Appendix. Modeling and Simulation Details

We describe the details of our modeling and simulation for estimating secondary infections in classrooms and projecting to the risk over the entire semester. In the Simulation section, we introduce the mechanism behind our simulation tool. In the Assumptions and Parameters section, we state the assumptions and parameter estimates in different components of the simulation. In the Mathematical Model for the Risk of Transmission over Short and Long Ranges section, we develop the mathematical model for the transmission probability between a source case and a susceptible person over different distances, which is an important component in the simulation. Throughout the appendix, we use # to denote "number of" and % to denote "fraction of."

Simulation

We implement a simulation tool in Python to simulate classrooms under different scenarios. We investigate the effect of different intervention measures (masking, increasing distancing, and ventilation) on transmission risk under varying simulation parameters (vaccination rate among students and vaccine efficacy). We describe the simulation mechanism in this section. In the Assumptions and Parameters section, we discuss the assumptions and values for these parameters in more detail.

Our simulation has two stages. In the first stage, we estimate the conditional probability that a student or instructor is infected in a one-hour class given there is one positive source case in the same class. In the second stage, we extrapolate the probability that a student or instructor becomes infected as a result of attending or teaching classes over the entire semester.

We omit the scenarios where there are two or more source cases in the same classroom at the same time. Such scenarios are relatively unlikely because prevalence is low. Moreover, the fact that a few positives occur in the same classrooms at the same time makes reality slightly more optimistic than our estimates: in our estimates, everyone else in the classroom is susceptible, whereas in reality, the other infectious individual cannot be infected again. In addition, the increase in the risk of infection created by adding a second positive is smaller than the increase created by adding a first positive.[1] The level of optimism introduced by this fact is extremely small, however, and our assumption produces nearly the same estimate as one that allows multiple positives to be in the same classroom.

Generating Classroom Seating Arrangements

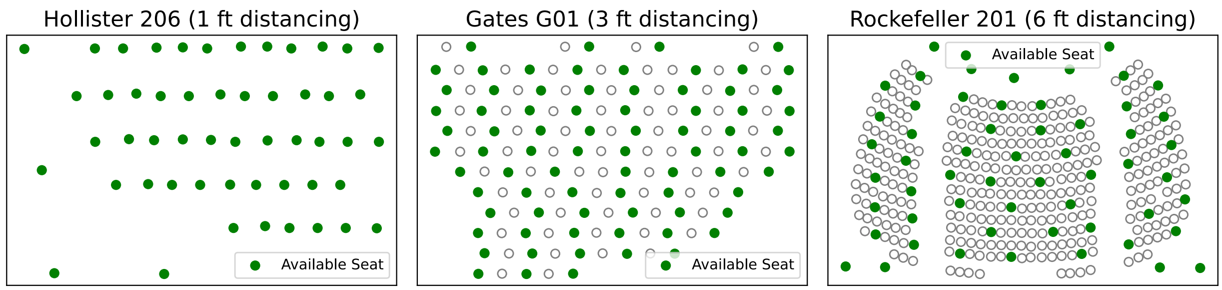
We simulate seating at different density levels by assigning a fixed number of individuals to classrooms of different sizes. We assume that an average class contains $n_0 = 50$ students and one instructor. From university floor plans, we identify three representative rooms—namely, Hollister 206, Gates G01, and Rockefeller 201—that correspond to roughly 1-, 3-, and 6-foot distancing, respectively, for 50 students. The corresponding seating capacities are presented in Table A.1.

Using the seating plan tool developed by the Committee on Teaching Reactivation Options (C-TRO) team (Greenberg et al. 2021), we identify seats in these classrooms and assign maximally distanced seats such that approximately 50 students can fit in each room. The generated seating plans are displayed in Figure A.1. The rooms used have extra space above and beyond what is required to accommodate the social distancing

Table A.1. Seating Capacity for Hollister 206, Gates G01, and Rockefeller 201

Room	Pre–COVID-19 capacity (1-ft distancing)	COVID capacity (6-ft distancing)
Hollister 206	52	12
Gates G01	156	22
Rockefeller 201	383	56

Figure A.1. (Color online) Seating Plans Generated for Hollister 206, Gates G01, and Rockefeller 201



Notes. The panels correspond to 1-, 3-, and 6-foot social distancing, respectively. The solid dots represent available seats (i.e., seats that are allowed to be occupied) and the empty circles are considered unavailable in our simulation tool.

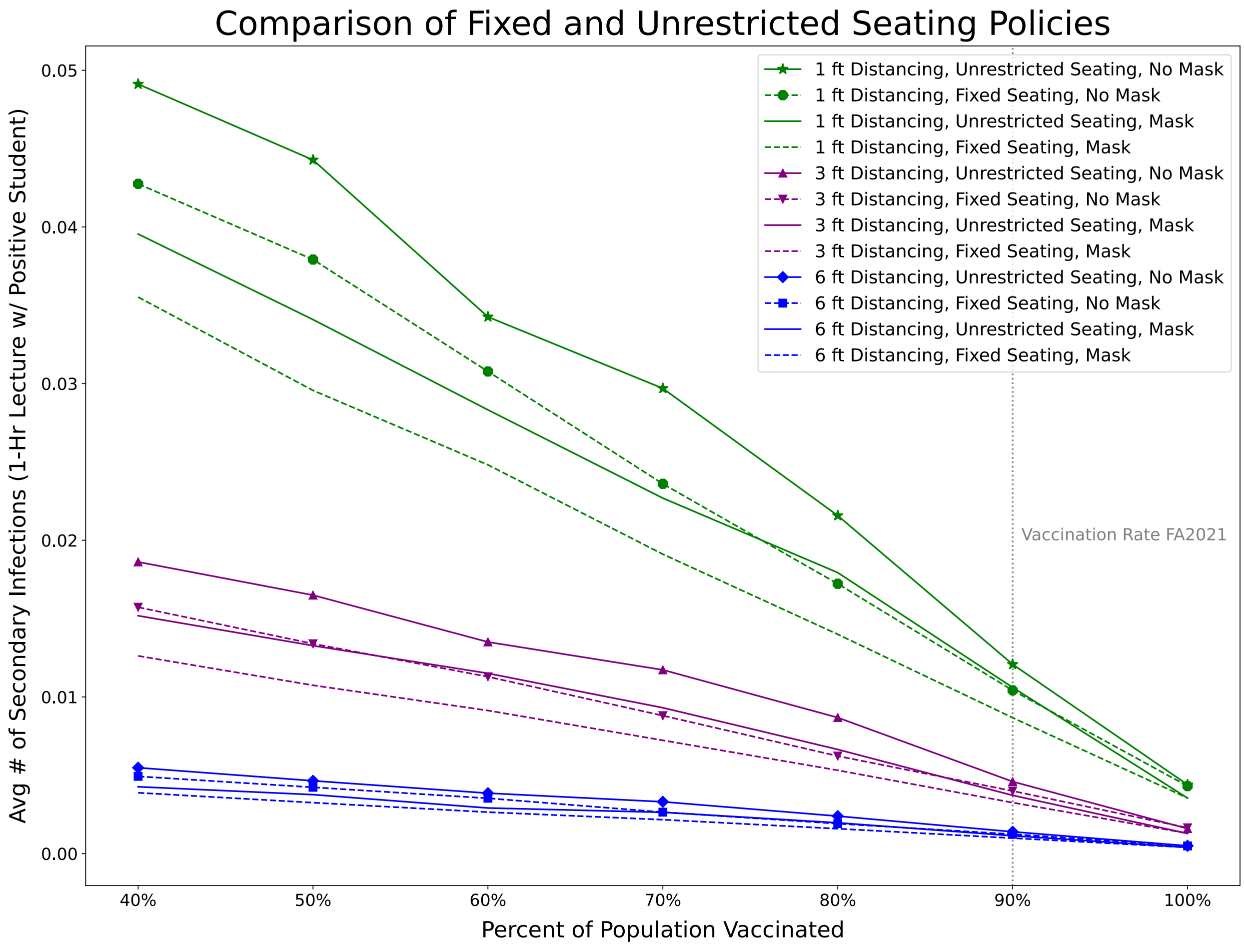
requirements we have assumed. This extra room does not offer additional benefit in our simulations as it is not used. We assume the instructor is standing in the front of the classroom, with at least 6 feet of distance separating them from all the students.

We initially consider two seating policies: (1) fixed seating, where students are randomly assigned to their own seats independent of their vaccination statuses; (2) unrestricted seating, where students can freely choose among the allowed seats. For unrestricted seating, we pessimistically assume that unvaccinated students tend to sit together, as the same demographic factors that cause students to be unvaccinated upon arrival to Ithaca may also create social connections. Because fixed seating is operationally difficult to implement, and initial simulations show that these two seating policies lead to comparable risk (Figure A.2),[2] Cornell University decided to adopt unrestricted seating. Figure A.3 shows examples of unrestricted seating arrangements at different distancing levels.

Stage 1: Simulating Infections in a Single Classroom

In the first stage, we estimate the conditional probability that a student or instructor is infected in a one-hour class given there is one positive source case in the same class. We do this by simulating the expected number of secondary infections that occur in a one-hour class with $n_0 = 50$ students, where initially a single student is infected. The simulation depends on the following configuration parameters: seating density, percentage of students vaccinated, and vaccine efficacy (VE). In particular, vaccine efficacy $VE = (v_{\text{source}}, v_{\text{susceptible}})$ is a tuple of two parameters characterizing the reduction in infectivity of a vaccinated source case and the

Figure A.2. (Color online) Initial Comparison of Fixed and Unrestricted Seating Policies



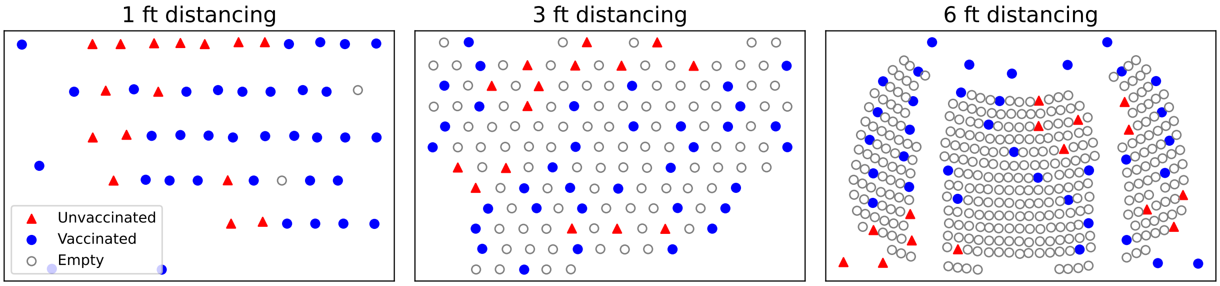
Notes. We show the average number of student secondary infections over a one-hour lecture with a positive student for different vaccination rates (from 40% to 100%), distancing levels (1-, 3-, and 6-foot distancing), and masking rates (0% and 100%). The differences between the solid lines (unrestricted seating) and dashed lines (fixed seating) are small and further decrease as vaccination rate increases. The vaccination rate we assume in our subsequent simulations, 90%, is indicated by the vertical dotted line. Here, masking is assumed to provide a 20% reduction in risk; this would be later updated to roughly range from 75% to 96% (Table A.2).

reduction in infection probability of a vaccinated susceptible person, respectively. We assume each student is vaccinated independently with probability equal to the vaccination rate among the students.

In this stage, we assume everyone is unmasked. This allows the flexibility to model varying masking rate among students as well as uncertainty in masking effectiveness in the next stage of the simulation.

We loop over all the possible combinations of seating density and VE (values of VE are discussed in detail in the Assumptions and Parameters section). For each configuration with fixed parameter values, one run of the simulation proceeds in the following steps:

Figure A.3. (Color online) Example Simulated Seating Arrangement of 50 Students at Different Levels of Social Distancing



Notes. At each distancing level, students are placed into available seats shown in Figure A.1. The triangles represent unvaccinated students, the solid dots represent vaccinated students, and the empty circles represent unavailable seats due to the distancing requirement or empty available seats. We assume pessimistically that unvaccinated students tend to sit together.

- Simulate a seating arrangement of $n_0 = 50$ students in the corresponding classroom.
- Choose one student uniformly at random as the source case and simulate their vaccination status.
- For each susceptible student, first simulate their vaccination status, then compute the probability p that they are infected over the one-hour lecture depending on their relative location to the source case. The mathematical model that produces this probability is derived in the Mathematical Model for the Risk of Transmission over Short and Long Ranges section.
- Return the sum of the individual infection probabilities in the last step divided by $(n_0 - 1)$. This is the average probability that a susceptible student becomes infected in a one-hour class given that there is an infectious student in the class, denoted $\eta_{\text{student}}(\text{density}, VE)$.
- We also compute the infection probability for the instructor, who is assumed to be subject to only the risk of aerosol transmission because of sufficient distancing, denoted $\eta_{\text{instructor}}(\text{density}, VE)$.

For each simulation configuration, we perform 500 replications to obtain the average conditional probability of infection over one hour given a positive case in the classroom.

Stage 2: Extrapolating Classroom Transmission Risk over the Semester

In the second stage, we extrapolate the probability of infection over the entire semester based on results in Stage 1 and additional parameters (prevalence, masking effectiveness, and fraction of population masked).

To deal with uncertainty in the parameter values, we impose independent prior distributions on each of them. We sample parameter configurations from the prior and compute the output at each configuration, which produces a distribution of the output. In particular, for each sampled parameter configuration (vaccine efficacy $VE = (v_{\text{source}}, v_{\text{susceptible}})$, masking effectiveness m, prevalence p) and a selected seating density, we perform the following computation for an average student:

- Get $\eta_{\text{student}}(\text{density}, VE)$ for the seating density and sampled vaccine efficacy from Stage 1 simulation results.

- Adjust for masking according to β_{masked}, the fraction of students wearing a mask, and masking effectiveness m, the reduction in transmission probability due to mask-wearing. We obtain the transmission probability adjusted for masking, conditioned on the existence of an infectious student in the class (we omit the terms (density, VE) for brevity):

$$\tilde{\eta}_{\text{student}} = \eta_{\text{student}}(\text{density}, VE) \cdot [\beta_{\text{masked}} \cdot (1-m) + (1-\beta_{\text{masked}})]. \tag{A.1}$$

This expression is an approximation in that we set m to be the reduction in transmission probability when both the source and susceptible individuals are masked, and we are not explicitly modeling the case where only one of them is masked. In practice, we assume $\beta_{\text{masked}} = 1$ in our modeling because masking compliance in previous semesters was very high (Cornell University 2021a).

- Under the assumption that infectious students are uniformly distributed across classes, the probability that a susceptible student attends class with an infectious student is given by

$$1-(1-p)^{n_0-1} \approx (n_0-1)\cdot p, \tag{A.2}$$

where the approximation is justified because prevalence, p, is small in practice; see below for further discussion of the approximation. Thus, the unconditional infection probability for a student over a one-hour lecture is given by

$$\tilde{\eta}_{\text{student}} \cdot (n_0-1) \cdot p.$$

- Extrapolate the infection probability over the semester. Suppose an undergraduate student spends on average τ_{UG} hours in class in a semester. The probability of infection over τ_{UG} hours is

$$Risk_{\text{student}} = 1-(1-\tilde{\eta}_{\text{student}} \cdot (n_0-1)\cdot p)^{\tau_{\text{UG}}} \approx \tilde{\eta}_{\text{student}} \cdot (n_0-1) \cdot p \cdot \tau_{\text{UG}}. \tag{A.3}$$

The above procedure does not eliminate already-infected students from the susceptible pool over time, so the estimated results are more pessimistic than reality. A similar computation can be done for faculty and graduate student instructors.

- We adjust for masking for instructors in a similar way as for students:

$$\tilde{\eta}_{\text{instructor}} = \eta_{\text{instructor}}(\text{density}, VE) \cdot [\beta_{\text{masked}} \cdot (1-m) + (1-\beta_{\text{masked}})].$$

- We assume faculty instructors teach a fraction β_{faculty} of all class hours, whereas graduate students teach the rest. Let n_{UG}, n_{faculty}, and n_{graduate} denote the number of undergraduates, faculty instructors, and graduate student instructors, respectively. The average hours a faculty instructor spends in class teaching in a semester is

$$\tau_{\text{faculty}} = \frac{n_{\text{UG}} \cdot \tau_{\text{UG}} \cdot \beta_{\text{faculty}}}{n_0 \cdot n_{\text{faculty}}},$$

where $n_{\text{UG}} \cdot \tau_{\text{UG}}/n_0$ gives the total undergraduate lecture hours to be taught by all instructors. Similarly, the average hours a graduate student instructor spends in class teaching in a semester is

$$\tau_{\text{graduate}} = \frac{n_{\text{UG}} \cdot \tau_{\text{UG}} \cdot (1-\beta_{\text{faculty}})}{n_0 \cdot n_{\text{graduate}}}.$$

- With the same reasoning as in Equation (A.3), we derive the risk for instructors over the semester:

$$Risk_{\text{faculty}} = 1-(1-\tilde{\eta}_{\text{instructor}} \cdot (1-(1-p)^{n_0}))^{\tau_{\text{faculty}}} \approx \tilde{\eta}_{\text{instructor}} \cdot n_0 \cdot p \cdot \tau_{\text{faculty}}.$$

$$Risk_{\text{graduate}} = 1-(1-\tilde{\eta}_{\text{instructor}} \cdot (1-(1-p)^{n_0}))^{\tau_{\text{graduate}}} \approx \tilde{\eta}_{\text{instructor}} \cdot n_0 \cdot p \cdot \tau_{\text{graduate}}.$$

In the next section, we give details on how we obtain estimates for τ_{UG}, n_{UG}, n_{graduate}, n_{faculty}, and β_{faculty}.

Quantification of approximation errors For simplicity, we use a first-order Taylor approximation in Equations (A.2) and (A.3). We quantify the approximation errors here.

For Equation (A.2), let $f(p) = 1 - (1-p)^{n_0-1}$, so that $f'(p) = df/dp = (n_0-1)(1-p)^{n_0-2}$. The first-order Taylor expansion around 0 is given by $f(p) = f(0) + f'(0)p + R_1(p) = (n_0-1)p + R_1(p)$. By Taylor's Theorem, for some c between 0 and p, the remainder $R_1(p) = f''(c)/2 \cdot p^2$. The second derivative is $f''(c) = -(n_0-1)(n_0-2)(1-c)^{n_0-3}$, so $|f''(c)| \le (n_0-1)(n_0-2)$. In our modeling, the prior distribution for the prevalence on campus is LogNormal$(-6.157, 0.413)$ with median value $\exp(-6.157) = 0.002$. Plugging in the median value for p, and noting that $n_0 = 50$, we derive the following bound for the magnitude of the remainder:

$$|R_1(p)| \le \frac{(n_0-1)(n_0-2)}{2} \cdot 0.002^2 = 0.005.$$

Indeed, we can verify that $1-(1-0.002)^{49} = 0.093$ and the approximation $49 \cdot 0.002 = 0.098$, with a 5% relative error.

For Equation (A.3), $\tilde{\eta}_{\text{student}} = 0.056$ for dense seating and 1 ACH without masking and other parameter values following Table A.2. By Taylor's Theorem, the bound for the magnitude of the remainder is

$$\frac{\tau_{\text{UG}}(\tau_{\text{UG}}-1)}{2} \cdot (\tilde{\eta}_{\text{student}} \cdot (n_0-1) \cdot p)^2 = 1.5.$$

We can verify that $1-(1-0.056 \cdot 49 \cdot 0.002)^{315} = 0.825$ and $0.056 \cdot 49 \cdot 0.002 \cdot 315 = 1.736$, so the absolute and relative approximation error is 0.911 and +105%, respectively. We acknowledge that the approximation provides a loose upper bound for the actual quantity being modeled, which is also directly computable, but we adopted the approximation values at the time of modeling. We were erring on the conservative side, producing estimates that were upper bounds of the true quantities being computed.

Assumptions and Parameters

This section presents our assumptions and estimates for the parameters used in the modeling. A summary is given in Table A.2. For parameters with high uncertainty, we design sensible prior distributions for their values rather than using a point estimate. From the joint priors on the parameters, we sample 10^5 parameter configurations and obtain a distribution for $Risk_{\text{student}}$, $Risk_{\text{faculty}}$, and $Risk_{\text{graduate}}$, respectively. We treat the median as a main point estimate and use the 5% and 95% quantiles as optimistic and pessimistic estimates. Running simulations at a large number of parameter configurations sampled from the priors enables a better understanding of how the possible outcomes are distributed.

1. **The Delta variant**

   (a) We assume the Delta variant had dominated all infections by the start of the fall 2021 semester.

   (b) The Delta variant is 2.4 times more transmissible than the nonvariant strain. The Alpha variant was approximately 50% more transmissible than the original SARS-CoV-2 (Washington et al. 2021), and Delta is approximately 60% more transmissible than Alpha (Callaway 2021). This gives a multiplicative increase of 1.5 * 1.6 = 2.4.

Table A.2. Simulation Parameters

Parameter	Value/Prior distribution
Transmissibility of the Delta variant	2.4 times that of Alpha
Respiratory activity	Breathing
$\beta_{\text{vaccinated}}$, fraction of students vaccinated	90%
v_{source}, reduction in a vac'd source's infectivity	See Table A.3
$v_{\text{susceptible}}$, reduction in a vac'd susceptible's infection prob.	See Table A.3
β_{masked}, fraction of students masked	100%
m, masking effectiveness	$\mathcal{N}(0.855, 0.0536)$, truncated to [0,1]
p, prevalence	LogNormal(−6.157, 0.413)
τ_{UG}, avg hours an undergrad spends in class in a semester	315
n_0, class size	50
n_{UG}, number of undergraduates	15,000
n_{faculty}, number of faculty instructors	850
n_{graduate}, number of graduate instructors	3,120
β_{faculty}, fraction of classes taught by faculty instructors	2/3

2. **Type of respiratory activity**

Respiratory activities of different intensities are associated with varying transmissibility. We assume breathing is the dominant type of respiratory activity for students attending lectures and that the effect of occasional speaking (e.g., asking and answering questions) is negligible. However, our simulation is able to handle activities such as talking and singing.

3. **Level of vaccination among undergraduates** ($\beta_{\text{vaccinated}}$)

By the start of the fall 2021 semester, 99% of undergraduate students and professorial faculty were fully vaccinated (Rosenberg 2021). Erring on the conservative side, we used 90% for the fraction of vaccinated undergraduates at the time the analysis was performed. We compute the infection probability for vaccinated and unvaccinated instructors separately.

4. **Vaccine efficacy** ($VE = (v_{\text{source}}, v_{\text{susceptible}})$) We base our estimates of vaccine efficacy on the literature available at the time of our modeling.

(a) For v_{source}, we assume vaccination reduces the infectivity of a source case by 0, 50%, and 71% with probabilities proportional to the sample size of the corresponding study (Brown et al. 2021, Harris et al. 2021, Levine-Tiefenbrun et al. 2021). See Table A.3.

(b) For $v_{\text{susceptible}}$, we assume vaccination reduces the risk of infection for susceptible individuals by 40%, 42%, 66%, 76%, 79%, and 88% with probability proportional to the sample size of the corresponding study (Fowlkes et al. 2021, Lopez Bernal et al. 2021, Pouwels et al. 2021, Puranik et al. 2021, Sheikh et al. 2021). See Table A.3.

Table A.3. Vaccine Efficacy Estimates in the Literature at the Time of the Analysis (Summer 2021)

	Study	Mean (%)	Sample size
$v_{\text{susceptible}}$	Pouwels et al. (2021)	40	199,411
	Puranik et al. (2021)	42	22,064
	Fowlkes et al. (2021)	66	2,840
	Puranik et al. (2021)	76	21,179
	Sheikh et al. (2021)	79	53,679
	Andrews et al. (2021)	88	15,871
v_{source}	Brown et al. (2021)	0	469
	Harris et al. (2021)	50	96,898
	Levine-Tiefenbrun et al. (2021)	71	4,938

(c) We have also applied uniform discrete distributions for both VE parameters (over three values for the vaccine efficacy in reducing viral load, and over six values for the vaccine efficacy in protecting against infections) and found the outcome to be of the same order of magnitude.

(d) Furthermore, our simulation for a single classroom requires specifying the vaccination status of the source case. This is simulated using a Bernoulli random variable with parameter $\mathbb{P}(\text{vaccinated} \mid \text{infected})$, which in turn can be deduced from $v_{\text{susceptible}}$ and $\beta_{\text{vaccinated}}$ using Bayes Rule. By definition,

$$\mathbb{P}(\text{infected} \mid \text{vaccinated}) = (1 - v_{\text{susceptible}}) \cdot \mathbb{P}(\text{infected} \mid \text{unvaccinated}).$$

By Bayes Rule,

$$\begin{aligned}\mathbb{P}(\text{vaccinated} \mid \text{infected}) &= \frac{\mathbb{P}(\text{infected} \mid \text{vaccinated}) \cdot \mathbb{P}(\text{vaccinated})}{\mathbb{P}(\text{infected} \mid \text{vaccinated})\mathbb{P}(\text{vaccinated}) + \mathbb{P}(\text{infected} \mid \text{unvaccinated})\mathbb{P}(\text{unvaccinated})} \\ &= \frac{(1 - v_{\text{susceptible}}) \cdot \beta_{\text{vaccinated}}}{1 - v_{\text{susceptible}} \cdot \beta_{\text{vaccinated}}}.\end{aligned}$$

5. **Masking** (β_{masked})

Effective July 30, 2021, Cornell University required all individuals, including fully vaccinated ones, to wear masks indoors (Cornell University COVID-19 Response 2021b). We assume perfect compliance with this mandate. Thus, we set $\beta_{\text{masked}} = 100\%$.

Jefferson et al. (2023) conducted a meta-analysis on the effectiveness of masking for reducing the spread of respiratory viruses during the COVID-19 pandemic. The results of the meta-analysis were inconclusive owing to the limits of the primary-source evidence included in the review (Soares-Weiser 2023). The included studies were conducted in various settings (e.g., hospitals, communities, and households), and masking compliance was low and the intensity and duration of interaction was high in many of these settings. On the other hand, students are usually fully compliant with masking in Cornell's classrooms. Therefore, the review conducted by Jefferson et al. (2023) would not have altered our decision to enforce masking.

6. **Masking effectiveness** (m)

We define masking effectiveness as the reduction in unmasked transmission probability due to masking. We assume two-way masking effectiveness, where both the source and susceptible are masked, to follow Normal(0.855, 0.0536), from studies on one-way masking effectiveness in the literature.

(a) We set 80% as an optimistic estimate for one-way masking effectiveness.

i. Masking of infectious individuals: Kumar et al. (2020) used fluid dynamics simulation and estimated that 12% of the airflow carrying virus particles leaks around the side of a mask. Wang et al. (2020) observed that masking by the primary case and family contacts before the primary case developed symptoms was 79% effective in reducing transmission.

ii. Masking of susceptible individuals: Konda et al. (2020) experimentally measured the filtration efficiency of masks made from different materials and found that many materials could block particles larger than 0.3 micrometers with at least 96% filtration efficiency. (Morawska et al. (2009) observed that most human expiratory activities generate droplets or aerosols with sizes larger than this.) Doung-Ngern et al. (2020) reported that people wearing a mask all the time during contact with a patient with COVID-19 are 84% less at risk of infection. Howard et al. (2021) noted that wearing masks provides additional protection by preventing touching the nose and mouth, which is another vector of transmission.

(b) We set 50% as a conservative estimate for one-way masking effectiveness. Konda et al. (2020) observed that improper mask-wearing (e.g., having a gap between the face and the mask) can result in a large decrease in filtration efficiency. Unmasking temporarily to eat or drink would also reduce the protection.

(c) Together, these imply that if masking is enforced, where the source and susceptible are both masked, transmission risk is reduced from the no-masking scenario by a multiplicative constant of $(1-80\%)^2 = 0.04$ to $(1-50\%)^2 = 0.25$. Thus, for two-way masking, we set the prior on the risk reduction factor to be Normal(0.855, 0.0536), truncated to $[0,1]$. The untruncated distribution is designed such that [0.75, 0.96] is the 95% symmetric confidence interval.

7. **Prevalence in the university population** (p)

We impose a prior distribution on prevalence, which is computed using the following procedure:

- We sample $N_{\text{infections}}$, the total number of infections over the entire semester from a distribution (details below).
- We conservatively assume each case is nonisolated and infectious for half a week, because surveillance testing requires that undergraduate students get tested twice a week on average. This is conservative because half a week is the maximum interval between tests, and Cornell took explicit measures to ensure consistent high compliance to testing (Cornell University COVID-19 Response 2021a).
- There are roughly 14 weeks in a semester. Thus, we approximate the prevalence at any time point during the semester as the total number of infected student-days divided by the total number of student-days in the semester:

$$\frac{N_{\text{infections}} \cdot 3.5 \text{ days}}{n_{\text{UG}} \cdot (14 \cdot 7) \text{ days}}.$$

This is a simplifying assumption as we do not model the temporal change of prevalence as the semester proceeds.

Before the start of the fall 2021 semester, we assumed that $N_{\text{infections}}$ follows LogNormal(6.791, 0.413). This is derived by setting the mode to be 750 and the 97.5% quantile to be 2000. Modeling results in early 2021 (Cornell COVID-19 Modeling Team 2021b) show that a variant with more than twice the transmissibility of the original strain would lead to approximately 500 student infections, under the same masking and social distancing conditions as the fall 2020 semester. To account for the relaxation of social distancing and masking (outside the classroom) in fall 2021, we increased this number by 50% and set 750 to be the mode.

By September 19, 2021, we observed 488 student cases since the start of the semester and 32 student cases in the past week. Assuming the constant rate of 32 cases/week for the remaining 11 weeks of the semester, we would then expect to see $488 + 32*11 = 840$ cases total. We repeated the simulations with $N_{\text{infections}}$ following the LogNormal(6.872, 0.372) distribution, derived by setting the mode to 840 and the 97.5% quantile to be 2000, and got similar results. At mid-October 2021, Cornell University's public COVID-19 dashboard showed 22 weekly cases (Office of Institutional Research & Planning, Cornell University 2021), so the assumed mode of 840 student cases was conservative.

• Prevalence is assumed to be proportional to $N_{\text{infections}}$ (which follows LogNormal(6.791, 0.413)), so it follows LogNormal(−6.157, 0.413), where the mean is shifted from that of $N_{\text{infections}}$ by the log of the proportionality constant.

8. **Average hours an undergraduate spends in class in a semester** (τ_{UG})

On average, students at Cornell University are enrolled for 15 credits with 45 hours of course-related work per week, including lecture time, non–lecture time in classrooms (e.g., recitation), and time spent outside of class on homework and other coursework. We assume half of the 45 hours is spent in the classroom. Over a 14-week semester, a student spends $45/2 \times 14 = 315$ hours in the classroom. Thus, we set $\tau_{\text{UG}} = \mathbf{315}$.

9. **Population sizes**

   (a) $n_{\text{UG}} = 15,000$ (Cornell University 2021b)

   (b) $n_{\text{faculty}} = 850$

The Ithaca campus has roughly 1,700 faculty members (Cornell University 2021b). We assume half of them teach undergraduate classes in a semester.

   (c) $n_{\text{graduate}} = 3,120$

In fall 2020, there were 6,239 graduate students (Office of Institutional Research & Planning, Cornell University 2022). We assume half of them work as TAs for classes or recitations. Note that some of these TAs may not interact face-to-face with students, for example, if their primary responsibility is grading, and that the risk calculated is the average risk across all TAs, including these individuals.

   (d) $\beta_{\text{faculty}} = 2/3$

On average, a course at Cornell University has three meetings per week, with two being lectures taught by faculty and one being a recitation led by a graduate student. Thus, we set this parameter to be 2/3.

10. **Sufficient distancing of instructors** We assume that faculty and graduate student instructors are sufficiently distanced from the students during lecture that their risk of infection only arises from long-distance transmission. We assume social distancing is maintained during one-on-one discussions between instructors and students and that they do not add significantly to the total interaction time between instructors and students.

Mathematical Model for the Risk of Transmission over Short and Long Ranges

In this section, we present the mathematical model for the transmission probability of COVID-19 depending on the relative location of the source case and susceptible individual. This is an important component in simulating infections in a single classroom, as described in Stage 1 in the Simulation section.

Exposure to respiratory fluids is a major mode of transmission of SARS-CoV-2. An infectious source case releases the virus through exhalation of virus-containing respiratory fluids (e.g., through speaking, coughing, or sneezing). A susceptible person becomes exposed to the virus if they inhale the virus-containing aerosols or fine droplets, if the virus particles deposit on their mucous membranes via larger droplets, or if they touch a contaminated surface and then touch their mucous membranes (Centers for Disease Control and Prevention 2021b). We consider the first two possibilities here. (Hand sanitizers and disinfecting wipes are provided in all classrooms, which reduces the risk from the third mode of transmission.) In particular, we model (1) *long-range transmission* via aerosols and fine droplets that suspend in the air and (2) *short-range transmission* via large droplets that eventually deposit after being emitted. Under the assumption that instructors are at least 6 feet away from all students, instructors are only at the risk of long-range transmission, whereas students are subject to the risk of both. The modeling of both types of transmission relies on the exponential dose-response model, introduced below.

Exponential Dose-Response Model

A dose-response model calculates the transmission probability as a function of *dose*, the amount of virus particles a susceptible person is exposed to. In the exponential dose-response model (Watanabe et al. 2010), the transmission probability given dose D takes the form

$$\mathbb{P}(\text{transmission}) = 1 - \exp(-c \cdot D), \tag{A.4}$$

where c is a positive constant. Observe that $\mathbb{P}(\text{transmission})$ is concave in the dose D, a fact used in the Simulation section to observe that the increase in risk created by adding a second positive in the classroom is smaller than the increase in risk created by the first positive.

Long-Range Transmission

We base our analysis of long-range transmission on the results from Schijven et al. (2021), who developed a model for predicting the transmission risk in an enclosed space due to aerosols only, under the assumption that emitted aerosols are dispersed across the entire room. The estimated risk depends on the aerosol-emitting activity (such as breathing, speaking, and singing), the level of ventilation of the room, the duration of interaction, and the virus concentration of the source case (measured in the number of virus particles per unit volume). One key property is that the risk due to aerosol transmission is uniform across all locations of the room, because the aerosols disperse quickly across space once emitted.

The risk of aerosol transmission over time T is estimated by the exponential dose-response model:

$$\mathbb{P}_{\text{aerosol}}(\text{transmission}, T) = 1 - \exp\left(-\frac{D(T)}{1440}\right),$$

where D denotes the *dose* (i.e., the amount of virus particles that a susceptible person receives from the infectious person over time), and 1,440 is the estimated average number of virus copies to cause illness, according to Schijven et al. (2021). The dose depends on the number of virus particles emitted over time T, which we denote $N(T)$, as follows

$$D(T) = N(T) \cdot \frac{\text{inhalation rate of the susceptible (volume / time)}}{\text{volume of the room}},$$

where $N(T)$ depends on the type of aerosol-emitting activity, the viral load of the infectious person, the ventilation condition of the room, and the duration of interaction T. It is estimated that breathing emits aerosols containing 3,300 virus RNA copies per hour (assuming a nominal viral load 10^8 copies per milliliter), whereas the value is higher for speaking and singing. We refer the readers to Equations (4) - (14) in Schijven et al. (2021) for details of the calculation.

We implement a few additional calculations when deploying this model for the classroom simulation:

- We average the risk over the distribution of the source viral load, which Schijven et al. (2021) estimated to be log-normal. In particular, we compute the weighted average of transmission risk given that the source case viral load is 10^k copies per milliliter, for $k = 5, 6, \ldots, 11$, with weights 0.12, 0.22, 0.3, 0.23, 0.103, 0.0236, and 0.0034, respectively.
- We take into account the effect of vaccination and masking, as described in the previous assumptions.
- We implement different ventilation conditions—namely, no ventilation (a conservative estimate of the amount of ventilation in naturally ventilated spaces), one air exchange per hour, and three air exchanges per hour. We model the amount of aerosols present in the room per hour as being reduced by a factor of two and four under the latter two conditions, respectively.

Short-Range Transmission

In this section, we first derive a mechanical model for the deposition of droplets over 2D space over short distances. Based on the mechanical model, we derive an expression for the amount of droplets that a susceptible person at a certain location relative to the source receives (equivalently, the amount of droplets that reach the susceptible person spatially). We then model the susceptible person's risk of infection due to exposure to viral droplets using the exponential dose-response model (Equation A.4). Finally, we estimate the model parameters from a dataset of transmissions on high speed trains in China (Hu et al. 2021). Table A.4 summarizes the notation for the functions and parameters used in developing the model.

We first make a fundamental assumption for model tractability.

Assumption 1. *The concentration of virus particles in droplets exhaled by a source case is uniform across all droplets of different sizes.*

Assumption 1 allows us to use the *volume* of viral droplets as a proxy for the amount of virus that a susceptible person is exposed to. This simplifies the calculation and allows us to better leverage existing results from the fluid dynamics literature. We next assume that the transmission of droplets is not blocked by any obstacles.

Assumption 2. *There are no obstacles between a source and a susceptible person, regardless of their locations. A susceptible person at distance r from the source receives all droplets that would deposit at distance r or further.*

Table A.4. Notation for Functions and Parameters in the Model for Transmission over Short Distances

Notation	Meaning
$\phi(r)$	Fraction of droplets that deposit at distance $\geq r$ meters from their location of emission in 1D
$\gamma_{1D}(r)$	Fraction of droplets deposited at distance r away from the source in 1D
$\gamma_{2D}(r,\theta)$	Fraction of droplets deposited at (r,θ) away from the source in the 2D model
α	Parameter defining cone of exposure over $[-\alpha, \pi+\alpha]$
$\phi_{2D,ind}(r,\theta;\alpha)$	Fraction of droplets emitted by the source case that reach an individual at (r,θ) assuming α-cone of exposure
$D(r,\theta,T)$	Dose of virus that a susceptible person at (r,θ) away from the source receives throughout interaction duration T
$N(x,y)$	Number of close contacts seated x rows and y columns away from an index case
$Y(x,y)$	Number of close contacts seated x rows and y columns away from an index case that were later confirmed as positive
$d(x,y)$	Distance between an index case and a close contact at relative location (x,y), computed from row-wise distance $d_r(x,y)$ and column-wise separation $d_c(x,y)$
$q_\alpha(x,y)$	Expected fraction of close contacts counted at (x,y) that are in the α-cone of exposure
$p_{c_2}((x,y),T \mid \text{in cone})$	Probability that a close contact in the cone of exposure at (x,y) is infected over duration T

Using fluid dynamics modeling and experimental data, Sun and Zhai (2020) estimated a function for the expected fraction of droplets that deposit at distance no less than r meters from their location of emission, assuming all droplets travel in the same direction, over a distribution of droplet sizes from a typical cough:

$$\phi(r) = -0.1819 \cdot \ln(r) + 0.43276. \tag{A.5}$$

This formula is valid for $r \in [0.04, 10.8]$ meters, at the two ends of which $\phi(r)$ is equal to 1 and 0, respectively. We let $r_{\min} = 0.04$ and $r_{\max} = 10.8$.

If the source case exhaled all droplets in one direction, then a susceptible person at distance r away in that exact direction would receive a fraction $\phi(r)$ of the droplets, whereas a susceptible person in all other directions would not receive any. In reality, however, the droplets emitted by an infectious individual may travel in multiple directions (Xie et al. 2009), putting surrounding neighbors at different angles at risk. As the droplets are being spread out in space, the 1D model in Equation (A.5) does not accurately capture the amount of droplets that reach one susceptible person in the vicinity.

Thus, we extend this 1D model to two dimensions. We design our 2D model of droplet deposition such that it depends on both the distance and angle of the susceptible person with respect to the source. To ensure our 2D model is consistent with the 1D model, we make the following assumption:

Assumption 3. *The fraction of droplets traveling beyond the entire radius-r circle around the source is the same as $\phi(r)$, the fraction beyond distance r in the 1D model.*

Now we derive the 2D model under Assumption 3. Let $\gamma_{1D}(r')$ denote the fraction of droplets deposited *at* distance r' away from the source in the 1D model. By definition,

$$\phi(r) = \int_r^{r_{\max}} \gamma_{1D}(r')dr'. \tag{A.6}$$

Let $\gamma_{2D}(r',\theta)$ denote the fraction of droplets deposited at distance r' and angle θ away from the source in the 2D model. By definition and Assumption 3,

$$\phi(r) = \int_r^{r_{\max}} \int_0^{2\pi} \gamma_{2D}(r',\theta) r' d\theta dr' = \int_r^{r_{\max}} \left[\int_0^{2\pi} \gamma_{2D}(r',\theta) d\theta \right] r' dr'. \tag{A.7}$$

The term in the bracket is the fraction of droplets that deposit over the entire circle of radius r'. From Equations (A.6) and(A.7), the 1D and 2D models should satisfy the following consistency condition:

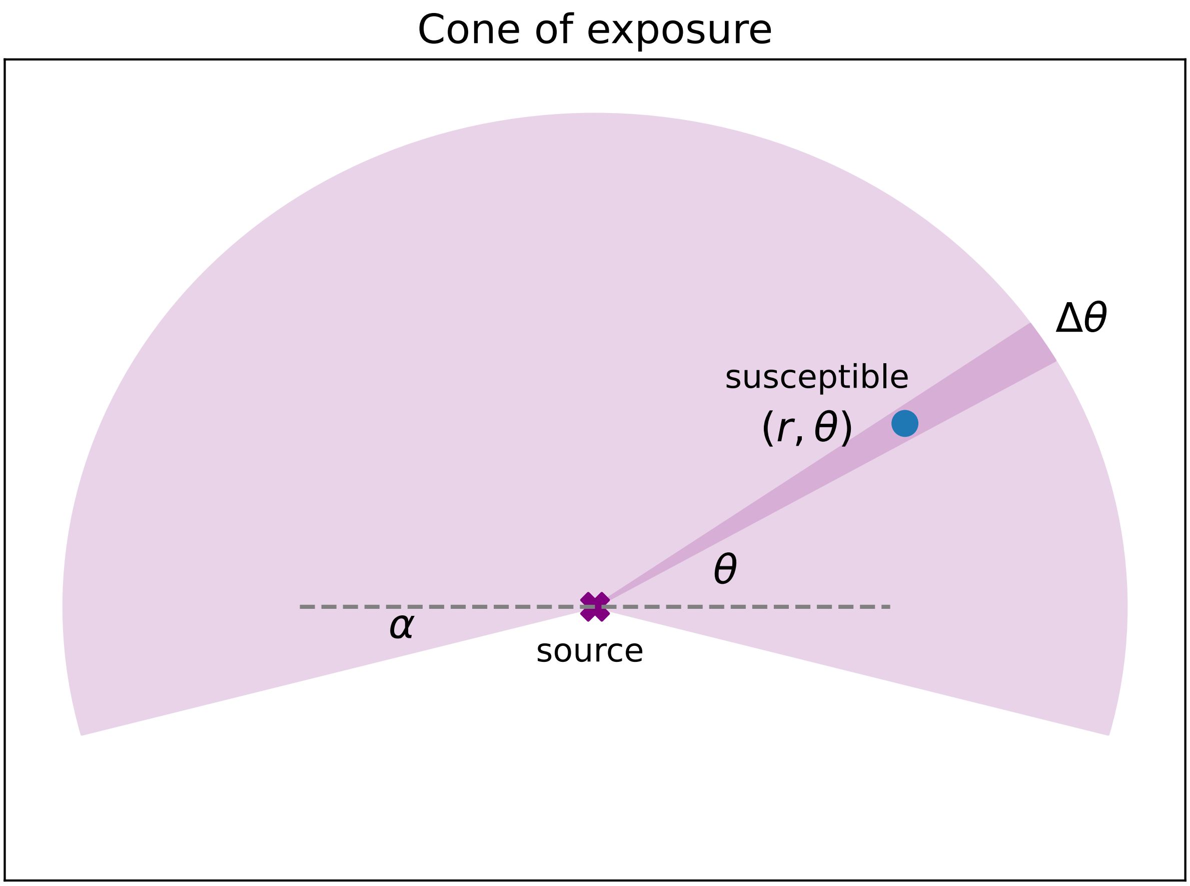
$$\int_0^{2\pi} \gamma_{2D}(r',\theta) d\theta = \frac{\gamma_{1D}(r')}{r'}. \tag{A.8}$$

Our goal is to model the transmission risk that one susceptible person at distance r and angle θ is subject to. Going from the fraction of droplets depositing over the entire circle, we next explicitly model the dependence of $\gamma_{2D}(r',\theta)$ on θ.

In realistic settings like classrooms or buses, a susceptible person could be seated in different directions from the source. We would naturally expect some directions to be riskier and others to be safer. For example, we would think of seats right behind the source as relatively safe, because the source is most likely facing and exhaling forward and the chair backs may block the droplets. On the other hand, it is possible that the source may turn their head from left to right when seated, so that droplets may potentially even reach someone sitting in rows behind them.

Based on these observations, we set up a "cone of exposure" model that quantifies the dependence of risk on angle θ. The cone of exposure, with parameter α (ranging from 0 to $\pi/2$), covers an arc of $(\pi/2+\alpha)$ degrees on both sides of the direction that the source case is facing. Hereafter, we call this an "α-cone of exposure." An illustration is given in Figure A.4. The source case is facing up and emits droplets in directions ranging from angle α behind on their left to angle α behind on their right. Susceptible cases sitting within this cone are at the risk of receiving the droplets; susceptible cases sitting outside of this cone are not at risk.

Figure A.4. (Color online) Cone of Exposure Model



Notes. The source case is represented with an "X," and the susceptible person is represented with a dot. The source emits virus uniformly over the cone extending from $-\alpha$ to $\pi + \alpha$. The susceptible person is located at distance r and angle θ away from the source. They occupy an angle of $\Delta\theta$ that scales inversely with r.

We choose the right-hand direction of the source to be of angle 0 and measure θ counterclockwise. We let $\mathbb{1}\{\theta \text{ in cone}; \alpha\}$ be an indicator of whether θ is in the α-cone of exposure—that is, $\mathbb{1}\{\theta \text{ in cone}; \alpha\} = \mathbb{1}\{\theta \in [-\alpha, \pi + \alpha]\}$. The amount of droplets depositing is positive for those in the cone and zero for those outside the cone. Next, we make a further assumption about droplet distribution within the cone.

Assumption 4. *At the same distance, droplets are distributed uniformly over all angles in the cone of exposure.*

This simplification aids the analysis, but in reality the distribution of droplets within the cone may depend on the angle in a more complicated way. We lack sufficient data to accurately estimate a complex angle-dependent droplet distribution model, and so a complex model might degrade accuracy rather than improve it. We thus adopt Assumption 4 in the spirit of regularization, understanding that the model we adopt approximates a more complex angle-dependent distribution model by replacing small droplet densities by 0 and large droplet densities by a constant.

With this cone-of-exposure model set up, we can model how the transmission risk depends on θ. We assume all individuals have comparable width. Then, for an individual at distance r and angle θ from the source, they occupy an arc whose central angle $\Delta\theta(r)$ is approximately inversely proportional to r and they receive all the droplets that would land in the sector across $[\theta, \theta + \Delta\theta(r)]$ and $[r, r_{\max}]$. Based on this insight, we derive $\phi_{2D,ind}(r,\theta;\alpha)$, the fraction of droplets emitted by the source case that reach such an individual. We call this quantity the *droplet reception factor*.

$$\begin{aligned}
\phi_{2D,ind}(r,\theta;\alpha) &= \mathbb{1}\{\theta \text{ in cone}; \alpha\} \int_r^{r_{\max}} \int_\theta^{\theta+\Delta\theta(r)} \gamma_{2D}(r',\theta) r' d\theta dr' \\
&= \mathbb{1}\{\theta \text{ in cone}; \alpha\} \int_r^{r_{\max}} \int_0^{2\pi} \gamma_{2D}(r',\theta) r' d\theta dr' \cdot \frac{\Delta\theta(r)}{\pi + 2\alpha} \\
&= \mathbb{1}\{\theta \text{ in cone}; \alpha\} \int_r^{r_{\max}} \gamma_{1D}(r') dr' \cdot \frac{\Delta\theta(r)}{\pi + 2\alpha} \\
&= \mathbb{1}\{\theta \text{ in cone}; \alpha\} \cdot \phi(r) \cdot \frac{\Delta\theta(r)}{\pi + 2\alpha} \\
&\propto \mathbb{1}\{\theta \text{ in cone}; \alpha\} \cdot \phi(r) \cdot \frac{1}{r},
\end{aligned} \tag{A.9}$$

where the second equality follows from Assumption 4, the third equality follows from Equation (A.8), and the fourth equality follows from Equation (A.6). Later, we will show that having an undetermined proportionality constant does not affect our results as long as the dependence on r and θ is modeled correctly.

Transmission probability calculation Given the expression for the droplet reception factor, we translate this to the probability of transmission using the exponential dose-response model in Equation (A.4). The dose that a susceptible individual at (r,θ) from the source receives is proportional to the fraction of the source's virus particles that they receive. By Assumption 1, this in turn is proportional to the droplet reception factor $\phi_{2D,ind}(r,\theta;\alpha)$.

Next, we observe that the dose is larger if the source case and susceptible person maintain the same relative location longer. We call this the *duration of interaction*, denoted T. For example, T is roughly one hour for a lecture. We make the following assumption about the droplet emission rate over time.

Assumption 5. *The amount of droplets a source case emits per unit time is constant.*

Under Assumption 5, the amount of droplets emitted, and hence the dose, is proportional to the duration of interaction T. Thus, the dose of virus particles that a susceptible person at (r,θ) away from the source receives can be expressed as

$$D(r,\theta,T) = c_1 \cdot \phi_{2D,ind}(r,\theta;\alpha) \cdot T,$$

where c_1 captures the proportionality of the dose to the droplet reception factor and the duration of interaction. We further absorb into c_1 the proportionality relation within the droplet reception factor (Equation A.9), yielding another constant c_2, and derive the final expression of the transmission probability for a susceptible person at (r,θ) away from a source case for a duration of interaction T:

$$\mathbb{P}_{\text{droplet}}(\text{transmission}, r, \theta, T; \alpha, c_2) = 1 - \exp\left(-c_2 \cdot \mathbb{1}\{\theta \text{ in cone}; \alpha\} \cdot \frac{\phi(r)}{r} \cdot T\right). \tag{A.10}$$

As a sanity check, we can see that if a susceptible person at (r,θ) is not in the cone of exposure, the transmission probability is 0.

Table A.5. Number of Cases and Total Number of Passengers Who Cotraveled with an Index Patient (Hu et al. (2021), table S1)

Rows apart	Columns apart					
	0	1	2	3	4	5
0	—	92/2,605	33/1,996	7/1,845	7/1,825	3/1,028
1	10/4,791	12/5,084	5/3,664	3/3,464	1/3,525	1/1,872
2	11/4,386	8/4,751	8/3,429	5/3,212	3/3,250	3/1,769
3	2/4,026	2/4,395	4/3,110	3/2,945	3/2,970	1/1,589

Note. Entries are $Y(x,y)/N(x,y)$, where x is the number of rows apart and y is the number of columns apart.

Parameter estimation The goal of this section is to first derive the likelihood for an empirically observed data set based on the model above and then find values of α and c_2 that maximize the likelihood.

Hu et al. (2021) studies 2,334 confirmed positive cases ("index cases") and 72,093 close contacts who had cotravel times of zero to eight hours from 12/19/2019 through 3/6/2020 on high-speed trains in China. They examine the association of attack rate with the spatial distance between pairs of index cases and close contacts. Here, a "close contact" was defined as a person who had cotraveled on a train within a three-row seat distance of an index case within 14 days before symptom onset. We treat a close contact as equivalent to a susceptible individual in our model. Table A.5 reports the number of close contacts, and, among them, those that were later confirmed as positive, at different seat locations with respect to an index case, within the period of study. We would like to derive the likelihood of observing the data in Table A.5.

We now introduce additional notation for formalizing our likelihood model. Let (x,y) denote the seat at x rows and y columns away from an index case, where x ranges from 0 to 3 and y ranges from 0 to 5. Let $N(x,y)$ denote the number of close contacts that are seated x rows and y columns away from an index case (hereafter we call this "at relative location (x,y)" for abbreviation). Let $Y(x,y)$ denote the number of close contacts at relative location (x,y) from an index case that were later confirmed as positive. Based on the train cabin layout given in Figure 1 in Hu et al. (2021), we calculate the separation between an index and a close contact at relative location (x,y). In particular, let $d_r(x,y)$ and $d_c(x,y)$ denote the row-wise and column-wise distance in meters. For all y, $d_r(x,y)=0.9x$; for all x, $d_c(x,y)$ is equal to 0, 0.5, 1.05, 1.6, 2.1, and 2.6 for $y=0,\dots,5$, respectively. We then calculate $d(x,y)=\sqrt{d_r(x,y)^2+d_c(x,y)^2}$ (Table A.6).

The data do not contain information about which *direction* the close contacts were seated with respect to the index cases. However, directionality information is crucial for our modeling. Thus, we make the following assumption about the symmetry of distribution of close contacts.

Assumption 6. *A close contact counted at (x,y) is equally likely to have been seated in all possible directions at location (x,y) away from the source.*

With slight abuse of notation, we let $(+x,+y)$ and $(+x,-y)$ denote the seats x rows in front of the index case and y columns to the right and left, respectively. Similarly, we let $(-x,\pm y)$ denote the seats x rows behind the index case and y columns to the right or left. Because we assume the cone of exposure has an

Table A.6. Distance $d(x,y)$ in Meters, Where x Is the Number of Rows Apart and y Is the Number of Columns Apart

Rows apart	Columns apart					
	0	1	2	3	4	5
0	—	0.5	1.05	1.6	2.1	2.6
1	0.9	1.03	1.38	1.84	2.28	2.75
2	1.8	1.87	2.08	2.41	2.77	3.16
3	2.7	2.75	2.90	3.14	3.42	3.75

angle larger than π, the seats $(+x,\pm y)$ are always in the cone of exposure. The seats $(-x,\pm y)$ are in the cone of exposure if and only if $\arctan(d_r(x,y)/d_c(x,y)) \le \alpha$.

Let $q(x,y;\alpha)$ denote the expected fraction of close contacts counted at (x,y) (which could be at $(\pm x,\pm y)$) that are in the α-cone of exposure. Based on Assumption 6,

$$q(x,y;\alpha) = \frac{1}{2} + \frac{1}{2}\mathbb{1}\{\arctan(d_r(x,y)/d_c(x,y)) \le \alpha\}. \tag{A.11}$$

We can calculate $q(x,y;\alpha)$ for all possible (x,y) pairs using Table A.6.

Next, let $p((x,y),T \mid \text{in cone}; c_2)$ denote the probability that a close contact in the cone of exposure at location (x,y) is infected. Based on Equation (A.10), we model this as

$$p((x,y),T \mid \text{in cone}; c_2) = 1 - \exp\left(-c_2 \cdot \frac{\phi(d(x,y))}{d(x,y)} \cdot \kappa_{\text{mask}} \cdot T\right), \tag{A.12}$$

where we include an additional factor of masking effectiveness κ_{mask}, owing to the fact that wearing a mask can reduce the amount of virus that a susceptible person is actually *exposed to* (compared with the virus in the droplets that *reach* where the person is) and that mask-wearing had been quite prevalent since late January of 2020 in China. Konda et al. (2020) estimated that a poorly fitting mask made of cotton or silk will reduce virus dose by approximately 30%. We heuristically select $\kappa_{\text{mask}} = 0.8$, assuming that the data set involved a mix of both masked and unmasked passengers.[3] We keep this constant separate from c_2 so that c_2 solely captures the way transmission probability depends on the unreduced dose. We set T to be 2.1 hours. This is the mean cotravel time over all pairs of index cases and close contacts in the data. Unfortunately, no information is given about the cotravel time for each individual pair.

We let $r(x,y;\alpha,c_2)$ denote the overall transmission probability at relative location (x,y) under parameters α, c_2. This is the product of the probability that a close contact at (x,y) is in the cone of exposure (Equation A.11) and the conditional probability that they become infected given that they are in the cone (Equation A.12):

$$r(x,y;\alpha,c_2) = q(x,y;\alpha) \cdot p((x,y),T \mid \text{in cone}; c_2).$$

For the likelihood model, we assume the number of transmissions at each (x,y) independently follows a binomial distribution:

$$Y(x,y) \sim \text{Binomial}(N(x,y), r(x,y;\alpha,c_2)).$$

Assuming the numbers of transmissions at each (x, y) are independent, the likelihood of all observations $\mathcal{D} := \{N(x,y), Y(x,y)\}_{\substack{x=0,\dots,3\\y=0,\dots,5}}$ is given by:

$$\mathcal{L}(\mathcal{D}) = \prod_{(x,y)} \binom{N(x,y)}{Y(x,y)} \cdot r(x,y;\alpha,c_2)^{Y(x,y)} \cdot (1 - r(x,y;\alpha,c_2))^{N(x,y)-Y(x,y)}.$$

We compute the log-likelihood and let c_3 denote the constant term that does not depend on α or c_2:

$$\ell(\mathcal{D}) = \sum_{(x,y)} [Y(x,y)\log(r(x,y;\alpha,c_2)) + (N(x,y) - Y(x,y))\log(1 - r(x,y;\alpha,c_2))] + c_3. \tag{A.13}$$

We next find values of α and c_2 that maximize the log-likelihood $\ell(\mathcal{D})$. Using a discretized grid search, we find that the log likelihood is maximized at $c_2 = 0.0135$ and multiple values of α. We choose the largest possible value $\alpha = 15$ degrees, or 0.26 radians, with the intention of being conservative.

Combining the Short- and Long-Range Transmission Risk

In our simulation, the risk for a susceptible individual is

$$\max\left(\mathbb{P}_{\text{droplet}}(\text{transmission}), \mathbb{P}_{\text{aerosol}}(\text{transmission})\right),$$

that is, the larger of the predicted risk due to short and long-range transmission. This is justified by the fact that we inferred the parameters for the droplet model assuming all infections in the data set result from short-range transmission; as such, the inferred model implicitly accounts for long-range transmission in the data set. In practice, the short-range transmission risk only dominates the long-range transmission risk at short distances (approximately within 3 meters), whereas the risk due to aerosol is uniform across all locations.

Finally, we recall that the computed risk so far is based on studies on the original virus strain. Thus, we multiply the calculated risk by 2.4 to account for the increased transmissibility of the Delta variant, as described in the Assumptions and Parameters section.

References for the Appendix

Brown CM, Vostok J, Johnson H, Burns M, Gharpure R, Sami S, Sabo RT, Hall N, Foreman A, Schubert PL, Gallagher GR, Fink T, Madoff LC, Gabriel SB, MacInnis B, Park DJ, Siddle KJ, Harik V, Arvidson D, Brock-Fisher T, Dunn M, Kearns A, Laney AS (2021) Outbreak of SARS-CoV-2 infections, including COVID-19 vaccine breakthrough infections, associated with large public gatherings — Barnstable County, Massachusetts, July 2021. *Morbidity and Mortality Weekly Report* 70(31):1059.

Callaway E (2021) Delta coronavirus variant: scientists brace for impact. *Nature* 595(7865):17–18.

Centers for Disease Control and Prevention (2021) Scientific brief: SARS-CoV-2 transmission. `https://archive.cdc.gov/#/details?url=https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/sars-cov-2-transmission.html`, Accessed: April 20, 2024.

Cornell COVID-19 Modeling Team (2021) Mathematical Modeling for Cornell's Spring Semester. `https://web.archive.org/web/20220611030559/https://covid.cornell.edu/_assets/files/general-audience-spring-modeling-20210216.pdf`, Accessed: April 20, 2024.

Cornell University (2021a) COVID-19 Faculty and Staff Town Hall, August 11, 2021. `https://www.cornell.edu/video/covid-19-faculty-staff-town-hall-august-11-2021`, Accessed: August 21, 2022.

Cornell University (2021b) University Facts: Cornell by the numbers. `https://www.cornell.edu/about/facts.cfm`, Accessed: August 28, 2022.

Cornell University COVID-19 Response (2021a) Impact of students' testing non-compliance and positive cases on your classes. `https://web.archive.org/web/20210611105403/https://covid.cornell.edu/updates/20210331-noncompliance-impact.cfm`, Accessed: April 20, 2024.

Cornell University COVID-19 Response (2021b) Indoor mask requirement for all individuals. `https://web.archive.org/web/20231129124449/https://covid.cornell.edu/updates/20210730-indoor-masks.cfm`, Accessed: March 22, 2024.

Fowlkes A, Gaglani M, Groover K, Thiese MS, Tyner H, Ellingson K, Cohorts HR (2021) Effectiveness of COVID-19 vaccines in preventing SARS-CoV-2 infection among frontline workers before and during B.1.617.2 (Delta) variant predominance — eight US locations, December 2020–August 2021. *Morbidity and Mortality Weekly Report* 70(34):1167.

Greenberg KK, Hensel T, Zhu Q, Aarts S (2021) An automated tool for optimal classroom seating assignment with social distancing constraints. *IIE Annual Conference. Proceedings*, 429–434 (Institute of Industrial and Systems Engineers (IISE)).

Harris RJ, Hall JA, Zaidi A, Andrews NJ, Dunbar JK, Dabrera G (2021) Effect of vaccination on household transmission of SARS-CoV-2 in England. *New England Journal of Medicine* 385(8):759–760.

Hu M, Lin H, Wang J, Xu C, Tatem AJ, Meng B, Zhang X, Liu Y, Wang P, Wu G, Xie H, Lai S (2021) Risk of coronavirus disease 2019 transmission in train passengers: an epidemiological and modeling study. *Clinical Infectious Diseases* 72(4):604–610.

Jefferson T, Del Mar CB, Dooley L, Ferroni E, Al-Ansary LA, Bawazeer GA, van Driel ML, Nair NS, Jones MA, Thorning S, Conly JM (2023) Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Database of Systematic Reviews* (1).

Konda A, Prakash A, Moss GA, Schmoldt M, Grant GD, Guha S (2020) Aerosol filtration efficiency of common fabrics used in respiratory cloth masks. *ACS Nano* 14(5):6339–6347.

Levine-Tiefenbrun M, Yelin I, Katz R, Herzel E, Golan Z, Schreiber L, Wolf T, Nadler V, Ben-Tov A, Kuint J, Gazit S, Patalon T, Chodick G, Kishony R (2021) Initial report of decreased SARS-CoV-2 viral load after inoculation with the BNT162b2 vaccine. *Nature Medicine* 27(5):790–792.

Lopez Bernal J, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, Stowe J, Tessier E, Groves N, Dabrera G, Myers R, Campbell CN, Amirthalingam G, Edmunds M, Zambon M, Brown KE, Hopkins S, Chand M, Ramsay M (2021) Effectiveness of Covid-19 vaccines against the B.1.617.2 (Delta) variant. *New England Journal of Medicine* 385(7):585–594.

Office of Institutional Research & Planning, Cornell University (2021) Monitoring Public Health (MPH) Dashboard. `https://web.archive.org/web/20211016002700/https://covid.cornell.edu/testing/dashboard/`, Accessed: April 20, 2024.

Office of Institutional Research & Planning, Cornell University (2022) University Factbook - Student Enrollment. `https://irp.dpb.cornell.edu/university-factbook/student-enrollment`, Accessed: January 8, 2023.

Pouwels KB, Pritchard E, Matthews PC, Stoesser N, Eyre DW, Vihta KD, House T, Hay J, Bell JI, Newton JN, Farrar J, Crook D, Cook D, Rourke E, Studley R, Peto TEA, Diamond I, Walker AS (2021) Effect of Delta variant on viral burden and vaccine effectiveness against new SARS-CoV-2 infections in the UK. *Nature Medicine* 27(12):2127–2135.

Puranik A, Lenehan PJ, Silvert E, Niesen MJ, Corchado-Garcia J, O'Horo JC, Virk A, Swift MD, Halamka J, Badley AD, Venkatakrishnan AJ, Soundararajan V (2021) Comparison of two highly-effective mRNA vaccines for COVID-19 during periods of Alpha and Delta variant prevalence. MedRxiv.

Rosenberg M (2021) Vaccination, testing, masking: Here's what you need to know for Fall 2021. `https://cornellsun.com/2021/08/16/vaccination-testing-masking-heres-what-you-need-to-know-for-fall-2021/`, Accessed: August 28, 2022.

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

Sheikh A, McMenamin J, Taylor B, Robertson C (2021) SARS-CoV-2 Delta VOC in Scotland: demographics, risk of hospital admission, and vaccine effectiveness. *The Lancet* 397(10293):2461–2462.

Soares-Weiser K (2023) Statement on "physical interventions to interrupt or reduce the spread of respiratory viruses" review. `https://www.cochrane.org/news/statement-physical-interventions-interrupt-or-reduce-spread-respiratory-viruses-review`.

Washington NL, Gangavarapu K, Zeller M, Bolze A, Cirulli ET, Schiabor Barrett KM, Larsen BB, Anderson C, White S, Cassens T, Jacobs S, Levan G, Nguyen J, Ramirez JM, Rivera-Garcia C, Sandoval E, Wang X, Wong D, Spencer E, Robles-Sikisaka R, Kurzban E, Hughes LD, Deng X, Wang C, Servellita V, Valentine H, De Hoff P, Seaver P, Sathe S, Gietzen K, Sickler B, Antico J, Hoon K, Liu J, Harding A, Bakhtar O, Basler T, Austin B, MacCannell D, Isaksson M, Febbo PG, Becker D, Laurent M, McDonald E, Yeo GW, Knight R, Laurent LC, de Feo E, Worobey M, Chiu CY, Suchard MA, Lu JT, Lee W, Andersen KG (2021) Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *Cell* 184(10):2587–2594.e7, ISSN 0092-8674.

Watanabe T, Bartrand TA, Weir MH, Omura T, Haas CN (2010) Development of a dose-response model for SARS coronavirus. *Risk Analysis: An International Journal* 30(7):1129–1138.

Xie X, Li Y, Sun H, Liu L (2009) Exhaled droplets due to talking and coughing. *Journal of the Royal Society Interface* 6(Suppl 6):S703–S714.

Endnotes

[1] This follows (1) when the dose resulting from each positive is independent and identically distributed and (2) from concavity of the probability of infection as a function of the dose, given in the Mathematical Model for the Risk of Transmission over Short and Long Ranges section. In particular, let V_1 and V_2 be the (strictly positive) dose to a given susceptible person associated with the first and second positives in a classroom, so that the dose is V_1 if there is one positive in the classroom and $V_1 + V_2$ if there are two. We assume that V_1 and V_2 are independent and identically distributed after marginalizing over the random locations of the two positive individuals. Then let $P(v)$ be the probability of infection given a dose v, as given in the Mathematical Model for the Risk of Transmission over Short and Long Ranges section. Because P is concave, and also using that $P(0) = 0$, then for strictly positive V_1 and V_2, $P(V_1 + V_2) - P(V_1) \leq P(V_2) - P(0)$. Then, because V_1 and V_2 are identically distributed, $\mathbb{E}[P(V_1 + V_2) - P(V_1)] \leq \mathbb{E}[P(V_2) - P(0)] = \mathbb{E}[P(V_2)] = \mathbb{E}[P(V_1)]$. The left-hand side $\mathbb{E}[P(V_1 + V_2) - P(V_1)]$ is the increase in the risk of infection created by adding a second positive, and the right-hand side $\mathbb{E}[P(V_1)]$ is the increase in risk from adding the first positive.

[2] Figure A.2 was generated at the project's outset using outdated parameters. Back then, we did not have a full understanding of masking efficacy, so we assumed a 20% reduction in risk due to masking.

[3] We constructed this model prior to conducting the simulations described in the Simulation and the Assumptions and Parameters sections, so we used a point estimate for masking effectiveness. Nevertheless, the point estimate of 0.8 is well-aligned with the prior choice for masking effectiveness parameter m, discussed in the Assumptions and Parameters section.