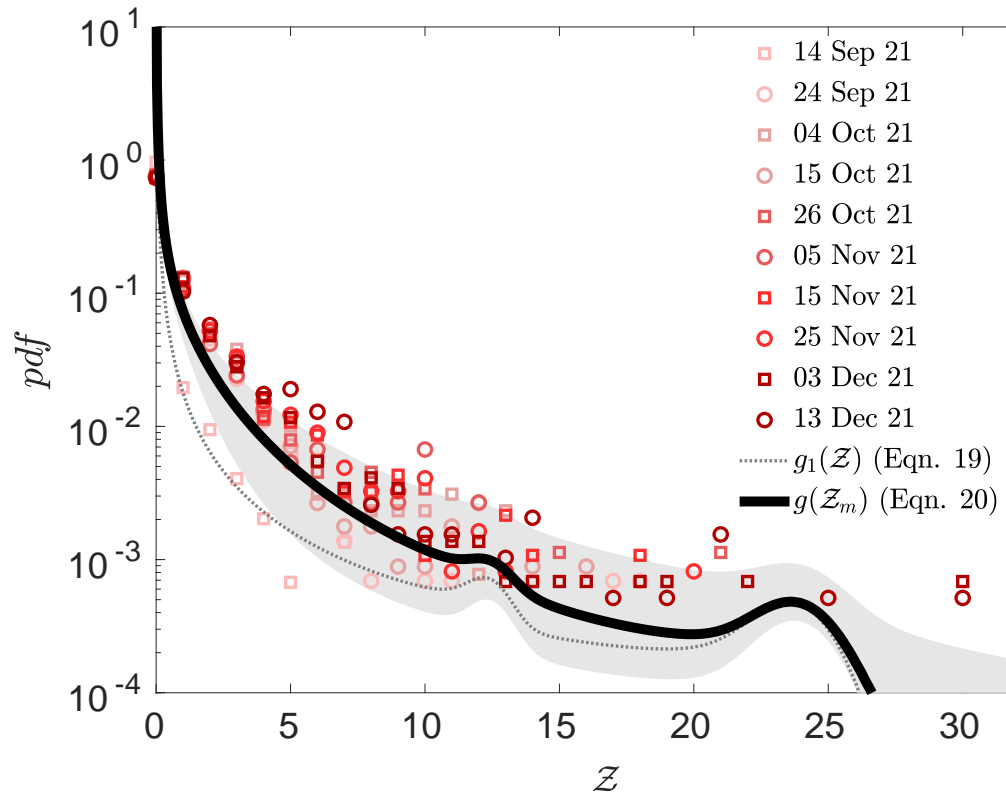
Fig 5. Model predictions for secondary infection distribution in schools compared with reported school infection data. Analytical PDF of secondary infections through long-range transmission $\mathcal{Z}$ (Eq. 19) (thin dotted line) and with added near field transmission contribution $\mathcal{Z}_m$ (Eq. 20) (solid black line), compared to reported infection data in Ontario public schools (red symbols). The shaded region provides an approximate measure for variation in $g(\mathcal{Z}_m)$ with a choice of different input values for certain parameters. The upper and lower bounds of the shaded region correspond to the cases listed in Table 2

parameters used to generate $g_1(\mathcal{Z})$ account for both school-specific data and δ-variant specific biological quantities. The computed $g_1(\mathcal{Z})$ PDF in Fig. 5 shown as a thin dotted line is compared to the reported data, clearly showcasing qualitative similarities while maintaining a degree of quantitative mismatch. In particular, the model notably underpredicts the PDF values at small $\mathcal{Z} \leq 10$, described by the much higher probability at zero, followed by a sharper gradient compared to the reported data. We hypothesized that this could be a limitation that arises from the assumption of homogeneous aerosol concentration, which cannot account for the increased number of infections in close proximity to the index case. The analytical PDF does successfully capture the long-tailed nature of the reported data, while also peaking twice in tandem with it, by virtue of the bimodal Gaussian nature of the classroom occupancy distribution. However, its tail diverges from the reported data and falls off at high $\mathcal{Z}$ ($\mathcal{Z} \geq 25$). This is because the data sample size is limited, causing any unlikely high-valued samples to skew the corresponding probability, somewhat artificially. Additionally, the reported data PDFs have points that go beyond the maximum class size considered in the model, reflecting infections incurred outside classrooms, and is a limitation of representing the disease spread dynamics within a school exclusively through a classroom.

To ensure improved quantitative predictions from our model, a near-field effect was

introduced – modeling the interactions of susceptibles with the localized, high virion $\qquad$ 458
density cloud ejected by the index case. This included coupling the well-mixed virion $\qquad$ 459
field with the ejected cloud represented by a simplified jet-puff model [6] to derive the $\qquad$ 460
PDF of secondary infections due to both long-range and near-field transmission $g(\mathcal{Z}_m)$ $\qquad$ 461
(Eq. 20). The modified PDF $g(\mathcal{Z}_m)$ is shown as a solid black line in Fig. 5 where an $\qquad$ 462
excellent quantitative match is immediately observable as $g(\mathcal{Z}_m)$ now passes through $\qquad$ 463
the cluster of red curves, and retaining the overdispersed nature of the older model. The $\qquad$ 464
magnitude of the gradient at low $\mathcal{Z}$ regime is smaller, while the probability of zero $\qquad$ 465
infections has shifted to lower values, mimicking the reported data. These observations $\qquad$ 466
support our near-field contribution hypothesis and suggest that several of the previously $\qquad$ 467
zero infection scenarios now involve a non-zero near-field infection count. The near-field $\qquad$ 468
model also includes outside-classroom infections and removes the previous limitation of $\qquad$ 469
a classroom-specific description. The success of this updated model strongly emphasizes $\qquad$ 470
the need for aerosolized transmission models to capture both near-field and far-field $\qquad$ 471
mechanisms, while putting forward a practical tool that for the first time, to the best of $\qquad$ 472
our knowledge, reproduces real-world infection spread dynamics in schools from an $\qquad$ 473
exclusively theoretical foundation. $\qquad$ 474

Finally, we revisit the fact that certain parameters used to compute the PDF of $\mathcal{Z}_m$ $\qquad$ 475
have no real-life data source and have been ascribed an intuitive approximate value i.e., $\qquad$ 476
their assigned values likely have an error margin. To account for this possible variation, $\qquad$ 477
two more cases have been run by modifying these particular parameters – (1) for an $\qquad$ 478
approximate lower bound scenario corresponding to low virion concentration exposure $\qquad$ 479
compared to the default case, and (2) for an approximate upper bound scenario $\qquad$ 480
corresponding to high virion concentration exposure. Table 2 lists such input parameters $\qquad$ 481
for all three cases that have been used in this study. Note that these bounding $\qquad$ 482
parameters do not have extreme values (e.g. zero speaking duration for an index case in $\qquad$ 483
one school day), as these cases are not meant to account for outlier situations; those are $\qquad$ 484
already handled by the PDF and its long tail, but rather intended to highlight a $\qquad$ 485
deviation in the approximate average values assigned. The $g(\mathcal{Z}_m)$ curves for these cases $\qquad$ 486
enclose the shaded region in Fig. 5, providing an approximate range within which the $\qquad$ 487
secondary infection PDF could vary. The shaded region itself lies relatively close to the $\qquad$ 488
band of real-life data curves. Even among the quantities varied to obtain the bounds, $\qquad$ 489
not all influence $g(\mathcal{Z}_m)$ equally; quantities like transition time to diffusion barely affect $\qquad$ 490
the PDF, whereas others like μ_j and σ_j that govern the distribution of susceptibles $\qquad$ 491
exposed to near-field transmission have a much stronger influence. This gives us a $\qquad$ 492
glimpse into how mitigation measures, that influence other input parameters, and their $\qquad$ 493
effectiveness at curbing disease transmission will present themselves through the $g(\mathcal{Z}_m)$ $\qquad$ 494
plot. Variation in the input parameters due to possible mitigation measures such as $\qquad$ 495
better masks, higher ventilation, lowered occupancy, vaccination, etc., would either shift $\qquad$ 496
$g(\mathcal{Z}_m)$ to lower values or modify the shape of the PDF, and thus provide insight and $\qquad$ 497
justification behind opting for certain mitigation strategies over others, such that they $\qquad$ 498
allow for efficient school operations during future pandemics. $\qquad$ 499

Discussion $\qquad$ 500

The present study puts forward a robust yet practical tool capable of predicting $\qquad$ 501
outbreak sizes of airborne diseases such as SARS-CoV-2 in schools or similar indoor $\qquad$ 502
locations, allowing for rapid informed implementation of 'precision' mitigation measures. $\qquad$ 503
This tool, in the form of an analytical model, was developed mostly from first principles $\qquad$ 504
and then examined against publicly available school infection data which guided further $\qquad$ 505
model improvements to boost its predictive capabilities. $\qquad$ 506

Model development broadly involved coupling the dose-response function governing $\qquad$ 507

Table 2. Parameters defining the three cases computed in this study

Symbol	Definition	Default	Lower bound	Upper bound
$\mathcal{T}$	Total speaking duration of the index case in one school day	2.4 hr	1.2 hr	3.6 hr
τ_j	Near-filed cloud exposure duration per person	60 s	30 s	120 s
t_{tr}	Puff-to-diffusion regime transition time (based on % of ejection velocity)	39 s (1%)	9 s (3%)	100 s (0.05%)
μ_j	Mean of $\ln(N_j)$ – parameter governing the susceptible count that undergoes near-field transmission	0.5	0.1	1.0
σ_j	Standard deviation of $\ln(N_j)$ – parameter governing the susceptible count that undergoes near-field transmission	1	0.75	1.25

the human body's response to incoming pathogens, with the physics of airborne transmission of said pathogens, to obtain the secondary infection count of an index case. First, we modeled the long-range transmission dynamics of the virus through a concentration equation in a well-mixed room, followed by expanding the model to an ensemble of rooms/locations, such that an analytical probability density function (PDF) of secondary infections could be derived. This stage saw the incorporation of several input parameters, among which the most dispersive ones – viral load and dose-response constant, were assigned distributions based on existing data. The choice of assigning an appropriate distribution instead of a mean value provided our model the flexibility of reacting to the biological processes that bring about such high dispersion in these quantities in the first place. The modeled PDF was then employed to compute the secondary infection distribution in Ontario public schools. The analytical nature of the predictive tool made this computation reasonably fast.

Our first set of comparisons between the modeled PDF based on homogeneous aerosol concentration and reported school data [47] displayed strong qualitative similarities plagued by a degree of quantitative mismatch. The similarities served as evidence of proper incorporation of the major governing factors of transmission, whereas the differences provided insights that led to the introduction of near-field transmission dynamics, modeled through a simple jet/puff model for the index case's ejected virion cloud. A convolution operation on the individual outputs of the long-range and near-field models generated the net secondary infection PDF. Another set of comparisons with reported school data showed an excellent quantitative match, emphasizing the importance of the near-field model. At this stage, we have a model that not only is practical to use due to its analytical tractability but can predict real-world SARS-CoV-2 spread in schools. Apart from predicting future outbreak sizes for SARS-CoV-2 variants in schools, this has potential application toward other airborne disease outbreaks within a variety of indoor locations due to the mostly generalized nature of its underlying description.

Furthermore, the model can also revise its prediction based on the mitigation measure implemented. The final PDF generated, is sensitive to all its component parameters to varying degrees, each of which brings in some form of mitigation strategy. Reduction in inhaled or exhaled volume is affected by masks. Ventilation and ambient effects influence the virion concentration loss term in the formulation. Occupancy and room volume capture the role of population density in enclosed spaces. The physiological properties, viral load, and dose-response, would be a possible avenue to introduce the effects of developed immunity due to vaccinations and repeat infections. A change in any of these parameters in the model would pre-emptively inform the user

about the efficacy of a proposed strategy and allow for the implementation of focused mitigation measures for future pandemics. This functionality allows our model to be a proactive predictive tool applicable to future airborne disease spread in schools and similar venues.

Acknowledgments

References

1. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438(7066):355–359.

2. Morawska L, Milton DK. It is Time to Address Airborne Transmission of COVID-19. Clinical Infectious Diseases. 2020;doi:10.1093/cid/ciaa939.

3. Allen JG, Marr LC. Recognizing and controlling airborne transmission of SARS-CoV-2 in indoor environments. Indoor air. 2020;30(4):557.

4. Prather KA, Marr LC, Schooley RT, McDiarmid MA, Wilson ME, Milton DK. Airborne transmission of SARS-CoV-2. Science. 2020;370(6514):303–304.

5. Chaudhuri S, Kasibhatla P, Mukherjee A, Pan W, Morrison G, Mishra S, et al. Analysis of overdispersion in airborne transmission of Covid-19. Physics of Fluids. 2022;34:051914.

6. Chaudhuri S, Basu S, Saha A. Analyzing the dominant SARS-CoV-2 transmission routes toward an ab initio disease spread model. Physics of Fluids. 2020;32(12):123306.

7. Chen PZ, Bobrovitz N, Premji Z, Koopmans M, Fisman DN, Gu FX. Heterogeneity in transmissibility and shedding SARS-CoV-2 via droplets and aerosols. Elife. 2021;10:e65774.

8. Bhavnani D, James ER, Johnson KE, et al. SARS-CoV-2 viral load is associated with risk of transmission to household and community contacts. BMC Infect Dis. 2022;22(672).

9. Thornton GM, Fleck BA, Kroeker E, Dandnayak D, Fleck N, Zhong L, et al. The impact of heating, ventilation, and air conditioning design features on the transmission of viruses, including the 2019 novel coronavirus: A systematic review of ventilation and coronavirus. PLOS Global Public Health. 2022;2(7):1–25.

10. Ricolfi L, Stabile L, Morawska L, Buonanno G. Increasing ventilation reduces SARS-CoV-2 airborne transmission in schools: a retrospective cohort study in Italy's Marche region. arXiv e-prints. 2022; p. arXiv:2207.02678.

11. Bazant MZ, Bush JW. A guideline to limit indoor airborne transmission of COVID-19. Proceedings of the National Academy of Sciences. 2021;118(17).

12. Haas CN. Estimation of risk due to low doses of microorganisms: a comparison of alternative methodologies. American journal of epidemiology. 1983;118(4):573–582.

13. Haas CN. Action levels for SARS-CoV-2 in Air: Preliminary approach. Risk Analysis. 2021;41(5):705–709.

14. Watanabe T, Bartrand TA, Weir MH, Omura T, Haas CN. Development of a dose-response model for SARS coronavirus. Risk Anal. 2010;30(7):1129–1138.

15. Schijven J, Vermeulen LC, Swart A, Meijer A, Duizer E, de Roda Husman AM. Quantitative microbial risk assessment for airborne transmission of SARS-CoV-2 via breathing, speaking, singing, coughing, and sneezing. Environmental health perspectives. 2021;129(4):047002.

16. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, et al. Mobility network models of COVID-19 explain inequities and inform reopening. Nature. 2021;589(7840):82–87.

17. Hammerstein S, König C, Dreisörner T, Frey A. Effects of COVID-19-Related School Closures on Student Achievement-A Systematic Review. Frontiers in Psychology. 2021;12. doi:10.3389/fpsyg.2021.746289.

18. Sze To GN, Chao CY. Review and comparison between the Wells-Riley and dose-response approaches to risk assessment of infectious respiratory diseases. Indoor Air. 2010;20(1):2–16.

19. Wells WF, et al. Airborne Contagion and Air Hygiene. An Ecological Study of Droplet Infections. Airborne Contagion and Air Hygiene An Ecological Study of Droplet Infections. 1955;.

20. Riley E, Murphy G, Riley R. Airborne spread of measles in a suburban elementary school. American journal of epidemiology. 1978;107(5):421–432.

21. de Oliviera PM, Mesquita LCC, Gkantonas S, Giusti A, Mastorakos E. Evolution of spray and aerosol from respiratory releases: theoretical estimates for insight on viral transmission. Proceedings of the Royal Society A. 2021;477(20200584).

22. Pöhlker ML, Krüger OO, Förster JD, Berkemeier T, Elbert W, Fröhlich-Nowoisky J, et al. Respiratory aerosols and droplets in the transmission of infectious diseases. arXiv preprint arXiv:210301188. 2021;.

23. Yang Q, Saldi TK, Gonzales PK, Lasda E, Decker CJ, Tat KL, 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. 2021;118(21).

24. Jüni P, Baert S, Bobos et al P. Rapid Antigen Tests for Voluntary Screen Testing. Science Briefs of the Ontario COVID-19 Science Advisory Table. 2021;2(52).

25. Killingley B, Mann AJ, Kalinova M, et al. Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. Nat Med. 2022;28:1031–1041. doi:10.1038/s41591-022-01780-9.

26. Sender R, Bar-On YM, Gleizer S, Bernshtein B, Flamholz A, Phillips R, et al. The total number and mass of SARS-CoV-2 virions. Proceedings of the National Academy of Sciences. 2021;118(25):e2024815118.

27. Despres HW, Mills MG, Shirley DJ, Schmidt MM, Huang ML, Roychoudhury P, et al. Measuring infectious SARS-CoV-2 in clinical samples reveals a higher viral titer:RNA ratio for Delta and Epsilon vs. Alpha variants. Proceedings of the National Academy of Sciences. 2022;119(5):e2116518119.

28. Lin Y, Malott R, Ward L, et al. Detection and quantification of infectious severe acute respiratory coronavirus-2 in diverse clinical and environmental samples. Sci Rep. 2022;12(5418).

29. Benson R, et al.. Building Our Schools, Building Our Future - A Report from the Expert Panel on Capital Standards; 2010.

30. Koomen E, Nijman J, Nieuwenstein B, Kappen T. Tidal Volume in Pediatric Ventilation: Do You Get What You See? J Clin Med. 2022;11(1).

31. Barrett KE, Barman SM, Brooks HL, Yuan JXJ. Ganong's review of medical physiology. McGraw-Hill Education; 2019.

32. Morais FG, Sakano VK, de Lima LN, Franco MA, Reis DC, Zanchetta LM, et al. Filtration efficiency of a large set of COVID-19 face masks commonly used in Brazil. Aerosol Science and Technology. 2021;55(9):1028–1041.

33. EduCanada. Understand the Canadian education system; 2023.

34. Swyers K, King A. Even with upgrades and extra filters, classrooms aren't getting equal air quality, experts say. CBC News. 2021;.

35. Lai ACK, Nazaroff WW. Modeling indoor particle deposition from turbulent flow onto smooth surfaces. Journal of Aerosol Science. 2000;51(4):463–476.

36. Dabisch P, Schuit M, Herzog A, Beck K, Wood S, Krause M, et al. The influence of temperature, humidity, and simulated sunlight on the infectivity of SARS-CoV-2 in aerosols. Aerosol Science and Technology. 2021;55(2):142–153.

37. Department of Homeland Security US. Estimated Airborne Decay of SARS-CoV-2; https://www.dhs.gov/science-and-technology/sars-airborne-calculator;. Available from: https://www.dhs.gov/science-and-technology/sars-airborne-calculator.

38. Balachandar S, Zaleski S, Soldati A, Ahmadi G, Bourouiba L. Host-to-host airborne transmission as a multiphase flow problem for science-based social distance guidelines. International Journal of Multiphase Flow. 2020;132:103439.

39. Duguid JP. The size and the duration of air-carriage of respiratory droplets and droplet-nuclei. Epidemiology and Infection. 1946;44(6):471–479. doi:10.1017/S0022172400019288.

40. Loudon RG, Roberts RM. Relation between the Airborne Diameters of Respiratory Droplets and the Diameter of the Stains left after Recovery. Nature. 1967;213:95–96.

41. Loudon RG, Roberts RM. Droplet expulsion from the respiratory tract. American Review of Respiratory Disease. 1967;95(3):435–442.

42. Johnson GR, Morawska L, Ristovski ZD, Hargreaves M, Mengersen K, Chao CYH, et al. Modality of human expired aerosol size distributions. Journal of Aerosol Science. 2011;42(12):839–851. doi:https://doi.org/10.1016/j.jaerosci.2011.07.009.

43. Bohanon HR, Piade JJ, Schorp MK, Saint-Jalm Y. An international survey of indoor air quality, ventilation, and smoking activity in restaurants: a pilot study. Journal of Exposure Science & Environmental Epidemiology. 2003;13(5):378–392.

44. Ontario Government. Education Facts, 2020-2021;. Available from: https://data.ontario.ca/en/dataset/quick-facts-ontario-schools/resource/f8f45027-020c-43d9-8a06-3e548ae5a8e8.

45. Ontario Government. Ontario Regulation 132/12; 2020.

46. Madewell ZJ, Yang Y, Longini J Ira M, Halloran ME, Dean NE. Household Secondary Attack Rates of SARS-CoV-2 by Variant and Vaccination Status: An Updated Systematic Review and Meta-analysis. JAMA Network Open. 2022;5(4):e229317–e229317.

47. Ontario Government. COVID-19 data; 2021. Available from: https://data.ontario.ca/dataset/summary-of-cases-in-schools/resource/dc5c8788-792f-4f91-a400-036cdf28cfe8.

48. Ontario Government. Status of COVID-19 cases in Ontario; 2023. Available from: https://data.ontario.ca/en/dataset/status-of-covid-19-cases-in-ontario/resource/ed270bb8-340b-41f9-a7c6-e8ef587e6d11.