Original Paper

Predicting COVID-19 Transmission to Inform the Management of Mass Events: Model-Based Approach

Claire Donnat[1], BSc, MSc, PhD; Freddy Bunbury[2], BA, MRES, PhD; Jack Kreindler[3], MD; David Liu[1], BSc; Filippos T Filippidis[3], MPH, MSc, MD, PhD; Tonu Esko[4], PhD; Austen El-Osta[3], BSc, MSc, MPA, PhD; Matthew Harris[3], MBBS, MSc, FFPH, PGCE, DPhil

[1]Department of Statistics, University of Chicago, Chicago, IL, United States

[2]Department of Plant Biology, Carnegie Institution for Science, Stanford, CA, United States

[3]Faculty of Medicine, School of Public Health, Imperial College, London, United Kingdom

[4]Institute of Genomics, University of Tartu, Tartu, Estonia

Corresponding Author:
Claire Donnat, BSc, MSc, PhD
Department of Statistics
University of Chicago
5747 South Ellis Avenue
Chicago, IL, 60637
United States
Phone: 1 773 702 9890
Email: cdonnat@uchicago.edu

Abstract

Background: Modelling COVID-19 transmission at live events and public gatherings is essential to controlling the probability of subsequent outbreaks and communicating to participants their personalized risk. Yet, despite the fast-growing body of literature on COVID-19 transmission dynamics, current risk models either neglect contextual information including vaccination rates or disease prevalence or do not attempt to quantitatively model transmission.

Objective: This paper attempted to bridge this gap by providing informative risk metrics for live public events, along with a measure of their uncertainty.

Methods: Building upon existing models, our approach ties together 3 main components: (1) reliable modelling of the number of infectious cases at the time of the event, (2) evaluation of the efficiency of pre-event screening, and (3) modelling of the event's transmission dynamics and their uncertainty using Monte Carlo simulations.

Results: We illustrated the application of our pipeline for a concert at the Royal Albert Hall and highlighted the risk's dependency on factors such as prevalence, mask wearing, and event duration. We demonstrate how this event held on 3 different dates (August 20, 2020; January 20, 2021; and March 20, 2021) would likely lead to transmission events that are similar to community transmission rates (0.06 vs 0.07, 2.38 vs 2.39, and 0.67 vs 0.60, respectively). However, differences between event and background transmissions substantially widened in the upper tails of the distribution of the number of infections (as denoted by their respective 99th quantiles: 1 vs 1, 19 vs 8, and 6 vs 3, respectively, for our 3 dates), further demonstrating that sole reliance on vaccination and antigen testing to gain entry would likely significantly underestimate the tail risk of the event.

Conclusions: Despite the unknowns surrounding COVID-19 transmission, our estimation pipeline opens the discussion on contextualized risk assessment by combining the best tools at hand to assess the order of magnitude of the risk. Our model can be applied to any future event and is presented in a user-friendly RShiny interface. Finally, we discussed our model's limitations as well as avenues for model evaluation and improvement.

XSL•FO
RenderX

Introduction

Background

Evaluating the Safety of Live Events

More than a year after a global and unprecedented cancellation of live events in March 2020, the future of live events and the entertainment industry remains uncertain despite increasing vaccination rates and low community prevalence levels (at the time of writing). The main concern raised by these gatherings lies in their susceptibility to "super-spreading"—a scenario whereby a few contagious participants inadvertently infect a disproportionately large number of others [1-6] and that has been highlighted as a significant driver of the pandemic [7-10]. Despite the re-opening of live events in the United Kingdom on July 19, 2021, the threat of existing and emergent COVID-19 variants coupled with dwindling immunity from vaccination over time suggests that policy makers and event organizers will likely continue to struggle with the following 2 questions: (1) Is the COVID-19 transmission risk posed by these events tolerable? and (2) What additional safety measures can be feasibly deployed to reduce this risk?

The answer to these questions is inherently tied to the estimation of 2 quantities: the number of infections occurring at the event and the postevent secondary attack rate, or number of subsequent infections in the participants' social circles. Evaluating the safety (or lack thereof) of large public gatherings can then be reframed as quantifying the significance and magnitude of their effect on the distribution of the number of primary and secondary COVID-19 cases. Yet, despite the growing body of literature on COVID-19 risk evaluation and recent efforts to evaluate the safety of live events, this effect remains ill-characterized. Nevertheless, over the past several months, several calculators were developed to estimate this risk [11-14]. These methods can typically be placed in 1 of 3 categories: ranking heuristics, context-based heuristics, and transmission risk calculators.

Ranking Heuristics

These estimators typically rank events on a scale ranging from "low" risk to "high" risk based on the feedback of medical experts [13,15-17]. However, these heuristics do not take into account contextual information, including the prevalence. For example, the risk associated with an event would be classified as high regardless of whether it was held in August 2020 (background prevalence of 1 in 3000 individuals in the United Kingdom) or January 2021 (prevalence of 1 in 60 individuals [18]).

Context-Based Heuristics

These calculators estimate the probability of encountering 1 COVID-19 case based on the number of people attending an event [11,12]. While more context-aware than risk assessment charts, such estimators do not attempt to model transmission dynamics—which is undeniably one of the main unknowns in the spread of viral epidemics—and consequently rarely stratify risk by type of activity. To exemplify, a classical music recital of 1.5 hours for the BBC Proms would potentially be considered equally risky to a 3-hour concert in which participants could be expected to sing along.

Transmission Risk Calculators

Stemming from physics or fluid dynamics, these calculators focus on modelling the aerosolization and spread of microdroplets—typically in a closed or indoor environment [19-22]. These fine-grained models thus must be combined with extensive and often prohibitive simulations of crowd movements in order to model transmission dynamics during any given event.

Limitations of Existing Estimators

Regardless of their category, most of these models rely on a large number of input parameters, including (but not restricted to) the prevalence of the disease. While certain calculators attempt to bridge the gap between expert heuristics and physical models [11,23], they are not capable of predicting the risk of a future event. Moreover, all of these estimators provide point estimates—in other words, their output is a single number to quantify the risk. Given the uncertainty associated with all the inputs and the parametrization of the problem as well as the high stochasticity of viral transmission, the provision of a single consolidated outcome or number can potentially be misleading. This is because a singular focus on the expected outcome precludes consideration of the distribution of all possible outcomes, including worst-case scenarios. In the context of COVID-19, where the majority of new cases has been shown to be caused by a minority of index cases [24-26], the modelling of tail events and potential super-spreader phenomena takes on significant importance for risk assessment [26,27].

Mitigating Transmission Risk

Meanwhile, with the increasing vaccination rates in several countries around the world, a few initiatives have begun to evaluate the outbreak risk associated with live events empirically [28-31]. This is because vaccinated individuals may still be infected with SARS-CoV-2 [32,33], and even antigen-test based screening of ticket holders offers no guarantee due to false negatives [34,35]. The estimation of what constitutes an admissible level of risk thus poses a difficult conundrum to the live event industry. To begin answering these questions, the CAPACITY study [36]—a partnership between Certific (a private, remote testing, health status, and identify certification service) and Imperial College London—aims to predict and measure the outcomes of full capacity live events while ensuring rigorous implementation and alignment to current public health and recommended safety measures. Central to this study is the provision of a streamlined and efficient pre-event screening protocol of all ticket holders using professionally witnessed rapid at-home antigen tests followed by postevent monitoring based on antigen tests, surveys, and safety recommendations (see Multimedia Appendix 1). In this setting, providing risk estimates not only becomes essential in communicating to the ticket holders their own level of risk so that they may make an informed decision of whether to attend the event but also necessary to inform event managers and policy makers on the likelihood of an outbreak task that serves here as the motivating application behind this paper.

A Working Example: Concert at the Royal Albert Hall

In order to understand and illustrate the potential challenges that arise in the risk estimation for the CAPACITY study, we

considered as an example a concert at the Royal Albert Hall (RAH) and demonstrate how to estimate the associated risk assuming a near capacity attendance of 5000 in the main concert hall, which has a volume of 86,650 m^3 [37], with a dwell time of 3 hours. Attendees will be assumed to be a cross-section representative of the general British public and will be required to have a negative COVID-19 antigen test result within 2 days prior to the event, as well as satisfying other self-declared symptoms and exposure-risk questions. Vaccination status would be requested, but not required, for attendance, and full compliance with mask wearing was assumed in our default example.

Goals and Contributions

The objectives of our modelling approach were threefold: (1) enable the quantitative comparison of different activities and event characteristics, (2) estimate the efficacy of various safety protocols, and (3) provide a predictive risk assessment (ie, the risk associated with a scheduled future event). To this end, we delineated our approach into 3 sequential steps (see Figure 1): (1) estimating the number of contagious participants, (2) evaluating the transmission dynamics, and (3) comparing the risk of holding the event with the null model (ie, if the event had not taken place). We illustrated the application of our risk modelling pipeline in the RAH example to highlight the risk's dependency on factors such as prevalence, mask wearing, number of attendees, and event duration. In particular, we demonstrated how this particular event held on 3 different dates corresponding to 3 distinct COVID-19 prevalence regimes in

the United Kingdom (stable low prevalence: August 20, 2020; high prevalence peak: January 20, 2021; medium declining prevalence: March 20, 2021) would likely lead to transmission events that were on par with community transmission rates (0.06 vs 0.07, 2.38 vs 2.39, and 0.67 vs 0.60, respectively; see Table 1). However, the 99th percentile of the prediction interval for the infections at the event would likely be substantially higher than the background rate (1 vs 1, 19 vs 8, and 6 vs 3, respectively), further demonstrating that sole reliance on vaccination and antigen testing to gain entry would significantly underestimate the tail risk of the event. However, we emphasize that the goal of this paper is not to present a novel "state-of-the-art" risk estimation procedure. This is because COVID-19 transmission mechanisms remain poorly characterized, and we acknowledge that our approach requires certain simplifications and assumptions that we discuss at length in the last section of this paper. Rather, faced with the need to provide a risk evaluation tool despite many unknowns, our estimation pipeline combined the best tools at hand to assess the order of magnitude of the risk—thereby opening the avenue for further work on contextualized COVID-19 risk assessment. Consequently, in providing a pipeline for risk estimation, our objective was twofold: (1) developing a publicly available platform to increase risk awareness and promote informed consent for event organizers and participants, while simultaneously (2) encouraging the data collection that is currently so desperately needed for risk assessment. Our model can be applied to any event occurring in the near future and is presented in a user-friendly RShiny interface [38].

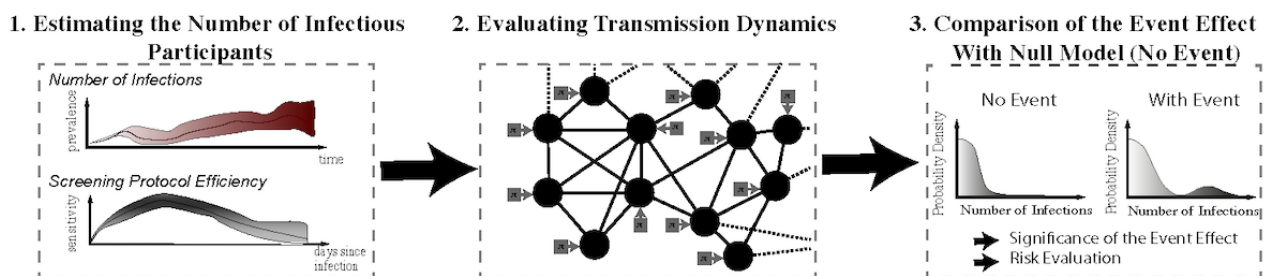**Figure 1.** Summary of our modelling pipeline.

Table 1. Quantiles of the number of transmission events for the Royal Albert Hall concert, by event date, assuming that all participants were wearing masks, so that the exhalation of particles is reduced by 70% and inhalation by 50%.

| Statistics	August 20, 2020		January 20, 2021		March 20, 2021	
	Event	Null	Event	Null	Event	Null
Median	0	0	1	2	0	0
Mean	0.06	0.07	2.38	2.49	0.67	0.63
1st percentile	0	0	0	0	0	0
2.5th percentile	0	0	0	0	0	0
97.5th percentile	1	1	10	7	3	3
99th percentile	1	1	19	8	6	3

Methods

Modelling the Risk of a Large Public Event

Step 1: Estimating the Number of Infectious Participants

Step 1a in our risk modelling procedure was determining the projected incidence, by predicting the number of infectious cases attending a given future event. COVID-19 forecasting is undeniably an involved task, as reflected by its impressive corresponding body of literature (eg, agent-based models or susceptible-exposed-infectious-removed models [39-49]). Predicting the number of new cases per day typically depends on the choice of a specific parameterization (eg, an exponential growth for computing the reproductive number R [50,51]), whose validity is severely hindered by continuous updates to public policies. To alleviate these concerns, we used a nonparametric k-nearest neighbor (kNN) approach. Using all trajectories of the disease incidence across countries and time since the beginning of the pandemic, we computed the k=100 closest trajectories (in terms of the l_2 loss) on time windows of 2 weeks. The historical trajectories of these kNNs were then used as a "dictionary of observed behaviors" to predict the daily incidence rate in the days leading to the event. We defer to Multimedia Appendix 2 for a more in-depth discussion of this

estimation procedure, a description of the parameter selection process, and an evaluation of its performance compared with standard epidemic prediction methods. To briefly summarize, our kNN approach provides a nonparametric, model-agnostic approach to epidemic prediction that is more robust for nonstationarity in public policies than model-based approaches. We show in Multimedia Appendix 2 that these parameters (k=100 neighbors, fitted on trajectories of 14 days) are optimal in allowing an accurate estimation of the trajectory while providing adequate coverage and uncertainty quantification. In fact, we show that, while standard methods fail to provide reliable uncertainty estimates, our kNN methods provide a coverage greater than 95%. Despite coming at the price of wider prediction intervals, our pipeline privileges methods that allow us to correctly estimate the uncertainty in its outputs—thereby more accurately reflecting the state of our knowledge (or lack thereof). Figure 2 presents a comparison of the projected incidences for our 3 dates of interest (August 20, 2020; January 20, 2021; March 20, 2021) for the RAH concert using 2 weeks of fitting and predicting 4 weeks in advance. Note the good coverage provided by our method (the convex hull of the 95% prediction intervals for the projected incidences contains the actual observations). These plots also highlight the importance and variability of the incidence, which varied by orders of magnitude between August 2020 and January 2021.

Figure 2. Projected incidence (average and 95% prediction interval) using a 100-nearest neighbor approach, which provides good coverage (observed trajectory lies within the 95% prediction interval). The black line denotes observed incidence rates, while the red denotes the predicted rates, based on an initial period of observation of 14 days; the prediction interval for the predicted incidence over the next 4 weeks is highlighted in dark grey.
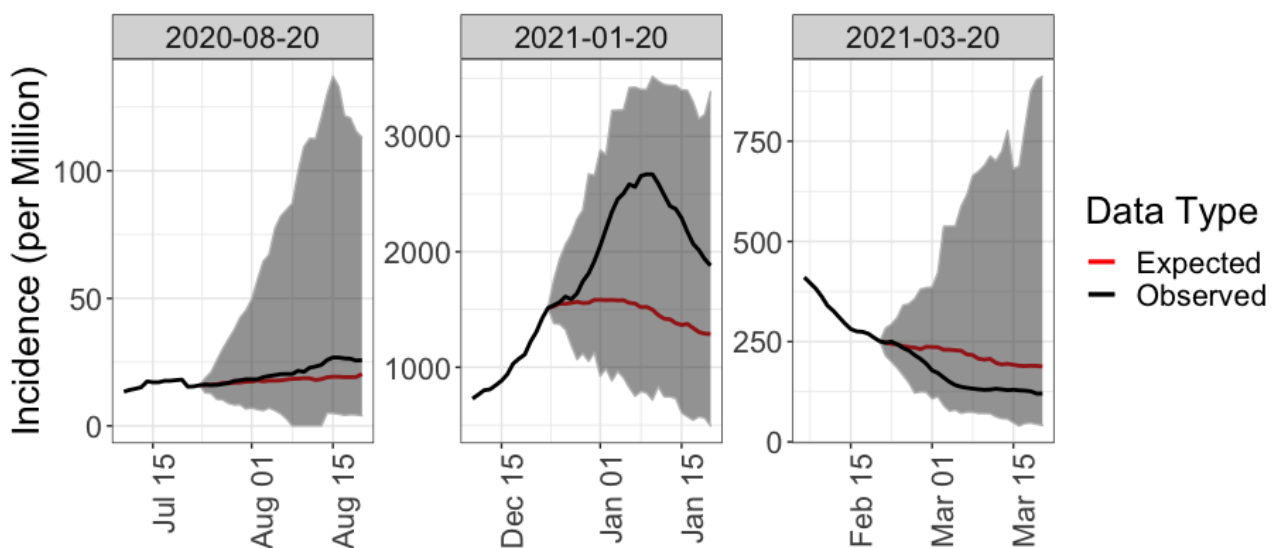
Step 1b was determining the under-ascertainment bias. The estimated number of new cases based on official incidence data will then need to be corrected for under-ascertainment. The latter refers to the downward bias of the reported prevalence in the population, due for instance, to limited testing capacity, low test sensitivity, or people being unwilling or unable to take a test. To this end, we compared the ratio of the number of deaths over reported cases (translated by 3 weeks) to an expected, age-stratified infection-fatality ratio [52] (see Multimedia Appendix 2 for more details). To highlight the potential importance of this correction step, the ascertainment rate for

the United Kingdom was evaluated as over 90% for August 2020 but below 40% for December 2020.

Step 1c was determining the number of infectious participants at the event. Having predicted the background daily incidence rate, we turned to the estimation of the number of infectious participants who will attend the event despite the screening protocols. For an infectious individual to attend the event in spite of the CAPACITY study's screening protocol, they must (1) have no COVID-19–like symptoms or fail to report them on the morning of the event, (2) receive a (false) negative result during antigen testing D at 2 days prior to the event, and (3) be

contagious (rather than simply infected) at the time of the event. We evaluate the joint probability of these events as follows and, for the sake of clarity, refer the reader to Multimedia Appendix 2 for an in-depth explanation of our estimation procedure.

Regarding symptom-check failure, one of the main challenges associated with the COVID- 19 crisis is the number of asymptomatic cases—that is, infected individuals who do not express symptoms and are thus unaware of their potential infectiousness. This group includes individuals that are either presymptomatic or completely asymptomatic during the course of their illness—the latter are estimated to represent roughly 25% of all cases [53]. For symptomatic patients, the probability of having symptoms on the day of the event is also a function of time since infection. To account for this temporal dependency, we used estimates of the incubation period (defined as the number of days between infection and symptom onset) from McAloon et al [54] and data on symptom duration from van Kampen et al [55] to estimate the probability for a ticket holder infected k days before the event to exhibit symptoms on the day of the event. A density plot of this probability is displayed in red in Figure 3A.

Figure 3. (A) Density of the COVID-19 incubation time and percentage culture positive and (B) probability that an individual is infectious (light grey), that the screening protocol will miss them (black), and that they will be missed and so attend the event (red) as a function of days since infection. The shaded regions denote the uncertainty of this estimate due to the uncertainty on the sensitivity of the test.
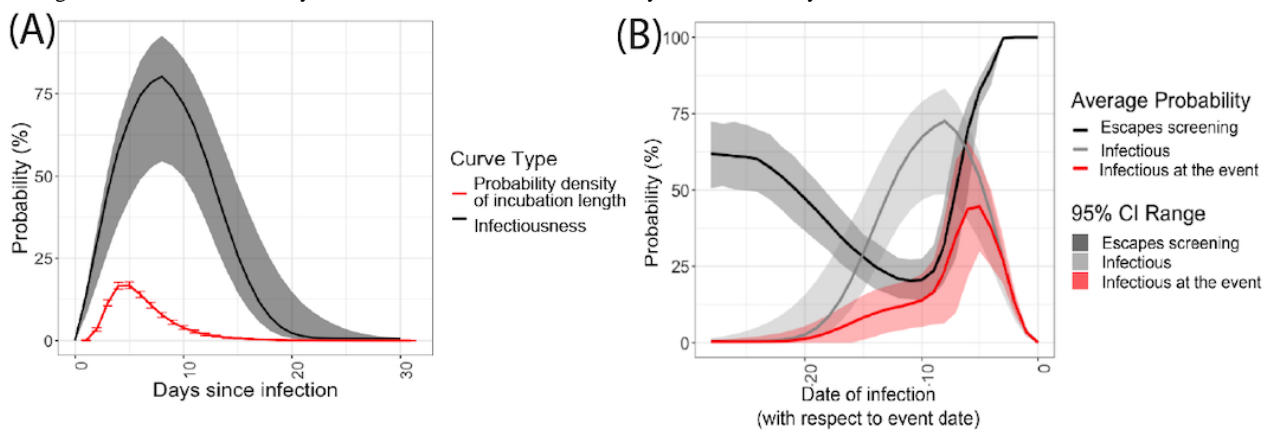
Regarding antigen test failure, the sensitivity of COVID-19 tests depends heavily on the time since infection—whether these are the gold-standard polymerase chain reaction (PCR) or lateral flow antigen assays [56]. Moreover, studies have shown that lateral flow antigen tests have much lower sensitivity on asymptomatic individuals than symptomatic: In particular, according to a recent Centers for Disease Control and Prevention report [57], rapid antigen testing has 80% sensitivity on symptomatic individuals, but only 40% sensitivity on asymptomatic individuals. Coupling the sensitivity estimates [56,57] with the distribution of the incubation period and estimated percentage of asymptomatic cases [53,54], for each individual infected at day k taking an antigen test D days before the event, the probability of getting through the filtering protocol is thus given by the formula:

$$p = P[Contaminated\ at\ time\ k]$$
$$= \left(1 - s_{t-k-D}^{(symptomatic)}\right)\left(p_{SC} \times p_{t-k}^{(symptom)} + \left(1 - p_{t-k}^{(symptom)}\right)\right)$$
$$+ p^{(asymptomatic)}\left(1 - s^{(asymptomatic)}\right)$$

where $s_{t-k-D}^{(symptomatic)}$ and $s^{(symptomatic)}$ are the sensitivities of the test taken D days before the event for a symptomatic participant infected $t-k$ days before the event and an asymptomatic individual, respectively. The parameter $p_{t-k}^{(symptom)}$ denotes the probability for a symptomatic individual to exhibit symptoms $t-k$ days after infection, whereas $p^{(symptom)}$ is the probability of being asymptomatic. Finally, the variable p_{sc} denotes the probability of the symptom check failing—namely, that the participant does not want to report their symptoms (see

Multimedia Appendix 2 for more details). The curve in black on Figure 3B shows the probability of the failure of the screening protocol as a function of days after infection. The shaded areas denote the uncertainty around this estimate due to the variability of the incubation time.

The infectiousness of the participants—that is, the propensity of an infected ticket holder to contaminate others—is a function of time since infection. In order to estimate this relationship, we build upon the existing literature studying the link between reverse-transcription PCR thresholds and cultivable virus [58,59]. The percentage of culturable viral material in the sample can indeed be used as a proxy for infectiousness. Using the estimated percentages of viable samples [58,59] as a function of time since symptom onset, compounded with distribution of the incubation period duration [54], we computed an estimate of the infectiousness as a function of time since infection (black curve in Figure 3A). A more complete description of this estimation procedure is presented in Multimedia Appendix 2. The results are presented in Figure 3B. The red line in Figure 3B shows the resulting probability for an infectious ticket holder to pass through the screening protocol and be allowed into the event. Note that ticket holders that have been infected 5 days before the event are the most likely to be infectious and let in the venue on the day of the event.

Step 1d was determining the number of participants at risk. Finally, the last quantity that we needed to infer before getting into the specifics of the transmission mechanisms was the number of participants at risk of being infected who present at the event. This requires a knowledge of the participants' COVID-19 susceptibility status (ie, has the participant already

had COVID-19 in the previous year, or has the participant been vaccinated?) While previous history could be imputed through additional questions (eg, previous positive test for COVID-19 and symptoms combined in a model such as in [60]), for the sake of simplicity, we only considered the vaccination status of the participants—thus leaving out the proportion of the population that had COVID-19 but was not yet vaccinated. This induces a risk estimate that is biased upward and is thus more conservative. We imputed missing data (cases where the participants have not filled in their vaccination status) using linear regression, expressing vaccination rate as a function of time. This assumes that vaccinations are operating at capacity (see Multimedia Appendix 2 for a longer discussion on the reasons for this approximation and further ways of improving this model). Having imputed the rate of new vaccinations $\pi_s, s=1\ldots t$ days leading to the event, we turned to the estimation of the number of individuals that are likely to be susceptible. Recent reports indicate that vaccine-acquired immunity is a function of both time since vaccination and number of doses [61]. To compute the effective number of participants at risk in the event, we used a compound Poisson distribution: On each day s in the weeks leading to the event, the number X of new participants vaccinated (having either their first or second dose) is expressed as a Poisson($\pi^{(dose\ j)}$), where $j \in \{1,2\}$. Each of

these newly vaccinated individuals then has a probability $\rho^{(dose\ j)}$ of being immune, depending on the date and dose j that they have received. The resulting number of immune people Z attending the event can thus be modelled as:

$$Z \sim Poisson\left(\sum_{j=1}^{2}\sum_{s=1}^{T}\pi_s^{(dose\ j)}\rho_{t-s}^{(dose\ j)}\right)$$

We discuss in Multimedia Appendix 2 how this estimation can easily be modified as the vaccination rates increase and the Poisson approximation becomes no longer valid.

Royal Albert Hall Example

For the RAH example, we present a comparison of each quantity for 3 different dates (see Table 2). Of note is that the screening safety protocol is effective in more than 60% of cases, that when combined with the expected infectiousness of participants and self-reporting of COVID-19–like symptoms, implies that 95% of infected cases are removed. We also note that prevalence is very important in determining the number of infectious cases at the event—thereby highlighting the importance of a context-aware risk calculator. The combined effect of the screening protocol and the natural time-dependent infectiousness of infected ticket holders means that the number of infectious participants at the event is likely to be very low (~ of the order of tens in times of extremely high prevalence).

Table 2. Comparison of the efficiency of the screening protocol and the number of infectious participants at the event by date.

Measurement	August 20, 2020	January 20, 2021	March 20, 2021[a]
Projected incidence (in 1,000,000)	20	1286	188
Number of infected participants	3.6	299.3	50.2
Number of infectious participants at the event	0.22	7.96	2.00
Percentage of caught cases, %	94	97	96
Number of susceptible participants	4996.4	4700.7	3860.4

[a]Vaccination rates started to account for a substantial proportion of the British public, so that the sum of the number of susceptible participants and the number of infected participants does not equate 5000.

Step 2: Modelling Transmission Dynamics

Having estimated the number of infectious participants at the event, the second major component of our model consists of estimating the number of transmission events during the event itself.

Identification of Transmission Mechanisms

More than a year after the start of the epidemic, the precise mechanisms by which COVID-19 is transmitted are still unclear. Aside from direct physical contact, experts continue to debate the significance of the following 2 main routes of infection: droplet transmission and airborne transmission.

In the scenario of droplet transmission, transmission happens through the inhalation of droplets (particles of 5-10 μm in diameter [62]) and typically occurs when a person is in close proximity (within 1 meter) of someone who has respiratory symptoms (eg, coughing or sneezing).

Increasing concerns around airborne transmission have been raised by a number of experts over the past few months [63,64].

Airborne transmission refers to the presence of the virus within droplet nuclei remaining in the air for long periods of time and with the potential to travel long distances [63] and penetrate more deeply in respiratory tracts. Airborne transmission has been estimated to be nearly 19 times more likely indoors than outdoors [65]. In the context of large public events, this transmission route thus has more diffusive power and hence could explain several super-spreader events (SSEs) [6], making it a major cause for concern [2,63,66-72].

While droplet emission is undeniably a source of concern and a major source of transmission, simple safety precautions such as mask wearing have been shown to efficiently control this transmission source [73,74]: It is estimated that face masks can block 80% of exhaled droplets and reduce inhaled droplets by up to 50% and so, on average, reduce the transmission probability by 70% [73]. Conversely, the evidence concerning the efficiency of standard protective equipment in filtering aerosol droplets varies widely across studies probably due to "variation in experimental design and particle sizes analyzed" [73]. Airborne transmission in indoor settings can thus represent

one of the main risk factors in live events, which we focus on modelling using the aerosol model proposed by Jimenez and collaborators [21,69,75]. This aerosol transmission model is currently one of the only COVID-19 transmission models that provide enough granularity to quantify the risk associated with an event. This recognized model has been used several times in the literature over the course of the pandemic [76], including to allow in-class teaching at the University of Illinois at Chicago [70]. Based on the Wells-Riley model [77-79], this estimator calibrates the quanta to known transmission events and considers important factors to compute a risk estimate, including event-specific (eg, number of people, local prevalence) and venue-specific (ventilation rate, size of the venue, UV exposure) variables. This Wells-Riley–based model relies on the evaluation of 3 quantities: (1) the quanta exhalation rate, which is contingent on the activity performed and the number of infectious participants; (2) quanta concentration, which is a function of the volume of the space, the room ventilation rate, and the quanta exhalation rate; and (3) quanta inhalation rate, which is a function of the quanta concentration and breathing rate associated with the activity performed. The probability for each susceptible individual to be infected can then be written as $p_{infection} = 1 - e^{-qinhalation}$. See Multimedia Appendix 3 for more details.

Modelling the Uncertainty of the Model

To estimate the uncertainty associated with this model, we used Monte-Carlo simulations. We simulated random input parameters (number of infectious and susceptible individuals) using the distributions and uncertainty estimates discussed in the previous section. In order to model the uncertainty associated with the aerosol transmission model, we added a sampling step at the end of the Jimenez and Peng pipeline. This allowed us to account for individual variations in infectious participants' ability to spread the disease and to remain consistent with the extensive literature on the heavy-tailed Pareto nature of COVID-19 transmission and superspreading [24-27]. For each infected participant, we sampled the number of quanta that they exhale using a Pareto distribution with shape $\theta = 1.16$ and rate $\eta = \theta/(\theta - 1)q^{exhalation}$. This produces a distribution centered around $q^{exhalation}$ but skewed to the right and heavy-tailed—thereby modelling the heterogeneity in infected participants' ability to spread. This choice of parameters allowed us to abide by the Pareto principle, according to which 80% of transmissions are due to 20% of those infected. In accordance with the uniform mixing assumption of the aerosol transmission models, susceptible participants then all inhale a quanta concentration that is a function of the sum of the exhaled quanta: All have an identical probability of becoming infected. In mathematical terms, infections are thus simulated using a binomial distribution such that $n_{infected} \sim \text{Binomial}(n_{susceptible}, 1 - e^{-qinhaled})$. We discuss the limitations of this approach and its assumptions in the discussion section of this paper.

The code for the model can be found online on the authors' Github [80].

Results

Step 3: Comparison With the Null Model

To quantify the effect of the event, it is necessary to put it in context of the background rate of infections: Even if the participants had not been to the event, they could have been infected elsewhere. In this null model, the number of infections is binomially distributed, such that the number infections Y is $Y \sim \text{Binom}(n_{susceptible}, \pi)$.

We present the results for the RAH example in Table 2. This table shows in grey the values of the different quantiles of this distribution. We note the skewed distribution that we obtain is expected given the modelling of the uncertainty around inhalation rate. If the event did not occur, then on each respective date, there would be an expected community transmission of 0.07 (95% prediction interval: 0-1), 2.5 (95% prediction interval: 0-7), and 0.63 (95% prediction interval: 0-3) events on August 20, 2020, January 20, 2021, and March 20, 2021, respectively. However, with the event taking place on these dates and calculating the expected number of infectious individuals, susceptible individuals, and transmission dynamics within the venue, the distribution of the number of transmission events would in general widen to 0.06 (0-1), 2.38 (0-19), and 0.67 (0-6) in that same order. In this case, it is important to note the similarity in mean transmission between the "event" and "no event" scenarios and their substantial deviation in the tails. This highlights the importance of modelling the distribution of the risk and highlighting its substantial heavy tails, rather than providing point estimates.

It is likely, although not inevitable, that the event will have an impact on the transmission and increase it irrespective of the level of the prevalence. However, for low levels of prevalence and higher vaccination rates, this substantially decreases. Having computed the number of expected transmission events, we can then compute several complementary metrics of interest including, for example, the secondary attack rate (SAR)—that is, the number of COVID-19 cases in the participants' community in both the null and event models. SAR can be calculated from the predicted reproductive rate (R) in the regions where the ticket holders dwell. In the United Kingdom, R rates are updated on a weekly basis at regional levels (eg, East Midlands, London) and available from the Office for National Statistics or can be derived from the kNN modelling previously described. An opportunity for further research would be to estimate SAR within households by gathering contextual data from ticket holders. Equally, estimates of hospitalizations and deaths might be possible based on individual characteristics and comorbidities; however, this is beyond the scope of the current article.

Evaluating the Effectiveness of the Screening Protocol

This risk modelling pipeline also allows comparison of different protocols and situations. For example, this pipeline highlights (1) the importance of event duration (the longer the dwell time at the event, the more at risk the participants) and (2) the importance of wearing masks. Table 3 quantifies the outcomes of holding the event on our 3 dates, assuming that either 0%,

50%, or 100% of participants are wearing masks or varying parameters such as the density or length of the concert. Figure 4 completes that analysis by providing a visual representation of the effect of these parameters on the distribution of the number of infections. The distributional nature of these results

is essential in highlighting nuances between scenarios: While holding an event at half capacity or for half the duration produces average transmission risks that are roughly similar, holding the event at half capacity seems to more substantially reduce the effect of the event in the tails of the distribution.

Table 3. Effect of different input parameters on the quantiles of the number of infections for an event at the Royal Albert Hall across all 3 dates.

Event	August 20, 2020, median, mean (99% CI)	January 20, 2021, median, mean (99% CI)	March 20, 2021, median, mean (99% CI)
No mask wearing, 3 hours, n=5000	0, 0.3 (0-4)	5, 9.9 (0-76)	1, 2.4 (0-21)
50% mask wearing, 3 hours, n=5000	0, 0.2 (0-3)	3, 5.5 (0-40)	1, 1.3 (0-13)
100% mask wearing, 3 hours, n=5000	0, 0.1 (0-1)	1, 2.4 (0-19)	0, 0.7 (0-6)
100% mask wearing, 1.5 hours, n=5000	0, 0.04 (0-1)	0, 1.4 (0-10)	0, 0.4 (0-3)
100% mask wearing, 3 hours, n=2500	0, 0.2 (0-1)	0, 0.9 (0-8)	0, 0.2 (0-3)

Figure 4. Boxplots showing the distribution of the number of infections across different scenarios, for our Royal Albert Hall event held on March 20, 2021: Where variables are not mentioned, the number of attendees is 5000, the duration is 3 hours, and the proportion of attendees wearing masks is 100%.
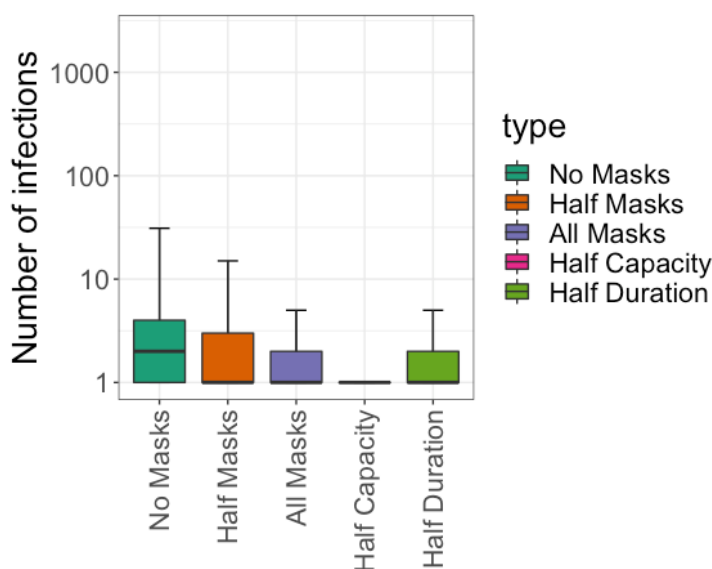
In addition to the aggregated risk that a live event presents, individual risk of transmission can be estimated and can be communicated to ticket holders so that they can gauge whether the risk of attending the event outweighs their desire to attend. For the first person to purchase a ticket, risk of transmission will be calculated based on their own immunity status (eg, vaccination, regional prevalence) and a synthetic population based on national prevalence at that time. As more bookings are assigned to ticket holders, the reliance on the synthetic population decreases as understanding of the number of susceptible and potentially infectious individuals attending the event increases. Therefore, the confidence in the risk score increases as the event draws closer and as the proportion of tickets sold increases. This can be reflected in the updated risk scores provided to ticket holders as the event approaches. The individual risk scores can be modified based on alternative scenarios imputed into the risk algorithm. For example, for an individual not yet vaccinated, their risk could be also presented as if they had been vaccinated, offering an opportunity for the individual to appreciate how vaccination could have modified their risk. Such an approach could form the basis for behavior

change interventional studies for promoting health literacy and tackling vaccine hesitancy (see Multimedia Appendix 1). By working in partnership with the live events organizer, individuals that chose to opt out can be reimbursed without delay and the ticket re-sold.

Discussion

The modelling we propose is based on prevalence estimates and screening protocols to calculate the number of infectious and susceptible individuals attending the event as well as transmission dynamics at the venue to predict the number of new infections. Our paper demonstrates the value of estimating attack rates from live events so that they can be appropriately managed. We also demonstrate how individual ticket holders can receive personalized risk scores for contracting COVID-19 at the event, which would, for the first time, enable genuine informed consent to be obtained. Although this methodology provides clear benefit to event organizers, local public health authorities, and individual ticket holders, our approach is based

on several assumptions that group in 2 categories: modelling assumptions and parameter sensitivity.

Modelling Assumptions

As they combine data and tools from different sources, the computations in our pipeline rely on assumptions at 3 main levels: predicting COVID-19 prevalence, assessing the efficiency of the screening protocol, and transmission at the event.

Predicting COVID-19 Prevalence

To predict future COVID-19 incidence, we chose a kNN approach as it yields a more robust prediction and better uncertainty quantification than most existing parametric methods. One of the downsides of this approach is that it might not generalize very well to entirely novel behaviors or viral variants—in which case well-parameterized methods may outperform our approach as knowledge of transmission, vaccination, and other relevant model parameters continues to improve. While prevalence predictions are important for event planners and attendees alike, on the day of the event, the more important metric is whether official case rates reflect actual cases (ie, the ascertainment rate). Historically, this rate has been low due to limited testing facilities, and our method to determine ascertainment using cases, deaths, and infection-fatality rates reflects this, but also indicates that ascertainment may exceed 100% in times of widespread testing and low prevalence. It was beyond the scope of this paper to further investigate ascertainment, but we expect that future research will clarify the impact of different test types, their false negative and positive rates, and their frequency of use in determining the ascertainment rate.

Assessing the Efficiency of the Screening Protocol

Our modelling framework assumes that events will screen participants with COVID-19 tests, such as virtually witnessed lateral flow antigen tests. Assessing the efficiency of this screening step requires the estimation of (1) the sensitivity of the test, (2) the probability of having symptoms, and (3) the probability of being infectious—all of these quantities being a function of days since infection. Our estimation of each of these quantities is based on published data—with the exception of the probability of symptom check failure (ie, the probability that a participant lies about their symptoms to get in). By default, we select this probability to be 50%, a choice that will be improved upon as the CAPACITY and other similar studies gather behavioral data. However, as shown in Multimedia Appendix 4, this factor has a relatively minor impact on the outcome of the model compared with the uncertainty of the other inputs. Of potentially greater concern is our assumption that the probability of testing negative 2 days before the event is independent (conditionally on time since infection) of a participant's infectiousness during the event. A potential avenue for improvement could consist of determining both test sensitivity and infectiousness as a function of viral load and estimating the joint probability of the viral load 2 days apart. However, the data required for this approach are—to the best of our knowledge—still lacking and given the variability of the viral load or PCR cycle threshold behavior, this conditional

independence assumption seemed a reasonable first-order approximation.

Transmission at the Event

The airborne transmission model that we use relies on a homogeneous (well-mixed) air hypothesis for an indoor environment. While several other models have been proposed (either breaking the room into compartments or using a distance index) to counter this hypothesis, we highlight (following the discussion by Jimenez and Peng [75]) that this is a first-order approximation: Some participants will have more risk and others less, so that at low quanta concentration, this effect will be averaged out. At very high concentration, the model will likely underestimate the number of infections, but given the efficiency of the screening protocol and density limitations, we do not expect this scenario to be common. Moreover, while this model was originally developed for indoor transmission, its application to an outdoor setting—where the ventilation rate can be considered infinite and transmission is more likely to occur through droplets rather than aerosolized particles—can nonetheless provide a conservative estimate of the risk. We are however currently working on developing a better model for outdoor transmission, relying on a modelling of droplet transmission in crowd bottlenecks. We leave the detail of this separate transmission model to future work. Finally, we note that our model is not tied to any specific transmission mechanism, and as our knowledge of COVID-19 transmission improves, we can refine and supplant the transmission dynamics with a superior alternative or another model that is deemed more suitable.

Parameter Sensitivity

While we try to limit the number of input parameters in our pipeline, the sensitivity of the estimates to these inputs (namely, the mask efficiency and population of interest) has to be studied. We refer the reader to Multimedia Appendix 4 for a quantitative sensitivity analysis and highlight our conclusions here. In terms of the model parameters, the greatest unknown consists in determining the efficiency of masks and protective equipment—the latter having been shown to vary depending on the mask type and activity. However, we hope to make use of the growing body of literature on the topic to update and refine this important factor. Second, our prediction framework assumes that participants at the event have the same probability of infection and vaccination as their regional average. However, this might not be the case as participation in the event may be an incentive to get vaccinated or conversely might select for less cautious subpopulations. The importance of this sampling frame assumption nonetheless decreases as participants' vaccination status and behavioral data from the CAPACITY study will result in more precise estimates.

Model Validation

Finally, one of the main current hurdles for developing risk estimators lies in the absence of quality data to validate and benchmark different transmission models—thereby making the task of validating our transmission pipeline a rather daunting task. Indeed, while we can (and have, see Multimedia Appendix 2) check the accuracy of the vaccination and prevalence

estimation step, the validation of the transmission model itself is inherently difficult: There are no, or very few, available datasets on COVID-19 spread following live events or rigorous accounts of SSEs, nor are there any statistics on how likely SSEs are. As such, the majority of SSEs that are documented currently (1) are generally not detailed enough to untangle the huge variability in context (eg, outdoors vs indoors, activity performed, background prevalence) and (2) suffer from selection bias—and might not be reflective of the general distribution of live events. To make up for the current lack of testing data, we resort here to the following 3 strategies: model checking, model validation on (scarce) existing data, and prospective data gathering.

For model checking, we begin by validating the behavior of our model estimates on documented SSEs [81]—that is, we confirm that the model outputs (1) present similar tail behavior as these documented SSEs and (2) are predicted as outlier SSE events by our model.

For model validation on (scarce) existing data, we also consider 2 documented live indoor concert events [82,83] and use the event parameters as well as the documented transmission statistics to verify that these numbers fall within the realm of feasible outcomes.

For prospective data gathering, finally, to overcome the lack of available data, we propose using the RShiny app [38] as a data collection platform and encourage users (event organizers and participants alike) who use the app to record their event in our dataset by filling in a survey [84]. This paves the way for a larger-scale and more detailed record of transmission events at large gatherings, as well as a more precise modelling of transmission dynamics.

This validation and model assessment step is further described in Multimedia Appendix 5.

Conclusion

A nuanced, data-driven system is required to assess risk at each event informed by the characteristics of all ticket holders and the background risk of transmission concurrent to the event, so that proportionate and specific action can be taken by event organizers and public health authorities. We have detailed our attempt to create such a system and have outlined its predictions and limitations. Our end-to-end risk model is provided in the form of an RShiny interface. At times of high prevalence, this type of system will ensure events likely to increase transmission can be halted. At times of low prevalence, this will ensure events can potentially continue to operate. Learning to live with SARS-CoV-2 will be about implementing systems that support hyperlocal, data-driven decisions so that far-reaching and highly damaging sector-specific lockdowns can be avoided as much as possible.

Conflicts of Interest

JK is the medical director and co-founder of Certific. None of the remaining authors have any competing interests.

Multimedia Appendix 1

Prediction.
[DOCX File , 537 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Model and assumptions of the Jimenez aerosol transmission model.
[DOCX File , 146 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Sensitivity analysis.
[DOCX File , 178 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Risk communication.
[DOCX File , 236 KB-Multimedia Appendix 4]

XSL•FO

RenderX

Multimedia Appendix 5

Model validation.

[DOCX File , 350 KB-Multimedia Appendix 5]

References

1. Lin J, Yan K, Zhang J, Cai T, Zheng J. A super-spreader of COVID-19 in Ningbo city in China. J Infect Public Health 2020 Jul;13(7):935-937 [FREE Full text] [doi: 10.1016/j.jiph.2020.05.023] [Medline: 32554034]

2. Majra D, Benson J, Pitts J, Stebbing J. SARS-CoV-2 (COVID-19) superspreader events. J Infect 2021 Jan;82(1):36-40 [FREE Full text] [doi: 10.1016/j.jinf.2020.11.021] [Medline: 33245943]

3. Nadal M, Lassel L, Denis M, Gibelin A, Fournier S, Menard L, et al. Role of super-spreader phenomenon in a Covid-19 cluster among healthcare workers in a Primary Care Hospital. J Infect 2021 May;82(5):e13-e15 [FREE Full text] [doi: 10.1016/j.jinf.2021.02.009] [Medline: 33581240]

4. Charlotte N. High rate of SARS-CoV-2 transmission due to choir practice in France at the beginning of the COVID-19 pandemic. J Voice 2020 Dec 23:1 [FREE Full text] [doi: 10.1016/j.jvoice.2020.11.029] [Medline: 33386191]

5. Atrubin D, Wiese M, Bohinc B. An outbreak of COVID-19 associated with a recreational hockey game - Florida, June 2020. MMWR Morb Mortal Wkly Rep 2020 Oct 16;69(41):1492-1493 [FREE Full text] [doi: 10.15585/mmwr.mm6941a4] [Medline: 33056952]

6. Hamner L, Dubbel P, Capron I, Ross A, Jordan A, Lee J, et al. High SARS-CoV-2 attack rate following exposure at a choir practice - Skagit County, Washington, March 2020. MMWR Morb Mortal Wkly Rep 2020 May 15;69(19):606-610 [FREE Full text] [doi: 10.15585/mmwr.mm6919e6] [Medline: 32407303]

7. Lewis D. Superspreading drives the COVID pandemic - and could help to tame it. Nature 2021 Feb 23;590(7847):544-546. [doi: 10.1038/d41586-021-00460-x] [Medline: 33623168]

8. Gómez-Carballa A, Bello X, Pardo-Seco J, Martinón-Torres F, Salas A. Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. Genome Res 2020 Oct 02;30(10):1434-1448 [FREE Full text] [doi: 10.1101/gr.266221.120] [Medline: 32878977]

9. Zhang Y, Li Y, Wang L, Li M, Zhou X. Evaluating transmission heterogeneity and super-spreading event of COVID-19 in a metropolis of China. Int J Environ Res Public Health 2020 May 24;17(10):3705 [FREE Full text] [doi: 10.3390/ijerph17103705] [Medline: 32456346]

10. Kochańczyk M, Grabowski F, Lipniacki T. Super-spreading events initiated the exponential growth phase of COVID-19 with higher than initially estimated. R Soc Open Sci 2020 Sep 23;7(9):200786 [FREE Full text] [doi: 10.1098/rsos.200786] [Medline: 33047040]

11. COVID-19 Event Risk Assessment Planning Tool. Georgia Institute of Technology. URL: https://covid19risk. biosci.gatech.edu/ [accessed 2021-10-16]

12. 19 and Me: COVID-19 Risk Score Calculator. Mathematica. URL: https://19andme.covid19.mathematica.org/ [accessed 2021-10-16]

13. What's your risk? ABC7 News. URL: https://tinyurl.com/5yfywcvr [accessed 2021-10-16]

14. MyCOVIDRisk App. Brown University. URL: https://digitalhealth.med.brown.edu/news/2021-05-31/mycovidrisk/ [accessed 2021-10-16]

15. Physician survey of COVID-19 risk for daily activities. Illinois State Medical Society. 2020 Sep 03. URL: https://tinyurl. com/2p95624a [accessed 2021-10-16]

16. Be Informed: Know Your Risk During COVID-19. Texas Medical Association. URL: https://www.texmed.org/uploadedFiles/ Current/2016_Public_Health/Infectious_Diseases/309193%20Risk%20Assessment%20Chart%20V2_FINAL.pdf [accessed 2021-10-16]

17. Know Your Risk This Holiday Season. Texas Medical Association. URL: https://www.texmed.org/uploadedFiles/Current/ 2016_Public_Health/Infectious_Diseases/309640_Winter_Risk_Assessment_Chart_COLOR.pdf [accessed 2021-10-16]

18. Coronavirus (COVID-19) Infection Survey: England. Office for National Statistics. URL: https://www.ons.gov.uk/ peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/coronaviruscovid19infectionsurveydata [accessed 2021-11-10]

19. Lelieveld J, Helleis F, Borrmann S, Cheng Y, Drewnick F, Haug G, et al. Model calculations of aerosol transmission and infection risk of COVID-19 in indoor environments. Int J Environ Res Public Health 2020 Nov 03;17(21):634-643 [FREE Full text] [doi: 10.3390/ijerph17218114] [Medline: 33153155]

20. Mittal R, Meneveau C, Wu W. A mathematical framework for estimating risk of airborne transmission of COVID-19 with application to face mask use and social distancing. Phys Fluids (1994) 2020 Oct 01;32(10):101903 [FREE Full text] [doi: 10.1063/5.0025476] [Medline: 33100806]

21. Peng ZJ, Jimenez J. Exhaled CO as a COVID-19 infection risk proxy for different indoor environments and activities. Environ. Sci. Technol. Lett 2021 Apr 05;8(5):392-397 [FREE Full text] [doi: 10.1021/acs.estlett.1c00183]

22. Wilson N, Corbett S, Tovey E. Airborne transmission of covid-19. BMJ 2020 Aug 20;370:m3206 [FREE Full text] [doi: 10.1136/bmj.m3206] [Medline: 32819961]

XSL•FO

RenderX

23. Risk tracker. microCOVID Project. URL: https://www.microcovid.org/tracker [accessed 2021-10-16]

24. Beare B, Toda A. On the emergence of a power law in the distribution of COVID-19 cases. Physica D 2020 Nov;412:132649 [FREE Full text] [doi: 10.1016/j.physd.2020.132649] [Medline: 32834250]

25. Althouse BM, Wenger EA, Miller JC, Scarpino SV, Allard A, Hébert-Dufresne L, et al. Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2. Cornell University. 2020 May 27. URL: https://arxiv.org/abs/2005.13689 [accessed 2021-10-16]

26. Cirillo P, Taleb NN. Tail risk of contagious diseases. Nat. Phys 2020 May 25;16(6):606-613. [doi: 10.1038/s41567-020-0921-x]

27. Donnat C, Holmes S. Modeling the heterogeneity in COVID-19's reproductive number and its impact on predictive scenarios. Journal of Applied Statistics 2021 Jun 22:1-29. [doi: 10.1080/02664763.2021.1941806]

28. Grenier E. Could Germany's COVID concert experiment help arenas hold large events again? Deutsche Welle. 2020 Aug 23. URL: https://www.dw.com/en/could-germanys-covid-concert-experiment-help-arenas-hold-large-events-again/a-54661902/ [accessed 2021-10-16]

29. Schulten L. Dutch researchers test ways to party during the pandemic. Deutsche Welle. 2021 Mar 22. URL: https://www.dw.com/en/dutch-researchers-test-ways-to-party-during-the-pandemic/a-56953021/ [accessed 2021-10-16]

30. Covid: Barcelona hosts large gig after testing crowd. BBC News. 2021 Mar 28. URL: https://www.bbc.com/news/world-europe-56556451/ [accessed 2021-10-16]

31. Moritz S, Gottschick C, Horn J, Popp M, Langer S, Klee B, et al. The risk of indoor sports and culture events for the transmission of COVID-19 (restart-19). medRxiv. 2020. URL: https://doi.org/10.1101/2020.10.28.20221580 [accessed 2021-10-16]

32. The Possibility of COVID-19 after Vaccination: Breakthrough Infections. Centers for Disease Control and Prevention. 2021 Sep 07. URL: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/effectiveness/why-measure-effectiveness/breakthrough-cases.html/ [accessed 2021-10-16]

33. Tinker B, Fox M. So far, 5,800 fully vaccinated people have caught Covid anyway in US, CDC says. CNN Health. 2021 Apr 15. URL: https://edition.cnn.com/2021/04/14/health/breakthrough-infections-covid-vaccines-cdc/index.html [accessed 2021-10-16]

34. Deeks J, Raffle A, Gill M. UK government must urgently rethink lateral flow test roll out, warn experts. BMJ. 2021 Nov 1. URL: https://tinyurl.com/f8akyswb [accessed 2021-10-16]

35. Halliday J. Rapid Covid testing in England may be scaled back over false positives. The Guardian. 2021 Apr 15. URL: http://www.theguardian.com/world/2021/apr/15/rapid-covid-testing-in-england- [accessed 2021-10-16]

36. Harris M, Kreindler J, El-Osta A, Esko T, Majeed A. Safe management of full-capacity live/mass events in COVID-19 will require mathematical, epidemiological and economic modelling. J R Soc Med 2021 Jun 19;114(6):290-294 [FREE Full text] [doi: 10.1177/01410768211007759] [Medline: 33870766]

37. Beranek L. Concert Halls and Opera Houses: Music, Acoustics, and Architecture. New York, NY: Springer; 2012.

38. What are the parameters for the event? Shiny Apps. URL: https://homecovidtests.shinyapps.io/aerosol_transmission_model/ [accessed 2021-10-16]

39. Chatterjee K, Chatterjee K, Kumar A, Shankar S. Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model. Med J Armed Forces India 2020 Apr;76(2):147-155 [FREE Full text] [doi: 10.1016/j.mjafi.2020.03.022] [Medline: 32292232]

40. Grant A. Dynamics of COVID-19 epidemics: SEIR models underestimate peak infection rates and overestimate epidemic duration. medRxiv. 2020. URL: https://www.medrxiv.org/content/10.1101/2020.04.02.20050674v1 [accessed 2021-10-16]

41. He S, Peng Y, Sun K. SEIR modeling of the COVID-19 and its dynamics. Nonlinear Dyn 2020 Jun 18;101(3):1-14 [FREE Full text] [doi: 10.1007/s11071-020-05743-y] [Medline: 32836803]

42. Gupta R, Pandey G, Chaudhary P, Pal S. SEIR and Regression Model based COVID-19 outbreak predictions in India. medRxiv. 2020 Apr 03. URL: https://www.medrxiv.org/content/10.1101/2020.04.01.20049825v1 [accessed 2021-10-16]

43. Wu J, Leung K, Leung G. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet 2020 Feb;395(10225):689-697. [doi: 10.1016/s0140-6736(20)30260-9]

44. Zhao S, Chen H. Modeling the epidemic dynamics and control of COVID-19 outbreak in China. Quant Biol 2020 Mar 11;8(1):1-9 [FREE Full text] [doi: 10.1007/s40484-020-0199-0] [Medline: 32219006]

45. Akbarpour M, Cook C, Marzuoli A, Mongey S, Nagaraj A, Saccarola M, et al. Socioeconomic network heterogeneity and pandemic policy response. NBER Working Paper. 2020 Jun. URL: https://www.nber.org/papers/w27374 [accessed 2021-10-16]

46. Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia. Nat Commun 2020 Nov 11;11(1):5710 [FREE Full text] [doi: 10.1038/s41467-020-19393-6] [Medline: 33177507]

47. Kai D, Goldstein G, Morgunov A, Nangalia V, Rotkirch A. Universal Masking is Urgent in the COVID-19 Pandemic: SEIR and Agent Based Models, Empirical Validation, Policy Recommendations. Cornell University. 2020 Apr 22. URL: https://arxiv.org/abs/2004.13553 [accessed 2021-10-16]

48.   Rockett RJ, Arnott A, Lam C, Sadsad R, Timms V, Gray K, et al. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. Nat Med 2020 Sep 9;26(9):1398-1404. [doi: 10.1038/s41591-020-1000-7] [Medline: 32647358]

49.   Silva PC, Batista PV, Lima HS, Alves MA, Guimarães FG, Silva RC. COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions. Chaos Solitons Fractals 2020 Oct;139:110088 [FREE Full text] [doi: 10.1016/j.chaos.2020.110088] [Medline: 32834624]

50.   Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc Biol Sci 2007 Feb 22;274(1609):599-604 [FREE Full text] [doi: 10.1098/rspb.2006.3754] [Medline: 17476782]

51.   Dietz K. The estimation of the basic reproduction number for infectious diseases. Stat Methods Med Res 1993 Jul 02;2(1):23-41. [doi: 10.1177/096228029300200103] [Medline: 8261248]

52.   Bevand M. mbevand / covid19-age-stratified-ifr. GitHub. URL: https://github.com/mbevand/covid19-age-stratified-ifr [accessed 2021-11-10]

53.   He J, Guo Y, Mao R, Zhang J. Proportion of asymptomatic coronavirus disease 2019: A systematic review and meta-analysis. J Med Virol 2021 Feb 13;93(2):820-830 [FREE Full text] [doi: 10.1002/jmv.26326] [Medline: 32691881]

54.   McAloon C, Collins A, Hunt K, Barber A, Byrne AW, Butler F, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. BMJ Open 2020 Aug 16;10(8):e039652 [FREE Full text] [doi: 10.1136/bmjopen-2020-039652] [Medline: 32801208]

55.   van Kampen JJA, van de Vijver DAMC, Fraaij P, Haagmans BL, Lamers MM, Okba N, et al. Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (COVID-19). Nat Commun 2021 Jan 11;12(1):267 [FREE Full text] [doi: 10.1038/s41467-020-20568-4] [Medline: 33431879]

56.   Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in false-negative rate of reverse transcriptase polymerase chain reaction–based SARS-CoV-2 tests by time since exposure. Annals of Internal Medicine 2020 Aug 18;173(4):262-267. [doi: 10.7326/m20-1495]

57.   Pray IW, Ford L, Cole D, Lee C, Bigouette JP, Abedi GR, CDC COVID-19 Surge Laboratory Group. Performance of an antigen-based test for asymptomatic and symptomatic SARS-CoV-2 testing at two university campuses - Wisconsin, September-October 2020. MMWR Morb Mortal Wkly Rep 2021 Jan 01;69(5152):1642-1647 [FREE Full text] [doi: 10.15585/mmwr.mm695152a3] [Medline: 33382679]

58.   Singanayagam A, Patel M, Charlett A, Lopez Bernal J, Saliba V, Ellis J, et al. Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. Euro Surveill 2020 Aug;25(32):2001483 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.32.2001483] [Medline: 32794447]

59.   He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med 2020 Apr 15;26(5):672-675. [doi: 10.1038/s41591-020-0869-5]

60.   Donnat C, Miolane N, Bunbury F, Kreindler J. A Bayesian hierarchical network for combining heterogeneous data sources in medical diagnoses. Proceedings of the Machine Learning for Health NeurIPS Workshop 2020;136:53-84.

61.   Impact of COVID-19 vaccines on mortality in England: December 2020 to March 2021. Public Health England. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/977249/PHE_COVID-19_vaccine_impact_on_mortality_March.pdf/ [accessed 2021-10-16]

62.   Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations. World Health Organization. 2020 Mar 29. URL: https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations/ [accessed 2021-10-16]

63.   Morawska L, Milton D. It is time to address airborne transmission of coronavirus disease 2019 (COVID-19). Clin Infect Dis 2020 Dec 03;71(9):2311-2313 [FREE Full text] [doi: 10.1093/cid/ciaa939] [Medline: 32628269]

64.   Lewis D. Is the coronavirus airborne? Experts can't agree. Nature 2020 Apr 02;580(7802):175-175. [doi: 10.1038/d41586-020-00974-w] [Medline: 32242113]

65.   Nishiura H, Oshitani H, Kobayashi T, Saito T, Sunagawa T, Matsui T, MHLW COVID-19 Response Team, et al. Closed environments facilitate secondary transmission of coronavirus disease 2019 (COVID-19). medRxiv. 2020 Apr 16. URL: https://www.medrxiv.org/content/10.1101/2020.02.28.20029272v2 [accessed 2021-10-16]

66.   Asadi S, Bouvier N, Wexler AS, Ristenpart WD. The coronavirus pandemic and aerosols: Does COVID-19 transmit via expiratory particles? Aerosol Sci Technol 2020:1-4 [FREE Full text] [doi: 10.1080/02786826.2020.1749229] [Medline: 32308568]

67.   Zhang R, Li Y, Zhang AL, Wang Y, Molina MJ. Identifying airborne transmission as the dominant route for the spread of COVID-19. Proc Natl Acad Sci U S A 2020 Jun 30;117(26):14857-14863 [FREE Full text] [doi: 10.1073/pnas.2009637117] [Medline: 32527856]

68.   Bhagat RK, Davies Wykes MS, Dalziel SB, Linden PF. Effects of ventilation on the indoor spread of COVID-19. J Fluid Mech 2020 Sep 28;903:F1 [FREE Full text] [doi: 10.1017/jfm.2020.720] [Medline: 34191877]

69.   Miller SL, Nazaroff WW, Jimenez JL, Boerstra A, Buonanno G, Dancer SJ, et al. Transmission of SARS-CoV-2 by inhalation of respiratory aerosol in the Skagit Valley Chorale superspreading event. Indoor Air 2021 Mar 13;31(2):314-323 [FREE Full text] [doi: 10.1111/ina.12751] [Medline: 32979298]

70. Elbanna A, Wong G, Weiner Z, Wang T, Zhang H, Liu Z, et al. Entry screening and multi-layer mitigation of COVID-19 cases for a safe university reopening. medRxiv. URL: https://www.medrxiv.org/content/10.1101/2020.08.29.20184473v1 [accessed 2021-10-16]

71. Buonanno G, Stabile L, Morawska L. Estimation of airborne viral emission: Quanta emission rate of SARS-CoV-2 for infection risk assessment. Environ Int 2020 Aug;141:105794 [FREE Full text] [doi: 10.1016/j.envint.2020.105794] [Medline: 32416374]

72. Fennelly K. Particle sizes of infectious aerosols: implications for infection control. The Lancet Respiratory Medicine 2020 Sep;8(9):914-924 [FREE Full text] [doi: 10.1016/s2213-2600(20)30323-4]

73. Brooks JT, Butler JC. Effectiveness of mask wearing to control community spread of SARS-CoV-2. JAMA 2021 Mar 09;325(10):998-999. [doi: 10.1001/jama.2021.1505] [Medline: 33566056]

74. Peeples L. Face masks: what the data say. Nature 2020 Oct 06;586(7828):186-189. [doi: 10.1038/d41586-020-02801-8] [Medline: 33024333]

75. Jimenez JL, Peng Z. COVID-19 Aerosol Transmission Estimator. URL: https://docs.google.com/spreadsheets/d/16K1OQkLD4BjgBdO8ePj6ytf-RpPMlJ6aXFg3PrIQBbQ/edit#gid=519189277 [accessed 2021-10-16]

76. Harrichandra A, Ierardi AM, Pavilonis B. An estimation of airborne SARS-CoV-2 infection transmission risk in New York City nail salons. Toxicol Ind Health 2020 Sep;36(9):634-643 [FREE Full text] [doi: 10.1177/0748233720964650] [Medline: 33085569]

77. Riley E, Murphy G, Riley R. Airborne spread of measles in a suburban elementary school. Am J Epidemiol 1978 May;107(5):421-432. [doi: 10.1093/oxfordjournals.aje.a112560] [Medline: 665658]

78. Wells W. Airborne contagion and air hygiene: An ecological study of droplet infections. JAMA 1955 Sep 03;159(1):90. [doi: 10.1001/jama.1955.02960180092033]

79. Sze To GN, Chao C. Review and comparison between the Wells-Riley and dose-response approaches to risk assessment of infectious respiratory diseases. Indoor Air 2010 Feb;20(1):2-16 [FREE Full text] [doi: 10.1111/j.1600-0668.2009.00621.x] [Medline: 19874402]

80. donnate / aerosol_transmission_model. GitHub. URL: https://github.com/donnate/aerosol_transmission_model [accessed 2021-11-10]

81. Swinkels K. SARS-CoV-2 Superspreading Events from Around the World. URL: https://docs.google.com/spreadsheets/d/1c9jwMyT1lw2P0d6SDTno6nHLGMtpheO9xJyGHgdBoco/edit [accessed 2021-10-16]

82. Revollo B, Blanco I, Soler P, Toro J, Izquierdo-Useros N, Puig J, et al. Same-day SARS-CoV-2 antigen test screening in an indoor mass-gathering live music event: a randomised controlled trial. The Lancet Infectious Diseases 2021 Oct 16;21(10):1365-1372 [FREE Full text] [doi: 10.1016/S1473-3099(21)00268-1] [Medline: 26745243]

83. Llibre JM, Videla S, Clotet B, Revollo B. Screening for SARS-CoV-2 antigen before a live indoor music concert: An observational study. Ann Intern Med 2021 Jul 20:1. [doi: 10.7326/m21-2278]

84. COVID event risk survey. URL: https://tinyurl.com/2p8z5z6d [accessed 2021-10-16]

Abbreviations

ARC: Applied Research Collaboration
kNN: k-nearest neighbor
NIHR: National Institute for Health Research
PCR: polymerase chain reaction
RAH: Royal Albert Hall
SAR: secondary attack rate
SSE: super-spreader event

XSL•FO

RenderX