



**HAL**  
open science

## Assessment of the risks of viral transmission in non-confined crowds

Willy Garcia, Baptiste Fray, Alexandre Nicolas

► **To cite this version:**

Willy Garcia, Baptiste Fray, Alexandre Nicolas. Assessment of the risks of viral transmission in non-confined crowds. 2020. hal-03060324v1

**HAL Id: hal-03060324**

**<https://hal.science/hal-03060324v1>**

Preprint submitted on 13 Dec 2020 (v1), last revised 3 Nov 2021 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Assessment of the risks of viral transmission in non-confined crowds

Willy Garcia<sup>a</sup>, Baptiste Fray<sup>a,b</sup>, and Alexandre Nicolas<sup>a,1</sup>

<sup>a</sup> Institut Lumière Matière, CNRS & Université Claude Bernard Lyon 1 & Université de Lyon, F-69622, Villeurbanne; <sup>b</sup> École nationale des travaux publics de l'État (ENTPE), Université de Lyon, F-69518, Vaulx-en-Velin

This manuscript was compiled on December 13, 2020

**This work aims to assess the risks of Covid-19 disease spread in diverse daily-life situations (referred to as scenarios) involving crowds of unmasked pedestrians, mostly outdoors. More concretely, we develop a method to infer the global number of new infections from patchy observations of pedestrians. The method relies on *ad hoc* spatially resolved models for disease transmission via virus-laden respiratory droplets, which are fit to existing exposure studies about Covid-19. The approach is applied to the detailed field data about pedestrian trajectories and orientations that we acquired during the pandemic. This allows us to rank the investigated scenarios by the transmission risks that they present; importantly, the obtained hierarchy of risks is conserved across all our transmission models (except the most pessimistic ones): Street cafés present the largest average rate of new infections caused by an attendant, followed by busy outdoor markets, and then metro and train stations, whereas the risks incurred while walking on fairly busy streets (average density around 0.1 person/m<sup>2</sup>) are comparatively quite low. Models that assume isotropic transmission of the virus fail to reproduce these results. In scenarios with a moving crowd, we find that density is the main factor influencing the estimated infection rate. Finally, our study explores the efficiency of street and venue redesigns in mitigating the viral spread: While the benefits of enforcing one-way foot traffic in (wide) walkways are unclear, changing the geometry of queues substantially affects disease transmission risks.**

Covid-19 | Crowd dynamics | Epidemiology

**E**fficient collective action to curb the spread of epidemics in general, and of the current Covid-19 pandemic in particular, requires input from a variety of disciplinary fields, from microscale fluid dynamics (to understand the propagation of virus or bacteria-laden droplets (1, 2)) to macroscale epidemiology. At present, the weak link between these two scales hinders the prediction of how the SARS-CoV-2 virus at the origin of the pandemic will spread in a given crowd.

Gatherings of people are encountered both in enclosed spaces (such as restaurants, offices, private accommodation or fitness centers), where statistical data may be insightful *a posteriori* from an epidemiological standpoint (3–5), as well as in non-confined environments. Most Covid-19 outbreaks are certainly associated with indoor settings (6), but nonetheless a minority of clusters – at least a few percent, as a tentative estimate (5, 7–9) – reportedly originate in outdoor or mixed indoor/outdoor settings, e.g., on building sites. Viral transmissions outdoors raise a specific challenge because they are inherently hard to trace and document, but also hard to circumscribe, as they bring together unrelated people\*. These difficulties are a hurdle to the control of outbreaks.

Accordingly, recommendations to wear a face covering out-

side have been issued far and wide. Some cities in China, France, part of Germany, Italy, Poland, Portugal, Singapore, South Korea, Spain, some Swiss cantons, and a number of US states, among others, have put in place mask mandates for some, or all, outdoor activities. Let us say from the outset that mask-related policies may have a broader impact than their chief purpose of physically warding off infections (11–13): Widespread usage of face coverings attracts every one's attention to the health situation and may thus promote stronger vigilance and abiding by other sanitary measures. Furthermore, these policies are constrained by the legal possibilities in place in each country, the indirect consequences of the measures, public perception, and an imperative of simplicity. On the downside, there is some discomfort associated with wearing a mask and it might still be too early to assess the psychological impact of being surrounded by covered faces and not being able to recognize people's emotions.

Thus, a proper assessment of the risks of infections incurred by unmasked crowds in diverse non-confined spaces is much needed. Not only can it provide more objective foundations to public policies, but it is instrumental in better targeting the situations where risks are highest and masks are most crucial, thus justifying stricter local control, and determining if (and where) it makes sense to restrict pedestrian mobility on streets and sidewalks.

For the time being, there remains a gap between the thriving experimental and computational studies focused on measuring the emission and propagation of respiratory droplets (2, 14–21) and the investigations of disease spread at larger scales (22–24). The former have shed light on the complex, turbulent

## Significance Statement

Crowds of unrelated people in outdoor settings are a blind spot of Covid-19 outbreak statistics, because infections in these settings are inherently hard to trace. We develop a method to estimate the number of new infections in a crowd using partial observations of pedestrian motion and simple spatially resolved models for disease transmission via respiratory droplets; these models are specified to fit the information available about SARS-CoV-2. Combined with field data on crowds, this allows us to (i) rank daily-life situations involving crowds by the rate of new infections that they are expected to cause and (ii) assess the efficiency of some redesigning strategies in mitigating the spread of the virus.

A.N. conceived the research. W.G., B.F. and A.N. performed the experiments. W.G. and A.N. analyzed the data. A.N. and W.G. wrote the manuscript. All authors proofread the manuscript.

The authors declare no competing interests.

<sup>1</sup> To whom correspondence should be addressed (alexandre.nicolas@polytechnique.edu).

\* To wit, more than three out of four new infections in France cannot be linked to a reported Covid-19 case as of November 2020 according to national statistics (10).

dynamics that take place inside the exhaled puff and called into question both the binary distinction between falling droplets and tinier airborne aerosols (2), and the scientific basis of the 2-meter social distancing rule (25, 26). However, the translation of these results into epidemiologically relevant predictions is uneasy, and trailing. Poles part from these microscopic studies, risk assessments at the scale of a facility or venue by means of agent-based pedestrian simulations (24, 27, 28) or large-scale experiments (29) resort to particularly crude assumptions with regard to viral transmission. Generally, an individual is considered exposed to the disease when he or she comes in a given radius (e.g., 2 meters) around an infected person, regardless of their orientations, overlooking that their head orientations control the direction in which respiratory droplets are expelled. Moreover, pedestrian dynamics models are hardly designed to reproduce fine observables such as precise inter-pedestrian spacings with any reliability, neither in usual times nor in times of pandemic, when pedestrian behaviors and trajectories are altered to reduce infection risks (30, 31).

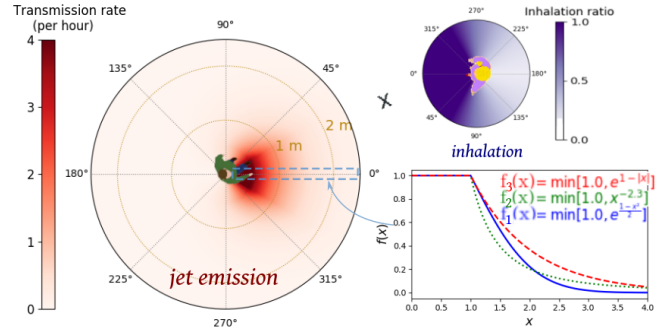
To overcome these strong limitations, we collected detailed field data <sup>†</sup>, during the pandemic in France, about pedestrian separations and orientations in diverse situations (hereafter called scenarios), either outdoors or in large, ventilated indoor facilities, and we developed a mathematically sound method to infer the hourly rate of new infections in each scenario from our partial observations. The method rests on simple *ad hoc* models for viral transmissions, which we introduce and fit to droplet emission data and existing exposure studies. While none of these simple models can claim to be accurate, they all converge towards a fairly robust ranking of the scenarios in terms of infection risks. The proposed framework is also useful to quantify the effect of enhanced physical distancing and to assess the mitigation efficiency of redesigning certain premises (one-way footpaths, queues, etc.).

## Coupling observations of crowds with spatially resolved models of viral transmission

While one can rely on estimated numbers of contacts between people to model the spread of an epidemic at regional or national scales (22, 23), more detailed information about the viral transmission route and the interactions between people is required to gauge how a virus will propagate in a given crowd. Short-range transmission via virus-laden respiratory droplets and long-range transmission via air-borne residues (aerosols) may both be possible, but the former is widely believed to prevail over the latter in non-confined environments, at least for influenza and the coronavirus (17, 19, 32). These droplets are exhaled when breathing, talking, shouting, panting, coughing or sneezing, mostly through the mouth but also through the nose (14, 16); the focus must thus be put on their propagation.

### Modeling viral transmission via respiratory droplets.

Unfortunately, detailed fluid dynamical studies of droplet and aerosol propagation in these scenarios would not only be extremely demanding computationally, but also hinge on very specific information that is neither available nor transferable between situations, e.g., ambient air flows (26), wind speed



**Fig. 1.** Spatially resolved model of disease transmission via virus-laden respiratory droplets. The transmission rate (Eq. 2) depends on the direction of the emitter’s head, the distance between the emitter and the inhaler, and the latter’s head orientation; these dependencies are all modeled with a decaying function,  $f_1$ ,  $f_2$ , or  $f_3$ . Optimistic parameters combined with  $f_3$  ( $ModOpt_3$ ) are used in the illustration.

(21), humidity, and speech details. Therefore, we opted for the development of coarser-grained, *ad hoc* models.

With the insight into droplet propagation provided by computational studies as well as experiments on expiration and inhalation (2, 14, 18, 21, 33), we modeled the disease transmission rate between a contagious individual  $E$  (the emitter) and a susceptible one (labeled  $R$ , for ‘receiver’) as a generic function  $\nu(r, \theta^E, \theta^R)$  that decreases with the horizontal distance  $r$  between the individuals’ heads and the orientations  $\theta^E$  and  $\theta^R$  of the emitter’s and receiver’s heads relative to the direction of the vector that connects them. We then defined a family of plausible parameter sets for the  $\nu$  functions, whose values span the entire spectrum from highly contagious (‘pessimistic’) to only mildly contagious (‘optimistic’); see the *Methods* for details. An example of such a function is depicted in Fig. 1. The spatial decay of the transmission function is such that  $\nu(r, \theta^E, \theta^R)$  becomes negligible past a few meters, except for the worst-case model describing uncovered sneezes (15). The typical time for infection in the event of a close ( $r = 50$  cm) face-to-face contact lies between 5 minutes (an extremely pessimistic estimate) and 30 minutes (except for sneezes), consistently with the epidemiological literature on SARS-CoV-2, outbreak reports and exposure case studies (8, 10, 13, 34–36).

The resulting models are representative of a disease that may be transmitted to multiple individuals within an hour in unfavorable cases (7, 37), but generally requires close and prolonged contacts for transmission: Even in households, reported attack rates lie in the range 5%-30% (38), often around 15% (39–41); moreover, casual episodic contacts at work or in the community, even face to face, do not necessarily trigger an outbreak (34, 39). More quantitatively, the spatial decay described by the fairly optimistic models agrees well with the decay of the concentration of droplets emitted by a coughing subject (33) (see Fig. S2). The reliability of the transmission models is further tested by mimicking a journey aboard a Chinese train (overlooking its confined nature), where transmission risks for passengers who sat close to an infected traveler were recently assessed using trip records (4). Overall, the simulated results, detailed in Appendix C, are compatible with the empirical data and the models featuring the most *optimistic* parameters display the best concordance. Finally, let us mention that the order of magnitude of the

<sup>†</sup> All our field data will be made freely available on a public repository. Meanwhile, they can be accessed under: alexandrenicolanet/separationspietons

parameters (especially the most optimistic ones) is consistent with the putative minimal infectious dose of SARS-CoV-2, estimated to be order 100 (42), granted that most exhaled droplets contain 0 or 1 viral copy (43) (see Appendix B).

Although these pieces of evidence favor the optimistic end of the parameter spectrum, our study is conducted with the whole gamut of plausible model parameters. This variety better reflects our current uncertainty with regard to the transmission of SARS-CoV-2, but also the established inter-individual and inter-case variability (14, 16), depending on physiology, talking characteristics, etc.

### Field measurements and inference of the rate of new infections.

We used a privacy-respective setup (see the *Methods* section) to film different scenarios with non-confined crowds (listed in Table S1) and we extracted the distances between pedestrians, the orientations of their heads, and their group membership. This was done by means of exhaustive semi-manual tracking, which allowed high accuracy and almost error-free processing. We compensated for undetected contacts with off-camera people by appraising the fraction of missed interactions as a function of their range and re-weighting the detected ones; Appendix A demonstrates the efficiency of this method.

For each scenario the time and space-resolved pedestrian measurements are then coupled to the above transmission models. This directly yields the instantaneous rate, abbreviated as  $\nu_{ij}(t)$ , at which a supposedly infected index patient  $i$ , that we will call Iago, transmits the disease to other pedestrians  $j$  around him via virus-laden droplets. Under the independent action hypothesis (IAH) (44, 45), each inhaled virus is equally likely to lead to an infection, with no cooperation or antagonism between viruses. It follows that, over the time interval  $[t_0, t_0 + \tau_i]$  over which Iago was filmed, he infected a number  $C_i^{(\tau_i)}$  of other people  $j$  (leaving aside his fellow group members  $G_i$ , whose possible infection is not specifically related to the scenario), given by a Wells-Riley-like equation (46), viz.

$$C_i^{(\tau_i)} = \sum_{j \notin G_i} S_j^0 \cdot (1 - e^{-N_{ij}}), \quad [1]$$

where  $S_j^0$  is the probability that  $j$  is susceptible (i.e., *not already* infected) at the beginning of the observation interval, and  $N_{ij} = \int_{t_0}^{t_0 + \tau_i} \nu_{ij}(t) dt$  is the cumulative transmission risk (47).

It is worth noticing that the IAH implies that infections are a stochastic process without threshold: Any encounter can potentially result in a new case, and multiple short interactions with various people are as risky as a single long one, and even riskier because, once infected, agent  $j$  can no longer be infected, viz.,  $S_j^0 : 1 \rightarrow 0$ . This saturation of the risk complicates the evaluation of  $C_i^{(\tau_i)}$  even if we assume that all pedestrians *except Iago's group  $G_i$  of co-walkers* are initially susceptible, because our videos only record part of Iago's wandering and may thus miss earlier off-camera interactions. Nonetheless, rigorous bounds on  $C_i$  can be set by noticing that, on the one hand,  $S_j^0 \leq 1$  and that, on the other hand,  $\sum_{j \notin G_i} (1 - S_j^0)$  cannot be larger than the number of people actually infected by Iago, which is related to  $C_i$  (see Appendix C). Finally, for comparison purposes,  $C_i^{(\tau_i)}$  is recast into an hourly rate of new infections  $C_i \equiv C_i^{(\Delta T)} = \frac{\Delta T}{\tau_i} C_i^{(\tau_i)}$  with  $\Delta T = 1$  h, assuming that the recorded videos are representative.

Static scenarios – namely, the cafés and waiting lines – are handled slightly differently, because then Iago's neighbors do not change significantly as time passes, in which case  $N_{ij} = \frac{\Delta T}{\tau_i} \int_{t_0}^{t_0 + \tau_0} \nu_{ij}(t) dt$  and  $S_j^0 = 1$  in Eq. 1. In a nutshell, the proposed framework enables us to quantitatively translate patchy observations, with undetected contacts, into an estimated global risk of viral spread.

### Ranking of scenarios by the risks of new infections.

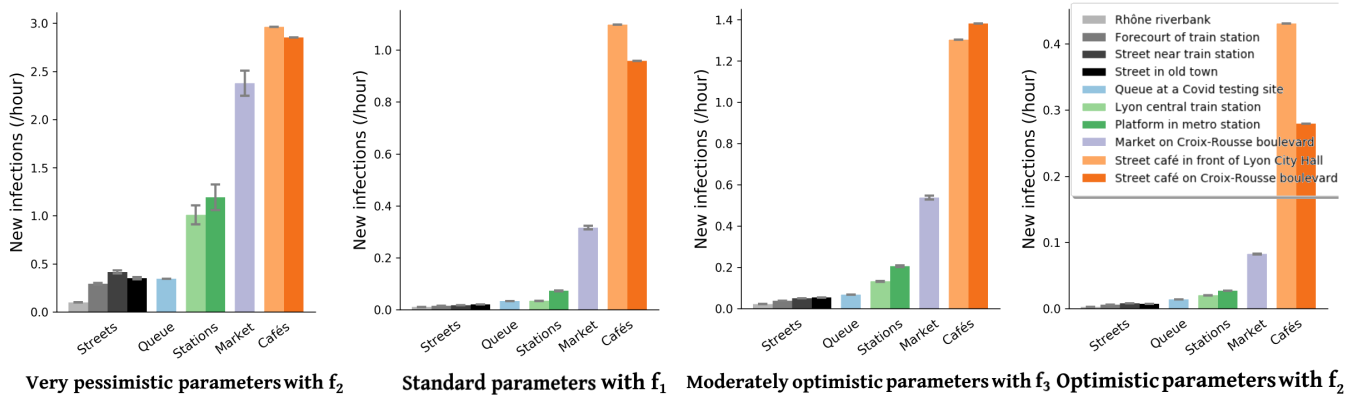
Inserting the collected field data into this framework, we obtain upper and lower bounds on the mean rate  $C = \langle C_i \rangle_i$  of new infections per hour for each scenario and each transmission model. Figure 2 presents a sample of results for four of these models. These results confirm the efficiency of the proposed bound-setting method, as the bounds are found to confine  $C$  to a narrow interval. Most importantly, the ranking of the different scenarios turns out to be robust, that is to say, largely preserved across models. This is our first major result.

Pursuing the analysis of Fig. 2, we observe that street cafés show the highest risks in terms of the mean number of new infections per hour, even though their tables were more spaced when the videos were shot than they used to be before the pandemic. These infections at cafés are easily rationalized by the close, face-to-face interactions between people sharing a table, let alone the increased emission of droplets associated with lively discussions and eating, which is overlooked here. This result is in line with case reports of high risks of viral transmission while dining and drinking (3, 37, 47). Next in line among the observed scenarios comes the outdoor market alley. Despite its high average density  $\rho \simeq 0.5$  ped/m<sup>2</sup>, this scenario never matches the level of risk at cafés, bar with the very pessimistic parameters corresponding to high contagiousness. Further down the list, crowd density explains the considerably higher risks at train and metro stations ( $\rho \simeq 0.25$  ped/m<sup>2</sup>) compared to fairly busy streets in Lyon ( $\rho \simeq 0.05 - 0.1$  ped/m<sup>2</sup>) and, to an even larger extent, the riverbank walkway that we filmed. Somewhat intriguingly, the estimated infection rate is at least as large, or even larger, at the observed testing site in Lyon than it is on these streets, although the overall density there is low and attendants were strictly asked to stay 2 meters apart from each other; yet, their relative proximity was prolonged over considerable time and, besides, they tended to turn and pace around a bit while waiting. One should however bear in mind that our models estimate the risks of viral transmission if no face mask is worn, whereas everybody was wearing a mask at the testing site that we filmed.

### Hourly rates of new infections.

Beyond this robust ranking, it is worth noticing that the estimated infection rates vary substantially, but not hugely across the full spectrum of plausible models. The pessimistic estimates are generally at most a factor 10 larger than the (possibly more relevant) most optimistic ones – if one excepts the sparse, comparatively low-risk situations, where the divergence about how very low the risk actually is may reach two to three orders of magnitude. Thus, it is reasonable to conclude that contagious Iago will infect a number of order 1 persons if he sits at a café for one hour, while he would probably cause significantly less than  $\sim 0.1$  new infections if he spent this time walking on a fairly busy street.

Nonetheless, these average rates mask the variety of pedestrian contacts in the different scenarios, which is better re-



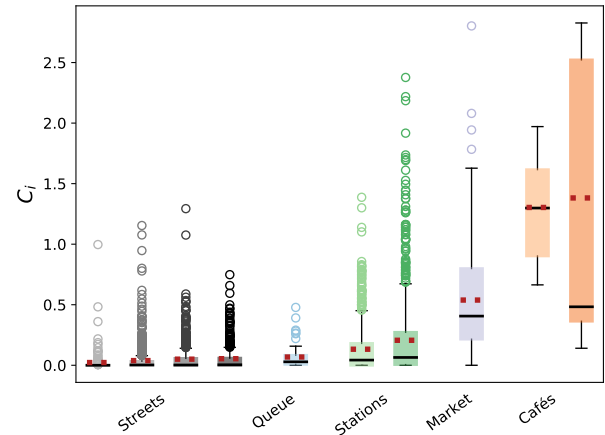
**Fig. 2.** Rates  $C$  of new infections per hour in the scenarios under study, estimated with four different parameter sets for the transmission model. Except in the static scenarios (cafés and queue), infections within groups have been dismissed and the error bars span the interval between the estimated lower bound  $\underline{C}^{(1\text{ h})}$  and  $\bar{C}^{(1\text{ h})}$ , while the filled bars represent  $\frac{1}{2} (\underline{C}^{(1\text{ h})} + \bar{C}^{(1\text{ h})})$ .

flected in the box plots of Fig. 3. The figure shows that, while the scenarios involving a moving crowd cause fewer infections than cafés on average, their rates of infections  $C_i$  are more dispersed and, unlike cafés, they feature many outliers on both ends, i.e., at  $C_i \simeq 0$  as well as at high- $C_i$ , the latter being pedestrians that fortuitously turn into super-spreaders because of their pattern of on-street contacts. As we shall see below, the blame does not necessarily rest on the pedestrian, but rather on the ebbs and flows of crowding in each observed situation.

Prior to that, let us remark that accounting for the directionality of droplet propagation and describing the orientations of pedestrians’ heads was crucial in our endeavor. Indeed, not only does an isotropic transmission model overestimate risks in crowds by a factor of at least 10 in comparison to its directional counterpart, but it also alters the ranking of scenarios: It predicts considerably more infections at the outdoor market than at the cafés (Fig. S4). Otherwise, such an inversion (along with high risk estimates) is only found for our worst-case transmission models, in particular the model that we introduced to mimic the effect of a contagious patient sneezing every few minutes without covering his or her sneezes. These observations may therefore call for revisions of the widespread risk assessment method whereby a pedestrian is considered exposed to the disease as soon as he stands within a certain radius around an index patient (24, 27–29). On the other hand, allowing infections within groups, as we did for the cafés, does not dramatically change the picture, even though it substantially heightens the risks associated with sparse situations, for instance, the riverbank walkway. This is not surprising because in these situations close contacts mostly occur between group members.

### Key determinants of the transmission rate.

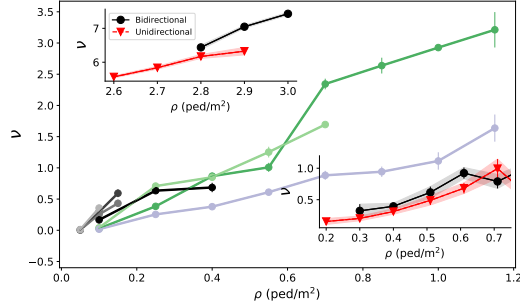
To better understand the observed disparities, we need to identify the key variables that determine the level of risk. Figure 4 confirms the intuition that the instantaneous pedestrian density  $\rho(t)$  is a major determinant of the rate of viral transmission  $\nu(t) \equiv \left\langle \sum_{j \notin G_i} \nu_{ij}(t) \right\rangle_i$  (where the average is taken over all pedestrians observed at time  $t$ ), in that it controls how close each pedestrian is to their counterparts. (Note that all time-dependent variables have been averaged over intervals of



**Fig. 3.** Hourly rates of new infections  $C_i$  caused by the different pedestrians  $i$  in each scenario, as estimated with *ModOpt*<sub>3</sub>.

two seconds, to reduce the statistical noise.) The variation of  $\nu$  with  $\rho$  looks similar across scenarios, but is not strictly identical, which indicates that other scenario-dependent variables affect the transmission rate  $\nu$ . Furthermore, these variations become more muddled as one turns to more optimistic parameter sets, which is consistent with the idea that one then probes the configuration of the crowd at finer length-scales, owing to the shorter transmission range. The total flow rate could in principle play a role; however, at fixed density  $\rho$ , we found that  $\nu$  does not follow any clear trend with this flow rate at fixed density (Fig. S5).

On the other hand, head orientations naturally have a bearing on the risks of infection, as evinced by the failure of isotropic transmission models to reproduce our results, but we now show that in non-static scenarios these orientations can be practically inferred using only the trajectories. To do so, we notice that the head orientations of walking pedestrians (speed  $v > 0.3\text{ m} \cdot \text{s}^{-1}$ ) are approximately normally distributed around their walking direction, with a standard deviation around  $26^\circ$ , and decorrelate over one second. Therefore, we choose to ascribe angular orientations randomly drawn from this normal distribution to walkers, while their stationary



**Fig. 4.** Dependence of the rate of viral transmission  $\nu(t)$  on the instantaneous density  $\rho(t)$ , in different scenarios, using the *ModOpt<sub>3</sub>* parameter set. Refer to Fig. 2 for the color code. In the *insets*, the rates are distinguished depending on the directionality of the flow in the (*bottom right*) outdoor market scenario, (*top left*) corridor flow experiments of the BaSiGo project (48). The error bars and envelopes represent standard errors (i.e.,  $\pm \frac{\text{std}}{\sqrt{n}}$ , where  $n$  is the number of uncorrelated data points).

counterparts ( $v \leq 0.3 \text{ m} \cdot \text{s}^{-1}$ ) are considered purely randomly oriented; the random values are refreshed every second. Quite interestingly, this simple reconstruction of head orientations yields mean infection rates  $C$  per hour that agree very well with the *bona fide* values, with a relative difference generally lower than 15%, regardless of the transmission model. The correspondence between the individual  $C_i$  values for each pedestrian is of course imperfect with this method, but overall the differences are not extremely large (Fig. S6). These observations are particularly relevant to bolster risk assessments based on observed or (reliably) simulated pedestrian datasets in which head orientations are missing, as they most often are.

## Practical implications and concluding remarks

### Insight into the risk in the scenarios under study.

In summary, the foregoing risk assessments in non-confined environments can guide public decisions in times of pandemic, in that (irrespective of the transmission parameters that are used) they confirm the risks of infection incurred at cafés (3) and underline the key role of pedestrian density in determining the rate of viral transmission in moving crowds. Fairly busy streets, with densities up to  $\rho \simeq 0.1 \text{ ped/m}^2$ , are found to present risks that are not entirely negligible, but comparatively low, and these risks will manifestly reach even lower values for less busy streets. This suggests that the scant reports of outbreaks in outdoor walking crowds are not only due to the intricacy of tracing back these infections (due to unidentified contacts, recall biases, etc.), but also to the limited transmission of the virus in these conditions, even without face coverings. Nevertheless, this remark does not apply to very crowded settings such as markets or metro and train stations, which deserve particular attention.

As a matter of fact, the spatial resolution of our empirical data and models provides deeper insight into the new infections expected in the above scenarios; it can contribute to the debate about what physical distance should be prescribed between pedestrians in non-confined environments, and whether 2 meters or '6 feet' are enough from an epidemiological perspective (25, 26). Admittedly, the answer will heavily rest on the transmission model (which was here established in an *ad hoc* way), but the statistics of inter-pedestrian contacts in

the scenario also play a large role. Using the cumulative rate of transmission  $\int \nu(t) dt$  as a proxy for the incurred risk, we find that its dominant contribution comes from interactions within a distance of 1 meter (for instance, 70% at Bellecour metro station, with *OptMod<sub>3</sub>*), whereas transmission beyond 2 meters, albeit possible, accounts for only a few percent of the risk (5% at Bellecour), at most. It is worth remarking that this statement holds for all (non-static) scenarios and all investigated models, with the exception of the worst-case ones (involving highly pessimistic parameters or mimicking uncovered sneezes) when they are applied to crowded settings.

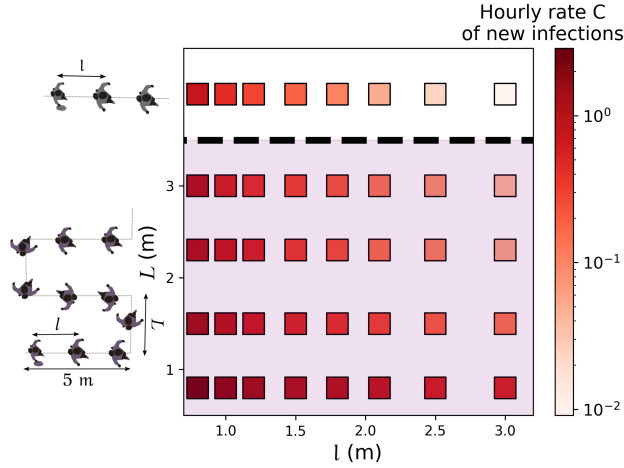
### Mitigation efficiency of redesigns.

Beyond these conclusions rooted in actual observations, the framework introduced here opens the door to evaluating the mitigation efficiency of hypothetical redesigns of streets and venues, consisting e.g. of enforcing one-way circulation on footpaths, a sitting plan at cafés, or increased spacing in queues. Since circulation plans have flourished during the pandemic, let us first explore the impact of one-way *vs.* two-way foot traffic on sidewalks and pedestrian streets.

To avoid potential situational biases, the question is investigated by separating the periods of time (binned in two-second intervals) when the flow was unidirectional from those when there were pedestrians going in both directions, in each given scenario – the market alley in the bottom right inset of Fig. 4. Since the transmission risks were found to depend on density, but not on the total flow rate (i.e., the sum of the directional flow rates across sections perpendicular to the main flow), we perform a comparison at fixed density. Our data (inset of Fig. 4) reveal only little benefit to switching from two-way to one-way traffic in our wide-path scenarios. To further test this somewhat surprising finding, we exploit the *controlled* experiments performed a few years ago by the German BaSiGo team<sup>‡</sup> to study unidirectional and bidirectional pedestrian flows in 4 to 5-meter-wide corridors (48); head orientations are reconstructed as explained above. The results, shown in the top left inset of Fig. 4, are in line with the aforementioned finding: In wide walkways, switching from two-way flow to one-way flow seems to only moderately reduce the risks, probably because head-on 'collisions' are rare in these self-organized flows.

Next, we turn to queues and how their arrangement affects transmission risks. On the basis of our observations at a testing site, we allow people in the line to move in a  $50 \text{ cm} \times 50 \text{ cm}$  rectangle around their central spot and consider their head orientations to be normally distributed around the queuing axis (with a standard deviation of  $22^\circ$ ) 75% of the time and purely random for the remaining 25% of the time, due to people turning around or having a look around. Both the positions and orientations are refreshed every second. We checked that this naive reconstruction of a whole scenario (which actually features two different, not strictly linear queues, one outdoors and one indoors) predicts infection rates that are within 50% of their actual values, for all models – except the worst-case ones, which bestow special importance to rare contact events that are overlooked in the reconstitution. Figure 5 illustrates the extent to which predicted infection rates vary when the spacing between queuing people or the queuing geometry are modified.

<sup>‡</sup> These extensive datasets are openly available under: <https://ped.fz-juelich.de/db/doku.php>



**Fig. 5.** Hourly rate of new infections in a linear queue (top row) and a winding queue, depending on the spacing between pedestrians and lines, as estimated with the *ModOpt*<sub>3</sub> parameter set. People move one spot forward every other minute. Note the logarithmic color scale on the graph.

### Current limitations and perspectives.

Our study pioneers the coupling of empirical crowd data to models of viral transmission at mesoscales. As such, it undoubtedly suffers from some limitations. To start with, the empirical data could be extended to include more scenarios and longer footage. For instance, in our recordings we did not observe any fortuitous encounter; such encounters may however be a factor of risk, due to the close contacts and talks that they induce.

Perhaps more crucially, the transmission model should be refined. The models used in this study are admittedly overly simple, even though this problem was partly warded off by ensuring the robustness of our conclusions across diverse model variants. More sophisticated models, which may differentiate transmission rates as a function of people’s activity (reflecting known variations in droplet emission (16, 18)) and account for the effect of the wind (21) and propagation dynamics (49), will afford more accurate estimation of the rate of new infections. In addition, fluid dynamics simulations of long-range aerosol propagation would make it possible to study enclosed spaces with poor ventilation, where our current models that discard the airborne transmission route can only provide lower bounds on infection risks.

Still, the most pressing task might be to generalize the transmission models to people who are (adequately or inadequately) wearing a mask (11), in order to determine how serious an issue very crowded streets really are in current times. It would be straightforward to account for the particle filtration efficiency of masks in the present framework, by simply multiplying the transmission rates by a reduction factor (say,  $\sim 20\%$  for cloth masks,  $\sim 10\%$  for surgical masks and  $\ll 5\%$  for N95 masks (33, 38), if *only the emitters* have their face covered), but this probably underestimates the efficiency of masks, which also reduce the reach of the exhaled puff (12), thereby probably shortening the range of transmission of droplets (33).

### Materials and Methods

	$T_0$ (min)	$r_0$ (m)	$\theta_0^E$	$\theta_0^R$
Optimistic	30	0.3	22.5°	45°
Moderately optimistic	15	0.5	30°	60°
Standard	10	0.5	30°	60°
Pessimistic	10	0.75	45°	60°
Very pessimistic	7.5	1	45°	60°
Extremely pessimistic	5	1	45°	90°
Uncovered sneezes	1.7	1.5	22.5°	60°

**Table 1.** Parameter sets used in the transmission model of Eq. 2. The angular values  $\theta_0^E$  and  $\theta_0^R$  correspond to the half-angles of the emission and inhalation cones in the horizontal plane.

**Video acquisition, pedestrian tracking, and contact rescaling.** Pedestrian flows and crowds were filmed in diverse scenarios in a discrete and passive way with a TomTom Bandit camera covered with a thin plastic layer (to deteriorate the quality of the image) and installed in a zenithal position. This privacy-respective setup was approved by the Data Protection Officer of CNRS.

After some pre-processing with the FFMpeg software to correct the ‘fish-eye’ effect, downsample the video, and select the area of interest (from  $8\text{ m}^2$  to  $46\text{ m}^2$ ), the positions and head orientations of all pedestrians were manually tracked at a rate of 2 frames per second (fps) with the help of a dedicated Python software; linear interpolation then increased the rate to 10 fps. Overall, close to 5,000 pedestrian trajectories were thus obtained. Pixel coordinates were converted into real-world coordinates with a geometric formula whose parameters are fit thanks to predefined calibration points at pedestrian height (see Appendix A. The estimated experimental error on the positions is typically below or around 20 cm, while that on the head orientations is lower than  $20^\circ$  in most videos.

Because of the limited field of view, some interactions with off-camera people were missed, especially at large separations. We compensated for this by numerically gauging the fraction of interactions thus lost, under the assumption of homogeneous density, and reweighting the detected contacts accordingly. We checked that this rescaling method successfully restores the genuine contact distribution, up to contact distances close to the size of the field of view (Fig. S1).

**Ad hoc spatially resolved models for viral transmission.** The proposed transmission models are premised on direct, fairly short-ranged propagation of droplets or aerosols (17). The process is directional in terms of the emitter’s relative orientation  $\theta^E$  (exhaled jet) and, to a lesser extent, that of the receiver,  $\theta^R$  (note that droplet emission and inhalation are not symmetric (18)). Assuming that these variables can be decoupled, we write the instantaneous transmission rate as

$$\nu(r, \theta^E, \theta^R) = \frac{1}{\tilde{T}_0} \bar{f}\left(\frac{r}{r_0}\right) \cdot \bar{f}\left(\frac{\theta^E}{\theta_0^E}\right) \cdot \bar{f}\left(\frac{\theta^R}{\theta_0^R}\right), \quad [2]$$

where  $\bar{f}(x)$  is a function such that  $\bar{f}(x \approx 0) = 1$  and  $\bar{f}(x)$  decays rapidly for  $x \gg 1$ . To be concrete, we tested the following three functions,

$$f_1(x) = \exp\left(\frac{1-x^2}{2}\right); \quad f_2(x) = |x|^{-m}; \quad f_3(x) = \exp(1-|x|)$$

Because of the limited accuracy of our positional measurements and the uncertainties about very near field transmission, the peaks of these functions are leveled at  $x \rightarrow 0$ , viz.,  $f_k = \min(1, f_k)$  for  $k = 1, 2, 3$ . The parameters  $r_0$ ,  $\tilde{T}_0$ ,  $\theta_0^E$ , and  $\theta_0^R$  in Eq. 2 and  $m$  in  $f_2$  were bounded using a broad scope of available empirical data, which suggest a characteristic distance  $r_0 \leq 1\text{ m}$ , an infection time (at  $r_0 = 0.5\text{ m}$ )  $T_0 = \tilde{T}_0 / \bar{f}(\frac{0.5}{r_0})$  between a dozen minutes and an hour, and an exponent  $m \approx 2 - 2.5$ . More details are given in Appendix B. Within these plausible bounds, we explored different sets of values, listed in Table 1.

**ACKNOWLEDGMENTS.** We are grateful to Isabelle Sabran (Ville de Lyon), Jessica Magraner (Cour d’appel de Lyon), and Frederic

Laurent (Hospices Civils de Lyon) for facilitating our collection of data with our privacy-respective cameras and to the French MODCOV19 initiative for supporting part of this work. We also thank Cécile Appert-Rolland for lending us some material and Marina Nicolas for proofreading. This work was funded by Agence Nationale de la Recherche (ANR program RA-COVID-19 V1) under project name *SeparationsPietons*.

---

DRAFT



## A. Empirical observations of pedestrian crowds

**Video acquisition and processing.** Videos of crowds were acquired from a top view in the non-confined settings described in Table S1, in a discrete and passive way that allowed us to observe natural behavior as pedestrians were mostly unaware of being filmed.

Scenario / Location	Date and time	Comments	Number of tracked pedestrians	Mean Density (ped/m <sup>2</sup> )
Pedestrian banks of the Rhône river, close to the Morand Bridge	Friday, July 3rd 2020, 4pm-5pm	Sunny and windy day	164	<b>0.042</b>
Plaza in front of Perrache hub (Hall - Level 1)	Friday, July 17th 2020, 9am-10am	Nice weather	1021	<b>0.038</b>
Part-Dieu train station - Ground level/passage area (indoors)	Friday, July 17th 2020, 11:20am-12:40pm	Sunny day	875	<b>0.22</b>
Busy street - Under the Passerelle Bouchut	Wednesday, July 29th 2020, 12pm-1pm	Hot and sunny day	800	<b>0.05</b>
Bellecour subway station - Platform of Line D (indoors)	Tuesday, July 28th 2020, 3pm-4pm	Hot and sunny day	849	<b>0.26</b>
Croix-Rousse boulevard - street cafés	Friday, September 11th 2020, 6:30pm-7pm	Nice weather	13	/
Saint-Jean street in the Old Town of Lyon	Monday, September 21st 2020, 1pm-2pm	Hot and cloudy day	481	<b>0.11</b>
Place des Terreaux - Bar/Restaurant terraces	Thursday, October 8th 2020, 5:40pm-6pm	Sunny day, cool weather	30	/
Croix-Rousse - Main market alley	Sunday, October 18th 2020, morning	Sunny day	183	<b>0.46</b>
COVID-19 testing site	Monday, October 19th 2020, 8:30am-10:30am	Outdoor waiting lines (cold day) - Indoor waiting lines (sport arena)	66	/

**Table S1. Details about the investigated scenarios. All sites are in the metropolitan area of Lyon, France; most are outdoors.**

After correcting for lens deformation, the trajectories and head orientations were extracted by extensive manual tracking with the help of a dedicated Python script and a touch screen. The pixel coordinates  $(x', y')$  are converted into real-world coordinates  $(X', Y')$  using the following geometric relation,

$$\begin{cases} X &= X_c + \gamma(\nu x - X_c) \\ Y &= Y_c + \gamma(\nu y - Y_c) \end{cases}$$

where  $\gamma = \frac{1}{1 - \sin \alpha \cdot \frac{y}{D}}$ ,  $\alpha$  is the angle between the camera axis and the vertical direction (we systematically found  $\alpha \simeq 0$ , reflecting the top view used in virtually all videos). Here, and  $(X, Y)$  and  $(x, y)$  are connected to  $(X', Y')$  and  $(x', y')$ , respectively, by rotations in Euclidean space. The unknown parameters in this relation are estimated by fitting at least 4 calibration points at predefined positions, where a team member stood in each scenario.

The expected uncertainty on positions created by failing to account for height differences ( $\delta h$ ) between individuals is expected to be around  $\delta x \simeq \tan \alpha' \cdot \delta h \approx 15$  cm in the present conditions, where  $\alpha'$  is the angle in which the pedestrian was filmed with respect to the vertical. In practice, by double-tracking some test pedestrians, we estimate that the uncertainty is typically below or around 20 cm. The same method allowed us to evaluate the standard deviation of the error on head orientations to  $19^\circ$ . Some videos, especially for the static scenarios, were filmed from a more distant viewpoint, in which case the error on the orientations is likely to be larger.

**Contact rescaling to compensate for undetected interactions.** The finite field of view of the camera results in some missed interactions with off-camera people. We can estimate the fraction  $\phi(r)$  of missed contacts depending on the separation distance  $r$  under an assumption of uniform density. This is achieved numerically by randomly placing points (pedestrians) in an area with the same dimensions as the view field and testing if secondary points inserted at a distance  $r$  of the first one in a random direction are inside or outside the area. Then, all detected interactions are multiplied by a factor  $\frac{1}{1 - \phi(r)}$  to compensate for those that went undetected; for practical purposes, we set a maximal value (20) to this factor, to avoid excessive amplification of the statistical noise at large separation distances  $r$ .

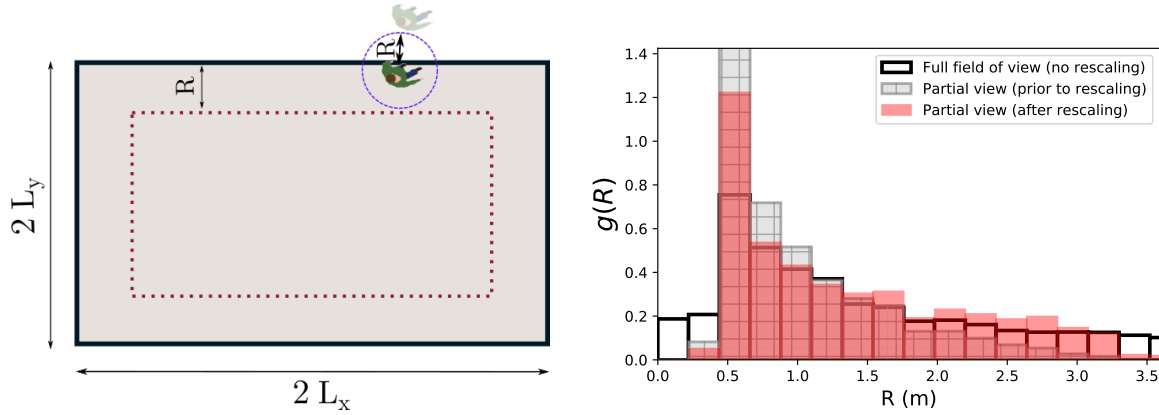
This method is put to the test by focusing on a small area of dimensions 3 m  $\times$  3 m in the view field of a given scenario and computing the radial distribution function  $g(r)$  of contacts (i.e.,  $g(r) \propto \frac{N(r)}{2\pi r}$  where  $N(r)$  is the number of contacts of range  $r$ ) in the 'partial' field before and after applying the rescaling method. Figure S1 demonstrates the efficiency of the method: After rescaling,  $g(r)$  gets much closer to the distribution measured in the original field of view (up to statistical noise, since only a fraction of the people are then observed).

## B. Ad hoc models for viral transmission

This supplementary section details and discusses the *ad hoc* functions that were introduced to model the transmission rate  $\nu$ , defined such that  $\nu dt$  is the probability for a *susceptible* individual to be infected by the index patient over the interval  $dt$ .

**Specification of transmission models.** We start by recalling the generic expression chosen for these functions (Eq. 2 of the main paper),

$$\nu(r, \theta^E, \theta^R) = \frac{1}{T_0} \bar{f}\left(\frac{r}{r_0}\right) \cdot \bar{f}\left(\frac{\theta^E}{\theta_0^E}\right) \cdot \bar{f}\left(\frac{\theta^R}{\theta_0^R}\right), \quad [\text{S.1}]$$



**Fig. S1.** Contact rescaling method. (*Left*) Sketch illustrating the possibility to miss interactions with off-camera people. (*Right*) Radial distribution function  $g(r)$  in the Perrache plaza scenario measured using the original field of view or a reduced field with and without application of the contact rescaling method.

emphasizing that  $\theta^E$  and  $\theta^R$  are the orientations of the pedestrians' heads, hence mouths and noses, relative to the vector that connects them, in the *horizontal plane*, i.e., irrespective of the difference of height. Accordingly,  $\theta^E = \theta^R = 0$  corresponds to a face-to-face interaction.

The formulation of Eq. S.1 is arbitrary to a large extent, but its main features are derived from experimental and numerical data. Experimental images of inhalation (50) and expiratory emissions during breathing, speaking (18), or coughing (15), as well as computational fluid dynamics simulations (17, 18, 21), brought to light strong *directed* transport of respiratory droplets over tens of centimeters, reaching a couple of meters; the exhaled cloud is *mostly enclosed* in a cone of half-angle between  $10^\circ$  and  $50^\circ$  at most, depending on the speech conditions (18), thus pointing to a significantly smaller characteristic angle  $\theta_0^E$ . In passing, note that such conical propagation is well described by Eq. S.1, in which  $r$  and  $\theta^E$  are decoupled. Inhalation is also directional, but, as ambient air from all around the mouth will be breathed in (18), we have systematically that its directionality was less marked, i.e.,  $\theta_0^R > \theta_0^E$  (see Table 1 of the main paper).

Regarding the characteristic distance  $r_0$ , a review of 172 observational studies (36) tentatively suggests that keeping a physical distance of 1 meter reduces the infection risk from 13% (very close contact) to 3% and down to 1.5% if a distance of 2 meters is maintained, although there is very large dispersion in the data. In the UK, the Scientific Advisory Group on Emergencies (SAGE) asserts that physically distancing by 2 meters instead of 1 reduces the risks by a factor between 2 and 10 (13). This translates into an exponent  $m \approx 2 - 3$  in function  $\bar{f}_2$ , a value  $r_0 \in [0.4\text{ m}, 1.4\text{ m}]$  if one uses function  $\bar{f}_3$  or  $r_0 \in [1\text{ m}, 2\text{ m}]$  if one uses function  $\bar{f}_1$ . However, at larger physical distances, contacts may have been less in a face-to-face orientation or shorter in duration, so these ranges ought to be taken with a grain of salt. Chen et al.'s theoretical work rather suggests  $r_0 \approx 0.2\text{ m}$  for talking and  $r_0 \approx 0.5\text{ m}$  for coughing (see Fig. 7 of (17)).

As for the time parameter, it is most easily expressed as  $T_0 = \bar{T}_0 / \bar{f}(\frac{0.5}{r_0})$  which is the characteristic duration for infection at a distance  $r = 0.5\text{ m}$ . Contact tracing routines generally consider a threshold duration between 5 and 15 minutes. Nonetheless, this value may be at the safe end of the spectrum, as exposure case studies have only reported very few cases with contacts lasting less than 15 minutes.

Finally, we also modeled uncovered sneezes, which are longer ranged than other expiratory events but also highly directional (15). Even though the emitted jet is reported to reach 7 or 8 meters, not all droplets are expected to travel this far, as seen in the experimental images of (15). Accordingly, we settled for  $r_0 = 1.5\text{ m}$ , which means that the risk at a distance of 7 meters is 1% of that at 50 cm with the  $\bar{f}_3$  function, 3% with  $\bar{f}_2$ , but close to zero with  $\bar{f}_1$ .

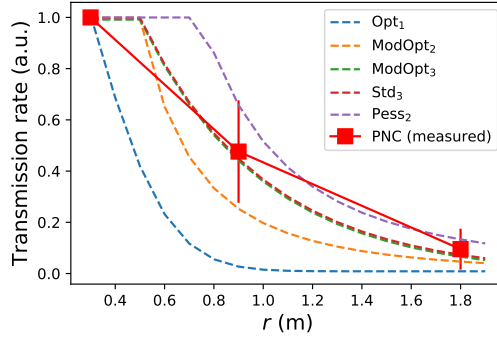
There is no denying that the foregoing *ad hoc* models and their parameters are but coarse approximations of the reality: They do not account for the effect of wind, ventilation, or humidity (21), nor do they describe the dynamics of exhaled jets and the time lag due to their propagation. Perhaps more importantly, the variability in droplet emission and viral shedding between individuals is overlooked, as is their dependence on their activity (breathing through the nose or the mouth, panting, talking loud, coughing, etc.), which is known to have a major impact (14, 16, 51). However, since virtually all scenarios under consideration (with the exception of the cafés) involve similar activities, i.e., very moderate physical activity and limited talking, we expect our approach to hold in a *statistical* sense, although different index patients may be associated with different parameter sets in the model. The characteristic duration  $T_0$  is set by noting that a single sneeze may expel 10 times as many respiratory droplets as 5 minutes of talking (43), but generally does not occur more than once every few minutes. This is particularly true if the focus is on *uncovered* sneezes. Accordingly, as a *worst-case* estimate, we set  $r_0 = 100\text{ s}$ .

**A. Consistency with the current knowledge about infection risks and droplet emission.** Let us now examine to what extent case reports may support the proposed models.

Generally, outbreak reports do not specify the very precise circumstances of the infections. Still, the attack rates in diverse settings are useful indications. Indeed, within households, reported attack rates range from 5% to 30% (38), for instance around 15% in (39–41). At work or in the community, casual episodic contacts, even face to face, do not necessarily trigger an outbreak of cases (34, 39). Furthermore, it was reported that at Solano County hospital (California, US) the majority of healthcare personnel did not get infected despite spending 10 to 50 minutes in the same room as a Covid patient, often within 2 meters and with no facial mask (35). All these pieces of evidence hint at an average time for infection  $T_0$  in a close face-to-face contact that is probably longer than  $\sim 10$  minutes.

Another general indication comes from the basic reproductive number  $R_0$  in the pre-pandemic context, which was around  $R_0 \approx 3$ . Contact pattern data in this usual, pre-pandemic context imply that only 7% of contacts lasting longer than 15 minutes lead to an infection. Assuming prolonged contacts of median duration 20 minutes that are mostly face to face (leaving aside height differences), this leads to a characteristic time for infection  $T_0$  of a couple of hours if the contact distance is only 50 cm, or varying between 20 minutes and 2 hours for a contact distance of 80 cm, depending on the functional choice  $\bar{f}_1$ ,  $\bar{f}_2$  or  $\bar{f}_3$  and the chosen value for  $r_0$ . Thus, this gives credence to the 'optimistic' end of our parameter spectrum.

Further insight is provided by empirical measurements of the particle number concentration of the respiratory droplets exhaled by a (healthy) subject (33). The concentration of droplets of sizes  $0.5\ \mu\text{m} - 20\ \mu\text{m}$  was measured at different distances in front of a subject during coughing, with and without face covering. In Fig. S2, we compare the spatial decay of these measurements (without face covering) to the predictions of some of our transmission models. Excellent agreement is found with the moderately optimistic and standard models



**Fig. S2.** Comparison between the spatial decay of the modeled transmission rate  $\nu(r, \theta^E = 0, \theta^R = 0)$  with distance  $r$  and the particle number concentration (PNC) measurements during coughing of (33). Arbitrary units have been chosen so that the value at  $r = 0$  m is always unity.

The following abbreviations are used:  $Opt_1$  (optimistic parameters combined with  $f_1$ ),  $ModOpt_2$  (moderately optimistic parameters combined with  $f_2$ ),  $ModOpt_3$  (moderately optimistic parameters combined with  $f_3$ ),  $Std_3$  (standard parameters combined with  $f_3$ ),  $Pess_2$  (pessimistic parameters combined with  $f_2$ ).

combined with  $f_3$ . Yet, one should bear in mind that the measurements were made while the subject was coughing, which may increase the transmission range. Therefore, these data once again support the most optimistic of our transmission models.

**B. Order of magnitude of droplet emission.** One may also wonder how compatible the model parameters are with the current fundamental knowledge about droplet emission. Exhaled droplets are very broadly distributed in size around a mean value of  $\sim 5 \mu\text{m}$  and contain  $\sim 0.1$  viral copy (mostly zero or one copy) per droplet (43). Talking for five minutes, or coughing once in these 5 minutes (which is a relevant frequency according to coughing statistics (2)) will expel about  $3 \cdot 10^3$  droplets. It is difficult to estimate what fraction of these will be breathed in by a person standing 50 cm apart, but, following the reasoning of (18), droplets should be diluted by a factor  $\sim 0.1$  at this distance, so it is reasonable to estimate that at most a few % of the droplets will be inhaled, a significant fraction of which may deposit in the nasopharynx, where they are most likely to lead to an infection (42). Assuming an infectious dose of  $\mathcal{O}(100)$  (42), i.e., that only one in a hundred virions will successfully invade a host cell and replicate, one arrives at a characteristic time for infection  $T_0$  of several hours ( $T_0 \lesssim 10\text{h}$ ). Notwithstanding how rough an estimate this is, it is worth noticing that it is once again rather in line with the very optimistic end of our spectrum of plausible values.

**C. Direct comparison with the collated statistics of infections aboard Chinese trains.** It is generally difficult to test transmission models against empirical data related to Covid-19 cases in a statistically meaningful way because the detailed interactions between the index patient and his or her contacts are seldom known. However, Hu et al.'s study of transmission risks among train passengers in China (4) provides a virtually unique opportunity to attempt such a direct comparison. Indeed, Hu et al. were able to retrieve the trip record of confirmed Covid-19 patients who had traveled on a train in the 14 days before the onset of the illness, between December 2019 and late February 2020. Having access to the detailed seating plan on their train coaches, the researchers then tracked reported Covid-19 cases among co-travelers and computed the infection probability depending on the distance (expressed as a number of rows and columns) between the index patients and the contacts. A marked increase in the infection risk was found for people seated in the same row as the index patient, especially in adjacent seats, for more than a couple of hours.

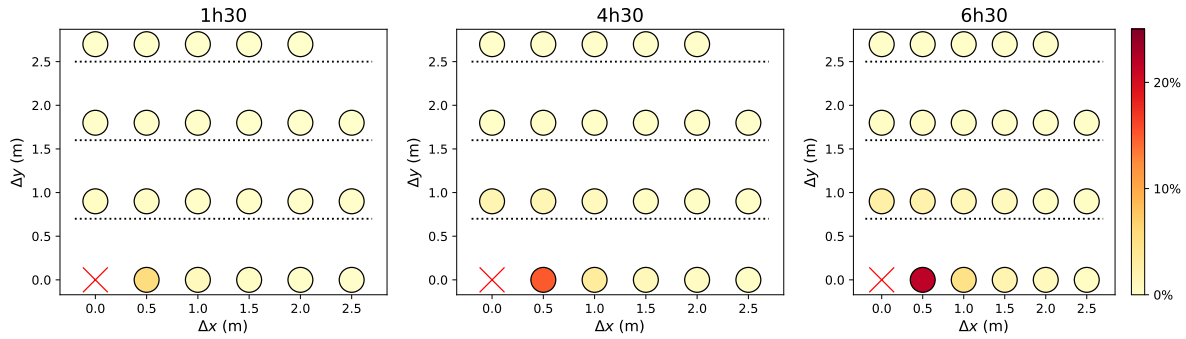
Leaving aside the fact that a train coach is an enclosed (but usually ventilated) space, the transmission models introduced above can be applied to these settings. Distances between seats and rows are precisely known. The central aisle is counted as one seat, as in the original study (4), while the barrier to transmission created by seat rows (and represented by dashed lines in Fig. S3) is arbitrarily considered to have an effect comparable to an additional distance of 1 meter. In the light of our empirical measurements on moving and queuing crowds (see the main text), the passengers' head orientation is assumed to be normally distributed around the axial direction, with a standard deviation of  $26^\circ$ ; we consider that the seated position suppresses the rotations of the whole torso that are occasionally observed in standing pedestrians.

Figure S3 presents the probabilities of infection evaluated on this basis using three different transmission models on the optimistic side of the parameter spectrum, with the help of Eq. S.3. These results are directly confronted with Fig. 2 of (4); the three of them are found to compare well as far as both the evolution of the risks with co-travel duration and their spatial pattern are concerned.

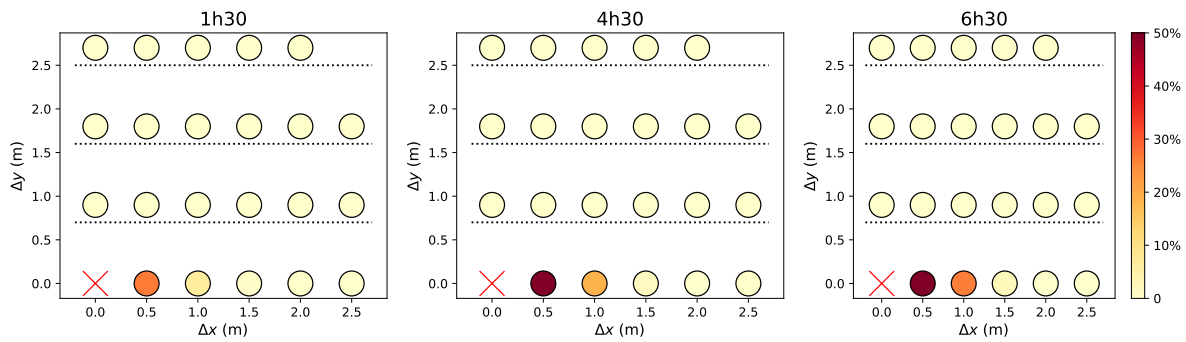
A note of caution should nevertheless be made about the absolute values of infection probabilities, which peak at about 10% in (4), compared to 25% to 80% here. This very notable difference may largely be due to the fact that all trips of the index patient in the two weeks before his or her diagnosis are taken into account in (4) whereas it is very unlikely that the patient was contagious during all this period. If the contagious period before the onset of symptoms spans two to three days (38, 52), i.e., about 15% to 20% of the two-week period, then applying this 15-20% ratio to the simulated infection probabilities yields values that become comparable to those reported by (4). These pieces of evidence thus give credence to our transmission models coupled with optimistic parameter sets.

Another caveat should however be mentioned regarding Hu et al.'s work (4): The researchers admit that they were not able to distinguish relatives from unrelated people in their data. Therefore, part of the reported infections may not have taken place aboard the train, but elsewhere.

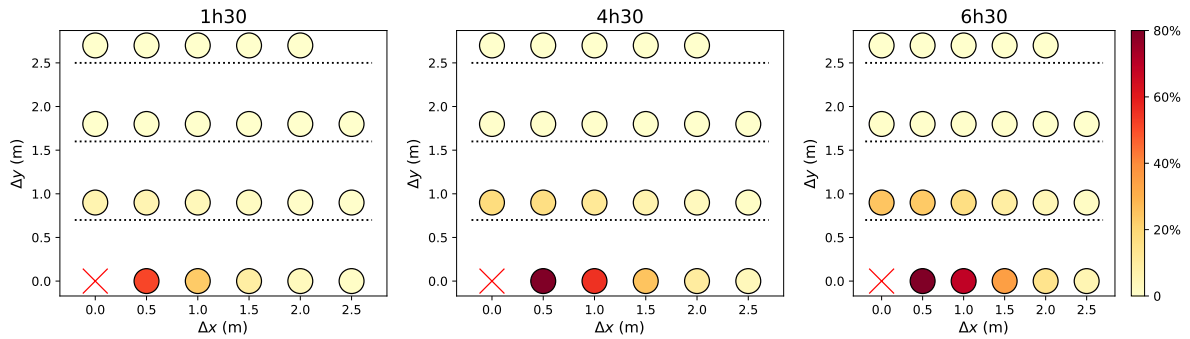
Optimistic parameters combined with  $f_2$ :



Moderately optimistic parameters combined with  $f_1$ :



Moderately optimistic parameters combined with  $f_3$  ( $ModOpt_3$ ):



**Fig. S3.** Simulated probability of infection of co-passengers aboard the train, depending on their proximity to the index patient (red cross) and the time spent together on the train.  $\Delta x$  denotes the *absolute* distance along each row and  $\Delta y$  denotes the distance along each column; the physical barriers created by seats are materialized as a dashed line. Model parameters are specified above each row of charts.

### C. Transmission

Our study relies on the independent action hypothesis introduced by Druett in the early 1950s (44) and partially validated by Zwart in virus-insect pathosystems a decade ago (45). It posits that there is no minimal infectious dose and that one can overlook cooperative or antagonistic interactions between virions in the host system and assign to each of them a finite probability (here denoted by  $\epsilon$ ) to cause an infection. This yields a Wells-Riley-like equation (46) for the time evolution of the contagion status  $I_j(t)$  of (initially susceptible) agent  $j$ , Juliette, infected by index patient  $i$ , Iago ( $I_j$  is the probability that Juliette is infected), viz.,

$$\dot{I}_j(t) = [1 - I_j(t)] \nu_{ij}(t), \quad [\text{S.2}]$$

$$I_j(t) = 1 - e^{-N_{ij}(-\infty, t)}, \text{ with } N_{ij}(s, t) = \int_s^t \nu_{ij}(t') dt' \quad [\text{S.3}]$$

where  $\nu_{ij}(t) = \epsilon n_{ij}(t)$  is the risk transmission rate and  $n_{ij}(t)$  is the number of viral copies exchanged between Iago and Juliette per unit time around time  $t$ . Accordingly, if Iago was filmed between time  $t_0$  and time  $t_0 + \tau_i$ , the number of new infections that were actually witnessed (or 'event  $R$ ' in the terminology of (47)) can be estimated to

$$\begin{aligned} C_i^{(\tau_i)} &= \sum_{j \notin G_i} I_j(t_0 + \tau_i) - I_j(t_0) \\ &= \sum_{j \notin G_i} S_j^0 \cdot [1 - e^{-N_{ij}(t_0, t_0 + \tau_i)}] \end{aligned} \quad [\text{S.4}]$$

where  $G_i$  contains Iago and all other agents that we visually identified as members of his group (family, friends, etc.) and  $S_j^0 = 1 - I_j(t_0) = e^{-N_{ij}(-\infty, t_0)}$  is the probability that Juliette was already infected at  $t_0$ . Note that similar assumptions about the rate of infections were made in a recent paper that we came across just before submission (47). Let us emphasize in particular that intra-group infections are discarded here, because group members are likely to have been infected outside the scenario of interest, whereas our purpose is to estimate *new* infections in the scenario.

**Scenarios involving a moving crowd.** To compute Eq. S.4, an issue must be overcome. Indeed, due to their limited temporal and spatial span, our observations may not capture all interactions between Iago and other pedestrians, in the non-static scenarios. Therefore, the infection status  $S_j^0$  at  $t_0$  is difficult to evaluate. To circumvent the issue, we derive upper and lower bounds on  $C_i$  that do not require specific information about the  $S_j^0$ .

The upper bound is straightforward, because  $S_j^0 \leq 1$  by definition, which implies that  $C_i^{(\tau_i)} \leq \bar{C}_i^{(\tau_i)} \equiv \sum_{j \notin G_i} 1 - e^{-N_{ij}(t_0, t_0 + \tau_i)}$ . This value can be converted into an hourly rate of new infections,  $\bar{C}_i^{(1h)}$ , by multiplying it with  $\frac{1h}{\tau_i}$ .

The lower bound requires more careful examination, wherein we will set a maximal value to the number of people that have *already* been infected at  $t_0$ , under the hypothesis that only Iago is contagious on the premises. Let  $H_i = \{j \text{ s.t. } j \notin G_i \text{ and } N_{ij}(t_0, t_0 + \tau_i) > 0\}$  be the group of unrelated pedestrians  $j$  with whom Iago came in contact while filmed, so that this number reads  $\sum_{j \in H_i} I_j(t_0)$ . If Iago and Juliette ( $j$ ) are unrelated, it is sensible to consider that they have not been close to one another for more than a duration  $\delta\tau$  in the global scenario (we will set  $\delta\tau = 5$  min for most scenarios), viz.,

$$\sum_{j \in H_i} I_j(t_0) \leq \sum_{j \notin G_i} 1 - e^{-N_{ij}(t_0 - \delta\tau, t_0)} \leq \bar{C}_i^{(\delta\tau)}, \quad [\text{S.5}]$$

where  $\bar{C}_i^{(\delta\tau)}$  is our upper-bound estimate for the number of infections during  $\delta\tau$ . Note that the foregoing inequality manifestly holds if Iago and Juliette's interaction took place in the interval  $[t_0 - \delta\tau, t_0]$ , but it will also hold should the interaction have occurred earlier, provided that our video is representative. Indeed, under this proviso, Juliette is equally likely to have already been infected at  $t_0$  as any *random* pedestrian in Iago's vicinity in the interval  $[t_0 - \delta\tau, t_0]$  (note that these *random* pedestrians are more numerous than those observed in  $[t_0, t_0 + \tau_i]$ , because it was almost always verified that  $\tau_i < \delta\tau = 5$  min). It follows from Eq. S.5 that

$$\sum_{j \in H_i} S_j^0 \geq \#H_i - \bar{C}_i^{(\Delta T)}.$$

One is thus left with an optimization under constraints, wherein one has to minimize (i.e., find a lower bound for)

$$C_i^{(\tau_i)} = \sum_{j \in H_i} S_j^0 \cdot [1 - e^{-N_{ij}(t_0, t_0 + \tau_i)}] \quad [\text{S.6}]$$

under the following constraints on the variables  $S_j^0$ ,

$$\begin{aligned} \forall j \in H_i, 0 \leq S_j^0 \leq 1 \\ \sum_{j \in H_i} S_j^0 \geq \#H_i - \bar{C}_i^{(\Delta T)}. \end{aligned}$$

This optimization problem may for instance be solved using Lagrange's multipliers. The minimum is reached at  $(S_j^0)_{j \in H_i}$  such that  $S_j^0 = 1$  for all  $j \in H_i$ , *except*

- the  $n \equiv \left\lfloor \bar{C}_i^{(\Delta T)} \right\rfloor$  indices  $j$  exhibiting the highest values  $N_{ij}(t_0, t_0 + \tau_i)$ , which are set to  $S_j = 0$ ,
- $j = n$ , which is set to  $S_n^0 = 1 - \left( \bar{C}_i^{(\Delta T)} - \left\lfloor \bar{C}_i^{(\Delta T)} \right\rfloor \right)$ .

In other words, this boils down to treating as infected the  $n$  unrelated people who interacted most closely with Iago on the video. The lower bound  $\underline{C}_i^{(\tau_i)}$  that we were seeking is obtained by computing Eq. S.6 with the above  $(S_j^0)_{j \in H_i}$ , which concludes our search.

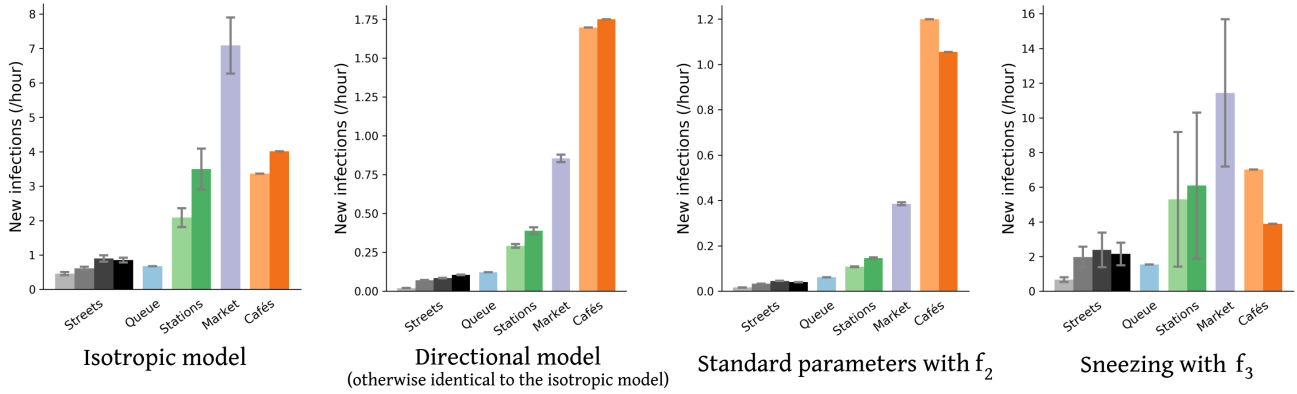
Finally, let us note that the foregoing reasoning can easily be extended to a population with partial immunity, by replacing the initial conditions  $S_j^0 = 1$  with  $S_j^0 = \alpha \cdot [1 - I_j(t_0)]$ , where  $\alpha$  is the fraction of immunized people.

**Static scenarios.** In the 'static' scenarios, i.e., the streets cafés and the queue, interactions are less changing. Therefore, we consider that the interactions that are observed on the videos are prolonged over the whole period of study, i.e., 1 hour to get an hourly rate. Accordingly, the risk increments  $N_{ij}(t_0, t_0 + \tau_i)$  in Eq. S.3 are simply multiplied by  $\frac{1}{\tau_i}$ , where  $\tau_i$  generally coincides with the duration of the video, and the hourly rate of new infections is given by

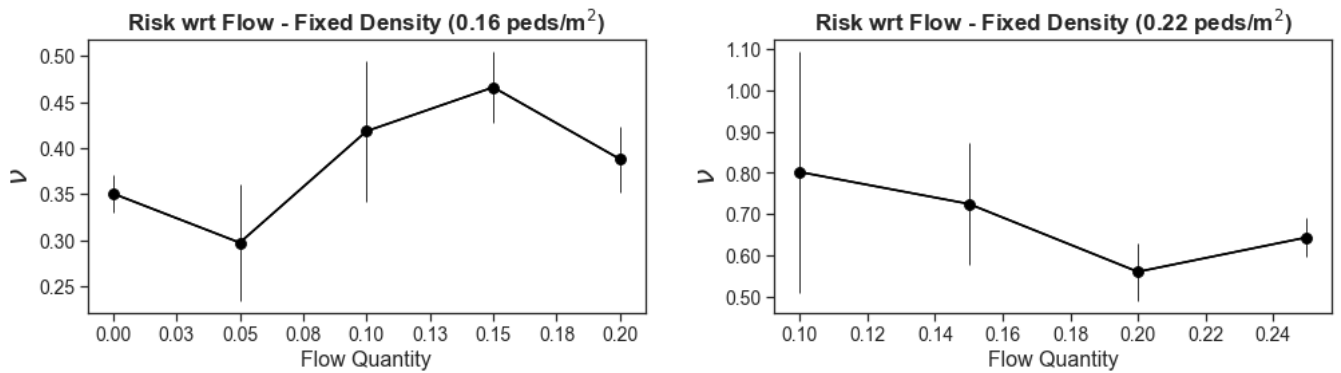
$$C_i^{(1\text{ h})} = \sum_j \left[ 1 - e^{-\frac{1}{\tau_i} \cdot N_{ij}(t_0, t_0 + \tau_i)} \right].$$

A remark should be made here. The sum is not restricted to people that are unrelated to Iago, contrary to the case of moving crowds, and we assume that everybody apart from Iago is initially susceptible,  $S_j^0 = 1$ . Indeed, infections at a café are counted starting from the moment when people meet and sit together. Besides, on these videos, we are not able to distinguish Iago's household members (who may have been infected beforehand) from his other relatives or friends, so everybody is assumed initially susceptible.

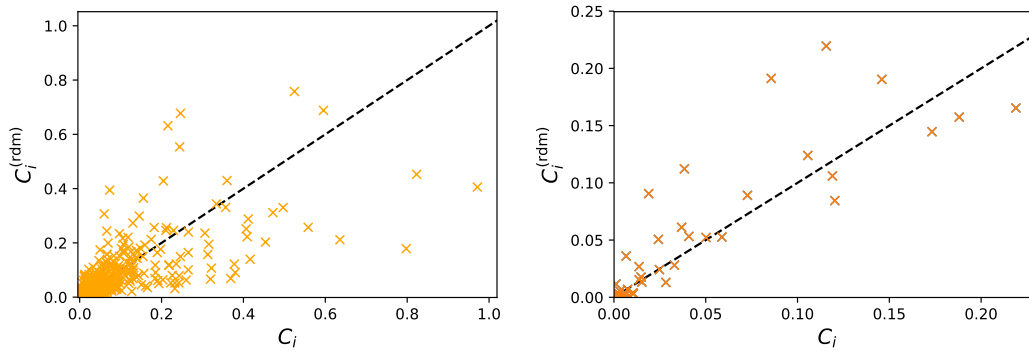
## D. Additional figures



**Fig. S4.** Hourly rates of new infections in the scenarios under study, estimated with four transmission parameter sets. The isotropic model and the directional one ( $\theta_0^E = \frac{\pi}{6}$ ,  $\theta_0^R = \frac{\pi}{3}$ ) share the same parameters  $T_0 = 15$  min and  $r_0 = 1$  m. The error bars span the interval between the lower bound  $\underline{C}^{(1\text{ h})}$  and  $\bar{C}^{(1\text{ h})}$ , while the filled bars represent  $\frac{1}{2} (\underline{C}^{(1\text{ h})} + \bar{C}^{(1\text{ h})})$ .



**Fig. S5.** Variations of the transmission rate  $\nu$  with the total flow rate in a street in the Old Town of Lyon (*left*) and along an outdoor market alley (*right*), at fixed pedestrian density, with *ModOpt*<sub>3</sub>.



**Fig. S6.** Correspondence between the *bona fide* individual-based hourly rates of new infections  $\bar{C}_i$  (upper bound) and the values  $\bar{C}_i^{(rdm)}$  reconstructed under the assumption of random head orientations (see the main text), on the plaza in front of the Perrache train station (*left*) and on the Rhône riverbank (*right*), with  $ModOpt_3$ .

1. Mittal R, Ni R, Seo JH (2020) The flow physics of covid-19. *Journal of fluid Mechanics* 894.
2. Bourouiba L (2020) The fluid dynamics of disease transmission. *Annual Review of Fluid Mechanics* 53.
3. Chang S, et al. (2020) Mobility network models of covid-19 explain inequities and inform reopening. *Nature* pp. 1–8.
4. Hu M, et al. (2020) The risk of covid-19 transmission in train passengers: an epidemiological and modelling study. *Clinical Infectious Diseases*.
5. Leclerc QJ, et al. (2020) What settings have been linked to sars-cov-2 transmission clusters? *Wellcome Open Research* 5(83):83.
6. Nishiura H, et al. (2020) Closed environments facilitate secondary transmission of coronavirus disease 2019 (covid-19). *MedRxiv*.
7. Shen Y, et al. (2020) Community outbreak investigation of sars-cov-2 transmission among bus riders in eastern china. *JAMA Internal Medicine*.
8. for Disease Prevention EC, Control (2020) Covid-19 clusters and outbreaks in occupational settings in the eu/eea and the uk, (European Centre for Disease Prevention and Control), Technical report.
9. Weed M, Foad A (2020) Rapid scoping review of evidence of outdoor transmission of covid-19. *medRxiv*.
10. France SP (2020) Covid-19 point épidémiologique hebdomadaire du 03 décembre 2020, (Santé Publique France), Technical report.
11. Leung NH, et al. (2020) Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nature medicine* 26(5):676–680.
12. Bahl P, et al. (2020) Face coverings and mask to minimise droplet dispersion and aerosolisation: a video case study. *Thorax* 75(11):1024–1025.
13. on Emergencies SAG (2020) Transmission of sars-cov-2 and mitigating measures (june 4th, 2020), Technical report.
14. Morawska L, et al. (2009) Size distribution and sites of origin of droplets expelled from the human respiratory tract during expiratory activities. *Journal of Aerosol Science* 40(3):256–269.
15. Bourouiba L, Dehandschoewerker E, Bush JW (2014) Violent expiratory events: on coughing and sneezing. *Journal of Fluid Mechanics* 745:537–563.
16. Asadi S, et al. (2019) Aerosol emission and superemission during human speech increase with voice loudness. *Scientific reports* 9(1):1–10.
17. Chen W, Zhang N, Wei J, Yen HL, Li Y (2020) Short-range airborne route dominates exposure of respiratory infection during close contact. *Building and Environment* p. 106859.
18. Abkarian M, Mendez S, Xue N, Yang F, Stone HA (2020) Speech can produce jet-like transport relevant to asymptomatic spreading of virus. *Proceedings of the National Academy of Sciences* 117(41):25237–25245.
19. Bao L, et al. (2020) Transmission of sars-cov-2 via close contact and respiratory droplets among haca2 mice. *The Journal of Infectious Diseases*.
20. Li T, Liu Y, Li M, Qian X, Dai SY (2020) Mask or no mask for covid-19: A public health and market study. *PLoS one* 15(8):e0237691.
21. Feng Y, Marchal T, Sperry T, Yi H (2020) Influence of wind and relative humidity on the social distancing effectiveness to prevent covid-19 airborne transmission: A numerical study. *Journal of aerosol science* p. 105585.
22. Ferguson N, et al. (2020) Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. *Imperial College London* 10:77482.
23. Ferretti L, et al. (2020) Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science* 368(6491).
24. Harweg T, Bachmann D, Weichert F (2020) Agent-based simulation of pedestrian dynamics for exposure time estimation in epidemic risk assessment. *arXiv preprint arXiv:2007.04138*.
25. Morawska L, Milton DK (2020) It is time to address airborne transmission of coronavirus disease 2019 (covid-19). *Clinical Infectious Diseases*.
26. Jones NR, et al. (2020) Two metres or one: what is the evidence for physical distancing in covid-19? *BMJ* 370.
27. Xiao Y, et al. (2020) Modeling indoor-level non-pharmaceutical interventions during the covid-19 pandemic: a pedestrian dynamics-based microscopic simulation approach. *arXiv preprint arXiv:2006.10666*.
28. Romero V, Stone WD, Ford JD (2020) Covid-19 indoor exposure levels: An analysis of foot traffic scenarios within an academic building. *Transportation Research Interdisciplinary Perspectives* 7:100185.
29. Moritz S, et al. (2020) The risk of indoor sports and culture events for the transmission of covid-19 (restart-19). *medRxiv*.
30. Pouw CA, Toschi F, van Schadewijk F, Corbetta A (2020) Monitoring physical distancing for crowd management: Real-time trajectory and group analysis. *PLoS one* 15(10):e0240963.
31. Ronchi E, Scozzari R, Frontè M (2020) A risk analysis methodology for the use of crowd models during the covid-19 pandemic, (Lund University, Department of Fire Safety Engineering), Technical Report 3235. LUTVDG/TVBB.
32. Freeman S, Eykelbosh A (2020) Covid-19 and outdoor safety: Considerations for use of outdoor recreational spaces. *National Collaborating Centre for Environmental Health*.
33. Li L, Niu M, Zhu Y (2020) Assessing the effectiveness of using various face coverings to mitigate the transport of airborne particles produced by coughing indoors. *Aerosol Science and Technology* pp. 1–11.
34. Burke RM, et al. (2020) Enhanced contact investigations for nine early travel-related cases of sars-cov-2 in the united states. *medRxiv*.
35. Heinzerling A, et al. (2020) Transmission of covid-19 to health care personnel during exposures to a hospitalized patient—solano county, california, february 2020, (US Centers for Disease Control and Prevention), Technical report.
36. Chu DK, et al. (2020) Physical distancing, face masks, and eye protection to prevent person-to-person transmission of sars-cov-2 and covid-19: a systematic review and meta-analysis. *The Lancet*.
37. Li Y, et al. (2020) Evidence for probable aerosol transmission of sars-cov-2 in a poorly ventilated restaurant. *medRxiv*.
38. Bar-On YM, Sender R, Flamholz AI, Phillips R, Milo R (2020) A quantitative compendium of covid-19 epidemiology. *arXiv preprint arXiv:2006.01283*.
39. Park SY, et al. (2020) Coronavirus disease outbreak in call center, south korea. *Emerging Infectious Disease Journal* 26(8).
40. Park YJ, et al. (2020) Contact tracing during coronavirus disease outbreak, south korea, 2020. *Emerging infectious diseases* 26(10):2465–2468.
41. Burke RM (2020) Active monitoring of persons exposed to patients with confirmed covid-19—united states, january–february 2020. *MMWR. Morbidity and mortality weekly report* 69.
42. Basu S (2020) Close-range exposure to a covid-19 carrier: transmission trends in the respiratory tract and estimation of infectious dose. *medRxiv*.
43. Poon WC, et al. (2020) Soft matter science and the covid-19 pandemic. *Soft matter* 16(36):8310–8324.
44. Druett H (1952) Bacterial invasion. *Nature* 170(4320):288–288.
45. Zwart MP, et al. (2009) An experimental test of the independent action hypothesis in virus–insect pathosystems. *Proceedings of the Royal Society B: Biological Sciences* 276(1665):2233–2242.
46. Sze To GN, Chao CYH (2010) Review and comparison between the wells–riley and dose–response approaches to risk assessment of infectious respiratory diseases. *Indoor air* 20(1):2–16.
47. Tupper P, Bourry H, Yerlanov M, Colijn C (2020) Event-specific interventions to minimize covid-19 transmission. *Proceedings of the National Academy of Sciences*.
48. Cao S, Seyfried A, Zhang J, Holl S, Song W (2017) Fundamental diagrams for multidirectional pedestrian flows. *Journal of Statistical Mechanics: Theory and Experiment* 2017(3):033404.
49. Das SK, Alam Je, Plumari S, Greco V (2020) Transmission of airborne virus through sneezed and coughed droplets. *Physics of Fluids* 32(9):097102.
50. Laverge J, Spilak M, Novoselac A (2014) Experimental assessment of the inhalation zone of standing, sitting and sleeping persons. *Building and environment* 82:258–266.
51. Fennelly KP (2020) Particle sizes of infectious aerosols: implications for infection control. *The Lancet Respiratory Medicine*.
52. Cevik M, Marcus J, Buckee C, Smith T (2020) Sars-cov-2 transmission dynamics should inform policy. *Available at SSRN 3692807*.