Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas

Alberto Aleta[a]🔟, David Martín-Corral[b,c,d]🔟, Michiel A. Bakker[e], Ana Pastore y Piontti[f], Marco Ajelli[f,g]🔟, Maria Litvinova[g]🔟, Matteo Chinazzi[f]🔟, Natalie E. Dean[h], M. Elizabeth Halloran[i,j]🔟, Ira M. Longini, Jr.[h], Alex Pentland[e], Alessandro Vespignani[a,f,1]🔟, Yamir Moreno[a,k,l,1]🔟, and Esteban Moro[b,c,e,1]🔟

Detailed characterization of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission across different settings can help design less disruptive interventions. We used real-time, privacy-enhanced mobility data in the New York City, NY and Seattle, WA metropolitan areas to build a detailed agent-based model of SARS-CoV-2 infection to estimate the where, when, and magnitude of transmission events during the pandemic's first wave. We estimate that only 18% of individuals produce most infections (80%), with about 10% of events that can be considered superspreading events (SSEs). Although mass gatherings present an important risk for SSEs, we estimate that the bulk of transmission occurred in smaller events in settings like workplaces, grocery stores, or food venues. The places most important for transmission change during the pandemic and are different across cities, signaling the large underlying behavioral component underneath them. Our modeling complements case studies and epidemiological data and indicates that real-time tracking of transmission events could help evaluate and define targeted mitigation policies.

COVID-19 | mobility | location | superspreading event

Without effective pharmaceutical interventions, the COVID-19 pandemic triggered the implementation of severe mobility restrictions and social distancing measures worldwide aimed at slowing down the transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). From shelter in place orders to closing restaurants/shops or restricting travel, the rationale of those measures is to reduce the number of social contacts, thus breaking transmission chains. Although individuals may remain highly connected to household members or close contacts, these measures reduce the connections in the general community that allow the virus to move through the network of human contacts. Some venues may attract more individuals from otherwise unconnected social networks or may attract individuals who are more active and thus have greater exposure. Understanding how interventions targeted at particular venues could impact transmission of SARS-CoV-2 can help us devise better nonpharmaceutical interventions (NPIs) that pursue public health objectives while minimizing disruption to the economy, the education system, and other facets of everyday life.

Although it is by now clear that NPIs have helped to mitigate the COVID-19 pandemic (1), most of the evidence is based on measuring the subsequent reduction in the case growth rate or secondary reproductive number. For example, econometric models were used to estimate the effect of the introduction of NPIs on the secondary reproductive number (2, 3). Other studies have shown directly (through correlations or statistical models) (4) or indirectly (through epidemic simulations) (5, 6) the relationship between mobility or individuals' activity and number of cases. Unfortunately, most of the data used so far do not have the granularity required to assess how social contacts and SARS-CoV-2 transmission events are modified by NPIs (7).

This is especially important given the heterogeneous spreading of SARS-CoV-2. Overdispersion in the number of secondary infections produced by a single individual was an important characteristic of the 2003 SARS pandemic (8) and has been similarly observed for SARS-CoV-2 (9). Several drivers of superspreading events (SSEs) have been proposed: biological, due to differences in individuals' infectiousness; behavioral, caused by unusually large gatherings of contacts; and environmental, in places where the surrounding conditions facilitate spread (10). Transmissibility depends critically on the characteristics of the place where contacts happen, with many SSEs documented in crowded, indoor events with poor ventilation. A characteristic of this overdispersion is that most infections (around 80%) are due to a small number of people or places (20%), suggesting that better-targeted NPIs or cluster-based contact tracing strategies can be devised to control the pandemic (11). Although several studies have provided insights on SSEs (7, 12), given their outsized importance for SARS-CoV-2, we need better

Significance

The characterization of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission risks across different settings remains unclear, including the roles of individual and setting heterogeneity. We integrate anonymized time-resolved mobility data with census and demographic data in the New York City, NY and Seattle, WA metropolitan areas to characterize the magnitude and heterogeneity of transmission events during the first COVID-19 wave. We simulate COVID-19 epidemic trajectories to study the impact of interventions, the part played by different settings in the infection spreading, and the role of superspreading events. Our results indicate that places are not dangerous on their own; instead, transmission risk is a combination of both the characteristics of the place/setting and the behavior of individuals who visit it.
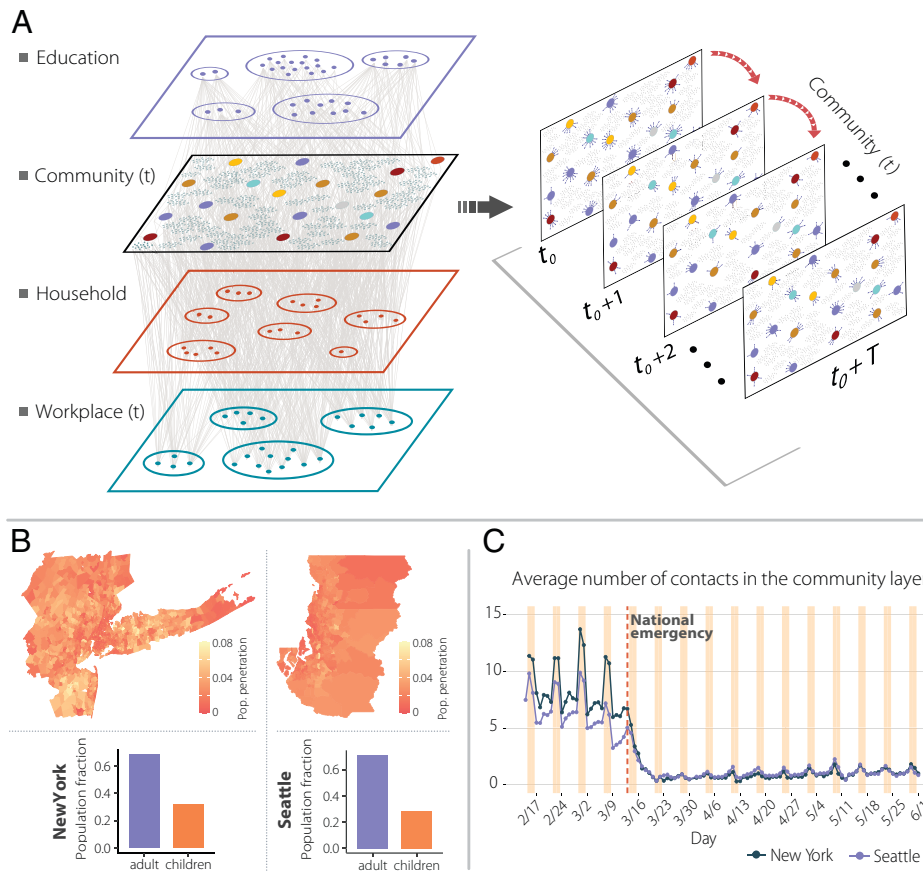
Fig. 1. Network components, New York and Seattle metropolitan areas population and social contacts dynamics at the community layer over time. (*A*) A schematic illustration of the weighted multilayer and temporal network for our synthetic population built from mobility data. There are four different layers; the school and household layers are static over time, and the combined workplace and community layers have a daily temporal component. (*B*) The geographic penetration (fraction of mobile devices by population) from our mobility data compared to the total population for the New York and Seattle metropolitan areas. (*C*) The average daily number of contacts in the community layer for both metropolitan areas.

information about where, when, and to what extent these SSEs happen and how they may be mitigated or amplified by NPIs.

In this paper we use a longitudinal database of detailed mobility and sociodemographic data to estimate the probability of contact and transmission between individuals in different places across the New York City, NY and Seattle, WA metropolitan areas, during the period from 17 February to 1 June 2020 (*SI Appendix*, section 1). Note that the metropolitan areas considered extend beyond the city limits for both locations. We selected these areas because of their large differences in COVID-19 epidemiology, population size, and density. The New York City metro area has a population of 20 million people, while the Seattle metro area has 3.8 million inhabitants. Moreover, the New York City metro area has a higher density (5,438 people per square kilometer, median by census tract) than Seattle (1,576 people per square kilometer). Finally, the number of reported COVID-19 cases/deaths during the study period in the New York City area was very large (223 per 100,000) compared to that in the Seattle area (24 per 100,000). Individual mobility data are sampled to be representative of the different census areas (census block groups) (Fig. 1). Probabilistic estimation of contact between individuals is weighted according to the likelihood of exposure between them in the different places around the metro areas. This defines a weighted temporal network consisting of four layers representing the probabilistic estimation of physical/social interactions occurring in 1) the community, 2) workplaces, 3) households, and 4) schools (Fig. 1). The community and workplace layers are generated

using 4 mo of data observed in the New York City and Seattle metropolitan areas from anonymized users who opted in to provide access to their location data, through a General Data Protection Regulation (GDPR)–compliant framework provided by Cuebiq (*SI Appendix*, section 1).

The data allow us to understand how infection can propagate in each layer by estimating the probability of transmission between individuals in the same setting, including schools, workplaces, households, and multiple locations in the community. Settings associated to the community are obtained from a large database of 375,000 locations in New York City and 70,000 locations in Seattle from the Foursquare public application programming interface (API). By measuring the probability that people interact in the different layers, we construct a probabilistic time-varying contact network of ω_{ijt} between individuals i and j on the same day t in the education, community, work, and household layers. Estimates of transmission in the community layer are done by extracting stays of users to the settings using different time and distance in the setting. Our results are independent of the particular choice of minimal time (5 or 15 min) and maximum distance to the setting (10 or 50 m); see Fig. 1 and *SI Appendix*, sections 1 and 2 for more information about the data and layers. Our model covers all possible interactions in urban areas and not just foot traffic to commercial locations that people visit (7), something especially important given the relevant role of households, schools, or workplaces in the transmission of SARS-CoV-2. It is important to note that the underlying data do not provide a direct
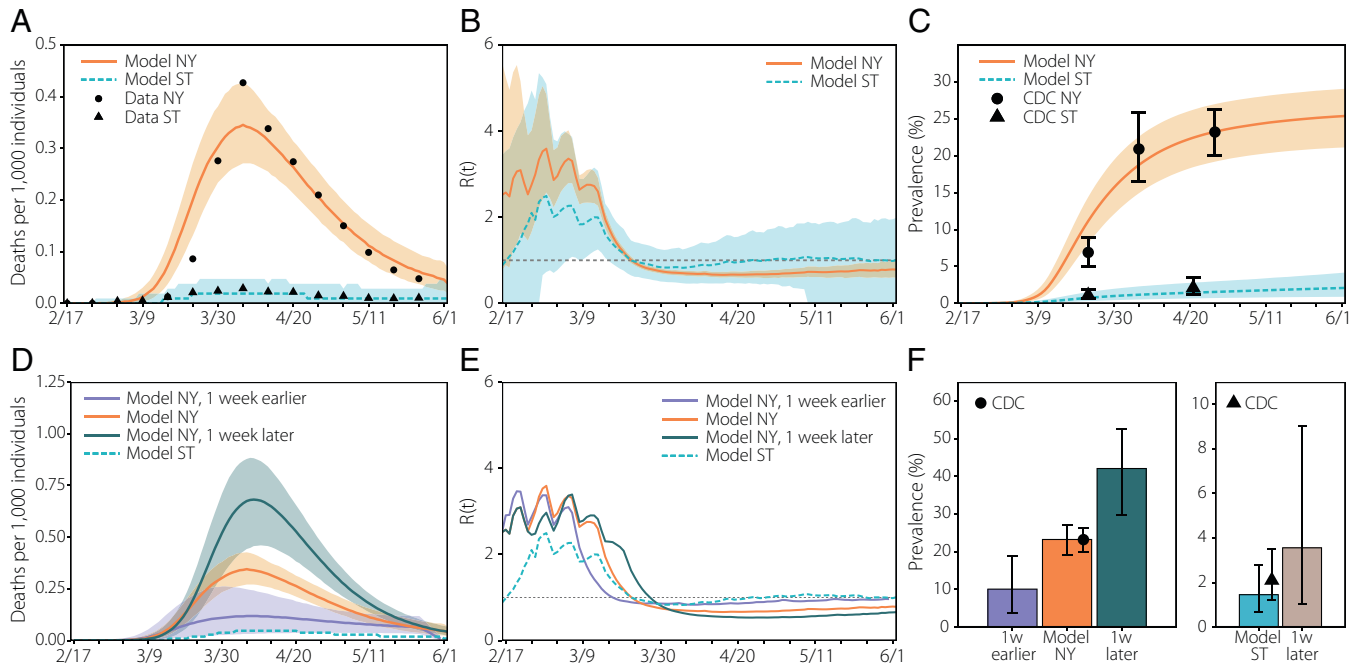
Fig. 2. Evolution of the first wave. (*A*) Weekly number of deaths in New York (NY) and Seattle (ST) metro areas. The dots/triangles represent the reported surveillance data used in the calibration of the models. The lines represent the median of the model ensemble for each location and the shaded areas the 95% CI of the calibrated model (17). (*B*) Evolution of the effective reproduction number according to the output of the simulation. The solid (dashed) line represents the median of the model ensemble and the shaded areas the 95% CI of the model. (*C*) Estimated prevalence in our model (median represented with solid/dashed lines and 95% CI with the shaded area) and values reported by the CDC (dots/triangles represent New York and Seattle data, respectively) (18). (*D*) Estimated number of deaths if the NPIs had been applied in New York 1 wk earlier/later. Solid (dashed) lines represent the median of the model ensemble and the shaded areas the 95% CI. (*E*) Estimated evolution of the effective reproduction number if the measures had been applied in New York 1 wk earlier/later. Solid (dashed) lines represent the median of the model ensemble. (*F*) Estimated prevalence in New York (*Left*) and Seattle (*Right*) if the NPIs had been applied in New York 1 wk earlier/later and in Seattle 1 wk later. The height of the bars represents the median of the model ensemble, while the vertical error bars represent the 95% CI. The dot/triangle shows the value reported by the CDC for the last week of April 2020.

measurement of contacts between individuals and the nature of these contacts (masked/unmasked, with conversation). Rather, our method uses these data to extrapolate the locations visited by each subject and the amount of time the subject spent there, to estimate the transmission probability between individuals, relaxing the homogeneous mixing assumption commonly used in mathematical modeling approaches. In simpler terms, our method does not detect directly colocation of individuals, but rather is a probabilistic estimation of the transmission between them according to the time they spend in the same places or layers.

To model the natural history of the SARS-CoV-2 infection, we implemented a stochastic, discrete-time compartmental model on top of the contact network ω_{ijt} in which individuals transition from one state to the other according to the distributions of key time-to-event intervals (e.g., incubation period, serial interval, etc.) as per available data on SARS-CoV-2 transmission (see *SI Appendix*, section 3 for details). In the infection transmission model, susceptible (S) individuals become infected through contact with any of the infectious categories (infectious symptomatic [IS], infectious asymptomatic [IA], and presymptomatic [PS]), transitioning to the latent (L) compartment, where they are infected but not infectious yet. Latent individuals branch out in two paths according to whether the infection will be symptomatic or not. We also consider that symptomatic individuals experience a presymptomatic phase and that once they develop symptoms, they can experience diverse degrees of illness severity, leading to recovery (R) or death (D). The value of the basic reproduction number is calibrated to the weekly number of deaths (see *SI Appendix*, sections 4, 5, and 7 for further information on the calibration process, on the model's

details, and for the sensitivity of our results toward different values of parameters used in the model).

Results

Impact of NPIs. Our data clearly show that the statistics of potential contacts in the two metro areas have changed due to the introduction of NPIs during the week of 15 March to 22 March (Fig. 1). A National Emergency was declared on 13 March, and the New York City School System announced the closure of schools on 16 March (13). The New York City mayor issued a "shelter in place" order in the city on 17 March (14), and nonessential businesses were ordered to close or suspend all in-person functions in New York, New Jersey, and Connecticut by 22 March. As we can see in Fig. 1 the individuals' total number of contacts decreased dramatically from around seven (in our community layer) to below two. In Seattle, the reduction of contacts started 1 wk earlier than in New York City, coinciding with earlier closing of some schools (15) and the Seattle mayor issuing a proclamation of civil emergency on 3 March (16).

In Fig. 2 we report numerical simulations of the epidemic curve that accurately reproduce the evolution of the incidence of new COVID-19–related deaths in both New York and Seattle metro areas, even though both cities were affected very differently by the epidemic in the first wave. The analysis identifies the impact of the reduction in the estimated number of contacts due to the implemented NPIs: In both the New York and Seattle metro areas, R_t dropped below one 1 wk after NPIs were introduced. To estimate the importance of timely implementations of NPIs in metropolitan areas, we have generated counterfactual scenarios
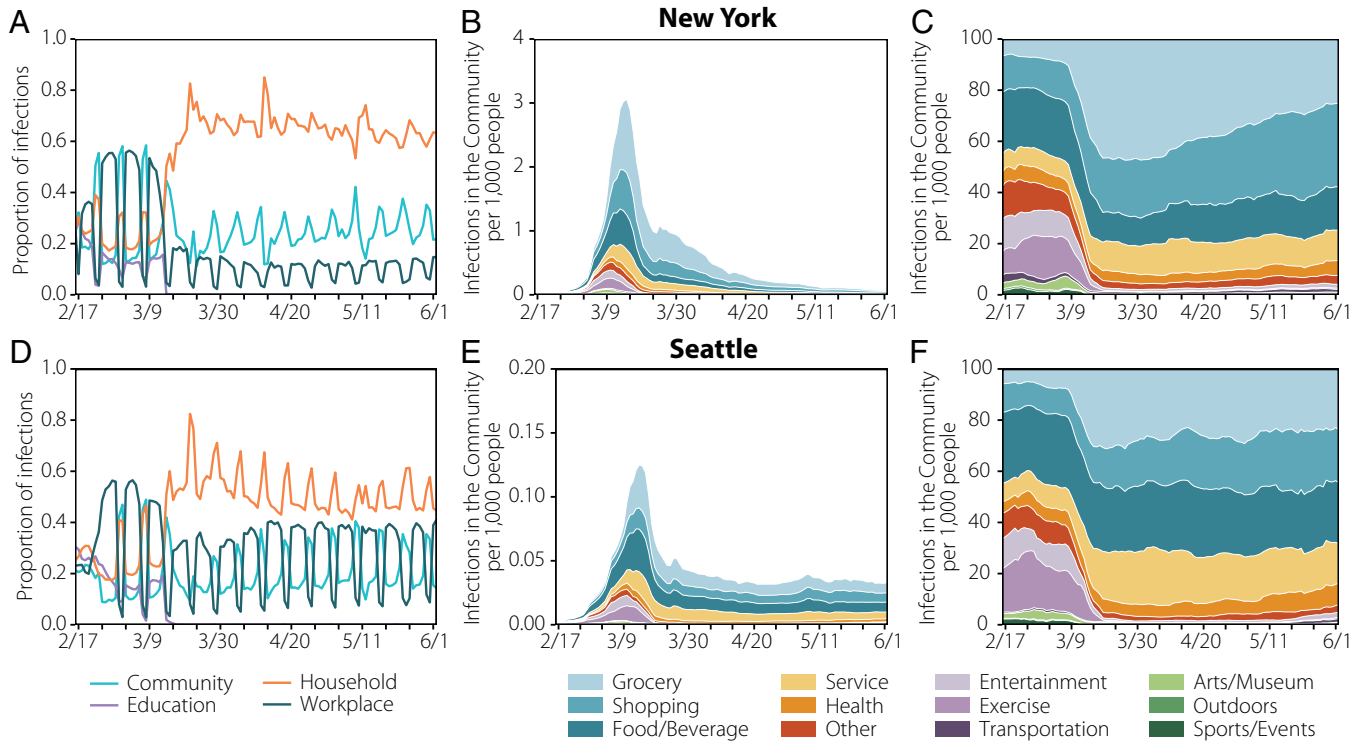
Fig. 3. Spatial spreading of the disease. (*A* and *D*) The share of infections across layers in New York (*A*) and Seattle (*D*). (*B* and *E*) The estimated location where the infections took place for New York (*B*) and Seattle (*E*) in the community layer. Note that the *y* axis is 20 times smaller in Seattle. The evolution has been smoothed using a rolling average of 7 d. (*C* and *F*) The distributions are normalized over the total number of daily infections, showing how infections were shared across categories in the community layer. The evolution has been smoothed using a rolling average of 7 d.

in which the NPIs and the ensuing reduction in the number of contacts could have happened 1 wk earlier or later than the actual timeline (19). The comparison between New York and Seattle is relevant, because we observed that the reduction in contacts in Seattle started to happen exactly 1 wk before that in New York. To this end we have shifted in time the contact patterns around the week where NPIs where introduced in both cities. The results for these scenarios are reported in Fig. 2*D*, where we see that a 1-wk delay in introducing NPIs could have yielded a peak in the number of deaths two times larger than the observed one (0.7 deaths per 1,000 people compared to the 0.35 per 1,000). This doubling in peak deaths following a 1-wk delay is also observed in the Seattle metro area and in the cumulative infection prevalence in the metro area. Conversely, a 1-wk earlier implementation of the NPIs timeline in the New York area could have reduced the death peak by more than a factor of 3, a result similar to that found using county-level simulations (19). In Seattle, implementing the NPIs 1 wk earlier would have prevented the first wave of infections. For this reason, the results are not shown in Fig. 2*F*.

Taxonomy of Transmission Events. The high resolution of our dataset allows us to estimate the relevance of different settings and the effects of NPIs on the transmission dynamic of SARS-CoV-2. People spent different times in each layer and place before and after the introduction of NPIs (*SI Appendix*, section 1). As a result, the number of infections varied significantly during the observed period. As we can see in Fig. 3, before NPIs were introduced, we estimate that most infections took place in the community and workplace layers. Once restrictions were implemented in both cities on 16 March, as expected, the proportion of infections in the household layer greatly increased, especially in the New York area. In Seattle, the numbers of infections in the workplace and household layers were comparable, probably because the number

of cases overall was lower than in New York. We can further stratify data by venue type in the community layer as in Fig. 3, by looking at the estimated top categories (see *SI Appendix*, section 1 for their definition) in terms of the number of total infections throughout the whole period. Before the NPIs were introduced, our model estimates that most of the infections in the community layer happened in food/beverage, shopping, and exercise venues. Also, a significant number of infections happened in art/museums and sport/events venues. After the introduction of NPIs, the number of infections in exercise, sports/events or art/museums venues decreases as expected. However, food, groceries, and shopping venues became the main community setting for transmission in both cities.

Superspreading Events. Our agent-based simulations also allow us to estimate statistically the transmission events by a single individual and estimate how many secondary infections the individual generates. In Fig. 4 we report the distribution of the number of secondary infections produced by each individual in the community layer only. This is driven by individual-level differences in activity and those individuals the individual might interact with. The distribution is highly skewed and can be modeled by a negative binomial distribution with dispersion parameters (k) of 0.16 (New York) and 0.23 (Seattle), in agreement with the evidence accumulated from SARS-CoV-2 transmission data (9, 10, 20, 21). As a result, SSEs are likely to be observed. We define a transmission event as a SSE if the individual infects in a specific location category more than the 99th percentile of a Poisson distribution with average equal to R (see ref. 8 and *SI Appendix*, section 6 for further details), here corresponding to an infected individual infecting eight or more others. Interestingly, if we compare the distribution of secondary infections produced before and after the introduction of NPIs, even though we see a
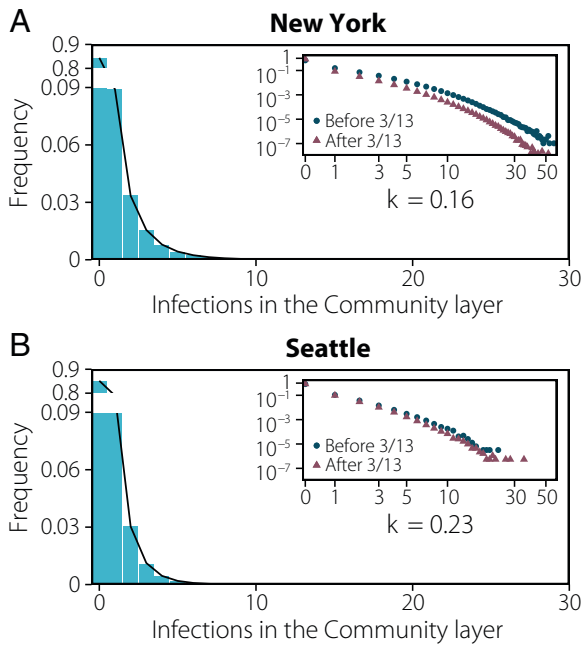
Fig. 4. Behavioral superspreading events. (*A* and *B*) Distribution of the number of infections produced by each individual in New York (*A*) and Seattle (*B*) up to the declaration of National Emergency. The distribution is fitted to a negative binomial distribution yielding a dispersion parameter of $k = 0.163$ [0.159 to 0.168] 95% CI and $k = 0.232$ [0.224 to 0.241] 95% CI, respectively. *Insets* represent the same distribution on the log scale and distinguishing infections that took place before the declaration of National Emergency on 13 March and after that date.

clear reduction of SSEs, we still find a heterogeneous distribution of secondary infections. Thus, the NPIs did not prevent the formation of SSEs, but only significantly lowered their frequency.

Consistent with this pattern of overdispersion in the number of transmission events, we find that the majority of infections are produced by a minority of infected people: ~20% of infected people were responsible for more than ~85% of the infections in both metro areas (*SI Appendix*, Fig. S9). However, note that a critical driver here of this phenomenon is that a large majority of infected people (85% in the community layer) do not infect any others in our simulations. Only a small fraction of infection events (0.08%) are made of eight (or more) secondary infections.

Transmission events and SSEs did not happen equally in different settings or along time or geography. In Fig. 5 we show the results of our simulations for the total number of infections produced in each category and the share of those infections that can be related to SSEs (*SI Appendix*, Table S2). The combination of those two features defines a continuous-risk map in which places can be at different types of risk: 1) low contribution from SSEs and low contribution to the overall infections, such as outdoor places; 2) larger contribution from SSEs but low contribution to the overall infections, such as sports/events, arts/museums or entertainment before the introduction of NPIs; 3) large contribution to the overall infections but with low contribution from SSEs, such as shopping or food/beverage venues after the introduction of NPIs; and 4) large number of infections and with large contribution from SSEs, such as groceries. This classification has important implications from a public health perspective. For instance, venues in risk 2 do not have a major contribution to the overall infections but might represent a challenge for contact tracing. Conversely, for categories in risk 3 it might be easier to trace chains of transmission but their total contribution is large. Note that this definition is not static, but changes over time due to the NPIs imposed by authorities. Indeed, looking at the weekly pattern of infections

(Fig. 5), we observe how some categories move to a different quadrant due to the behavior of individuals. Although we estimate that SSEs and infections were more likely in arts/museums and sports/events in New York and entertainment and grocery in both cities, our simulations show that the grocery category still greatly contributes to the total number of infections, but does not have as many SSEs after 16 March. On the other hand, we estimate that SSEs were rare before 9 March in Seattle, but their contribution doubled in the week of 9 to 15 March—when many individuals probably went for supplies amid preparation for the future introduction of NPIs. This observation includes implicitly a very important message: A place may not be inherently dangerous; rather, the risk is a combination of both the characteristics of the place/setting and the behavior of individuals who visit it. This suggests revisiting studies that find that settings could play always the same role in the evolution of the pandemic (7).

Discussion

Our results emphasize the intertwined nature of human behavior, NPIs, and the evolution of the COVID-19 pandemic in two major metropolitan areas. Specifically, our results suggest that heterogeneous connectivity and behavioral patterns among individuals lead naturally to differences in risk across settings and the generation of SSEs. In particular, the implemented partial or full closures of different settings (e.g., sport venues, museums, workplaces) had a dramatic effect in shaping the mixing patterns of the individuals outside the household (22, 23). As a consequence, the settings responsible for the majority of transmission events and SSEs varied over time. In absolute terms, the food and beverage setting is estimated to have played a key role in determining the number of both transmission events and SSEs in the early epidemic phase; however, this setting was among the first targets of interventions and thus its contribution became zero over time because of the introduced NPIs. On the other hand, settings such as grocery stores, which consistently provided a low absolute contribution to the overall transmission and SSEs, became, in relative terms, a source of SSEs during the lockdown when most other activities were simply not available. These findings suggest that there is room for optimizing targeted measures such as extending working time to dilute the number of contacts or the use of smart working aimed at reducing the chance of SSEs. That could be especially relevant to avoid local flareups of cases when the reproduction number is slightly above or below the epidemic threshold.

Although the overall picture emerging from studying Seattle and New York is consistent, it is important to stress that each urban area might have specific peculiarities due to local transportation, tourism, or other economic drivers differentiating the cities' life cycle. Our results suggest that a one-size-fits-all solution to minimize the spread of SARS-CoV-2 might have very different impact across cities. Furthermore, the results presented may not be generalized to rural areas. Although large parts of the Seattle metro area could be considered as rural, individual connectivity patterns may be differently constrained by the generally lower population density in some other parts of the country.

We note that less complex homogeneous-mixing models can be enough to reproduce aggregated features of the spread of SARS-CoV-2 in different cities (Fig. 2 and *SI Appendix*, section 7.10), and detailed (although still homogeneous-mixing) aggregate visitation patterns to places can be used to evaluate the average role of places in the spreading (7). However, the model proposed here incorporates both individual mobility behavior and the detailed description of home, school, and workplace multilayer temporal networks, thus allowing us to simultaneously capture key aspects
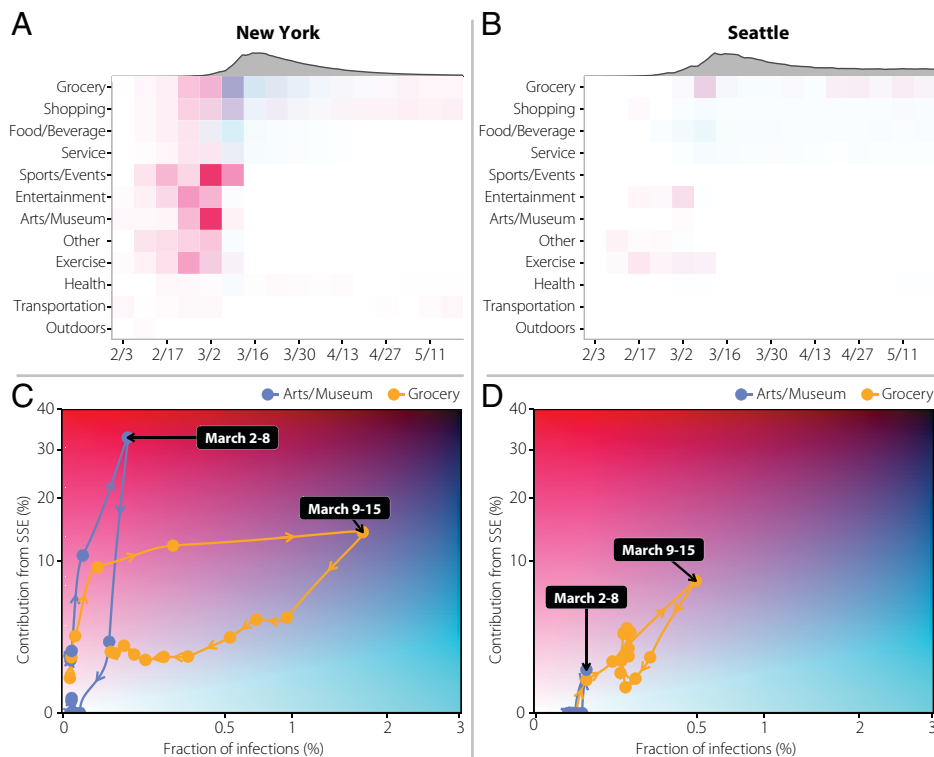
Fig. 5. Dynamics of SSEs. Risk evolves with time as a function of the behavior of the population and policies in place. (*A* and *B*) Risk posed by each category per week, defined using the corresponding map below. As a reference, the gray area on top shows the estimated weekly incidence. (*C* and *D*) The *x* axis represents the fraction of total infections that are associated with each category, while the *y* axis accounts for the share of those infections that can be attributed to SSEs in each category. Note that the fraction of infections is normalized over all the infections produced in all the social settings throughout the whole period. This defines a continuous-risk map in which places with few infections and low contribution from SSEs will be situated on the bottom left corner. Places where the number of infections is high but the contribution from SSEs is low are situated in the bottom right corner. Conversely, places with large contribution from SSEs but a low amount of infections are situated in the top left corner. Finally, places with both a large number of infections and an important contribution from SSEs are situated in the top right corner. The color associated to each tile in *A* and *B* is extracted from the position of the point in the plane defined in *C* and *D*. The points in *C* and *D* show the evolution of the position of the categories arts/museum and grocery for each week, with the arrows indicating the time evolution.

of COVID-19, such as contagion overdispersion (superspreading events, Fig. 4), the temporal evolution of the risk of infection by social setting (Fig. 5), or the impact of school closures or stay-at-home policies (Fig. 3). By having a better description of mobility patterns at the individual level, our methodology relies only on a minimal set of parameters, making it more generalizable to other locations of epidemic context than models that encode that behavior by fitting transmissibility parameters for places, residences, cities, or even temporal periods (7).

Our modeling analysis does not have the ambition to substitute field investigations, which remain the primary source of evidence. Some of the reported findings (e.g., the role of food and beverage venues or groceries) appear to be in agreement with epidemiological investigations (7, 24–27). Future empirical analyses could provide further validation of our findings. Our modeling investigation is based on real-time data on human mobility/activity that provide an indirect proxy for infection transmission. One of the strengths of this approach is that, different from epidemiological investigations, the data can be retrieved in real time and longitudinally, thus allowing us to quickly capture possible changes in the most relevant settings for transmission. Furthermore, our approach could help minimize the noisy and biased data collection related to massive transmission events (28). Yet, the approach used here is far from capturing all the finest details of human social contacts and thus the estimates on the contribution of different settings to SARS-CoV-2 transmission entail an unavoidable uncertainty.

To properly interpret our results, it is important to acknowledge the limitations of the assumptions included in our modeling

exercise. First, we have considered a decrease of the transmission probability in outdoor compared to indoor settings of January 2020 (29). Although this choice is guided by empirical evidence and our results are robust to this choice (*SI Appendix*, section 7), further studies better quantifying the relative risk of indoor vs. outdoor transmission are warranted. Second, our model neglects to consider differences in the behavior that people follow when in contact with each other. It is indeed possible that contacts between relatives and friends have a larger chance of resulting in a transmission event compared with interactions with strangers (30). Third, we do not model nursing homes, which were severely hit by the COVID-19 pandemic across the globe. However, although they represent a key setting to determine COVID-19 burden in terms of deaths and patients admitted to hospitals and intensive care units, they are possibly not central to capture the transmission dynamics of SARS-CoV-2 at the population level, which is the aim of this study. Although there is some location information from hospitals, we do not model them. Nonetheless, contact tracing studies from several countries have revealed that transmission within hospitals is relatively low, and hospital staff are more at risk from interactions with their coworkers (e.g., in the breakroom) or out in their communities (31, 32).

In conclusion, the majority of NPIs introduced in large urban areas in March 2020 were effective in dramatically slowing down the first wave of COVID-19 by greatly reducing the number of effective contacts in the population. Closing down schools, businesses, workplaces, and social venues, however, took (and still does take) an enormous toll on our economy and society. Our results and methodology allow for a real-time data-driven analysis

that connects NPIs, human behavior, and the transmission dynamic of SARS-CoV-2 to provide quantitative information that can aid in defining more targeted and less disruptive interventions not only at a local level, but also to assess whether local restrictions could trigger undesired effects at nearby locations not subject to the same limitations. Although nowadays the epidemiological landscape has dramatically changed by the introduction of vaccines, the spread of more transmissible variants, and the buildup of natural immunity, the results offered in this paper provide unique insights on the transmission pathways of SARS-CoV-2 and can be instrumental for the definition of location-based mitigation policies and for making informed decisions about high-risk activities.

Materials and Methods

We used individual-level mobility data of over 0.5 million individuals distributed in the New York and Seattle metropolitan areas during the months of February 2020 to June 2020 to estimate the day and type of venues where people might have interactions that yield transmission events. To do that we extracted from the mobility data the stays (stops) of people in a large collection of around 440,000 settings (33). With this information we built two synthetic populations, one for each metropolitan area, in which agents can interact in different settings: workplaces, households, schools, and the community (points of interest). We then explore the transmission of SARS-CoV-2 using a compartmental and stochastic epidemic model applied on top of this population.

The behavioral changes induced in the population by the introduction of several NPIs are naturally encoded in the mobility data, allowing us to characterize the effect of these interventions. We ran counterfactual simulations of our stochastic epidemic model to understand that effect. Furthermore, the resolution of these data allows us to characterize the spreading through different types of venues at different stages of the epidemic, depicting a complex picture in which the combination of both the characteristics of the place/setting and the behavior of individuals who visit it determine its risk.

Finally, the information about the statistical heterogeneity of the contact pattern of different individuals allows us to study the frequency and characteristics of behavior-related SSEs. We study the likelihood of finding a SSE per setting as a function of time by looking at the number of infections produced by each individual in each location. A full description of the materials and methods is provided in *SI Appendix*.

Author affiliations: [a]ISI Foundation, 10126 Turin, Italy; [b]Departamento de Matemáticas, Universidad Carlos III de Madrid, 28911 Leganés, Spain; [c]Grupo Interdisciplinar de Sistemas Complejos, Universidad Carlos III de Madrid, 28911 Leganés, Spain; [d]Zensei Technologies S.L., 28010 Madrid, Spain; [e]Connection Science, Institute for Data Science and Society, Massachusetts Institute of Technology, Cambridge, MA 02139; [f]Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA 02115; [g]Laboratory for Computational Epidemiology and Public Health, Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, IN 47405; [h]Department of Biostatistics, College of Public Health and Health Professions, University of Florida, Gainesville, FL 32611; [i]Biostatistics, Bioinformatics, and Epidemiology Program, Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; [j]Department of Biostatistics, University of Washington, Seattle, WA 98195; [k]Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, 50018 Zaragoza, Spain; and [l]Department of Theoretical Physics, Faculty of Sciences, University of Zaragoza, 50009 Zaragoza, Spain

Author contributions: A.A., D.M.-C., M.A., A.V., Y.M., and E.M. designed research; A.A., D.M.-C., and M.A.B. performed research; A.A., D.M.-C., M.A., A.V., Y.M., and E.M. analyzed data; and A.A., D.M.-C., M.A.B., A.P.y.P., M.A., M.L., M.C., N.E.D., M.E.H., I.M.L., A.P., A.V., Y.M., and E.M. wrote the paper.

1. M. U. G. Kraemer et al.; Open COVID-19 Data Working Group, The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
2. H. Badr et al., Social distancing is effective at mitigating COVID-19 transmission in the United States. *medRxiv* [Preprint] (2020). https://doi.org/10.1101/2020.05.07.20092353. Accessed 17 December 2020.
3. J. Y. Wu et al., Changes in reproductive rate of SARS-CoV-2 due to non-pharmaceutical interventions in 1,417 U.S. counties. *medRxiv* [Preprint] (2020). https://doi.org/10.1101/2020.05.31.20118687. Accessed 17 December 2020.
4. P. Cintia et al., The relationship between human mobility and viral transmissibility during the COVID-19 epidemics in Italy. *arXiv* [Preprint] (2020). https://doi.org/10.48550/arXiv.2006.03141. Accessed 17 December 2020.
5. J. Dehning et al., Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* **369**, eabb9789 (2020).
6. A. Aleta, Y. Moreno, Evaluation of the potential incidence of COVID-19 and effectiveness of containment measures in Spain: A data-driven approach. *BMC Med.* **18**, 157 (2020).
7. S. Chang et al., Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).
8. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
9. D. C. Adam et al., Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat. Med.* **26**, 1714–1719 (2020).
10. B. M. Althouse et al., Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control. *PLoS. Biol.* **18**, (2020).
11. A. Chande et al., Real-time, interactive website for US-county-level COVID-19 event risk assessment. *Nat. Hum. Behav.* **4**, 1313–1319 (2020).
12. R. Laxminarayan et al., Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science* **370**, 691–697 (2020).
13. E. Shapiro, New York City public schools to close to slow spread of coronavirus. *The New York Times*, 2020. https://www.nytimes.com/2020/03/15/nyregion/nyc-schools-closed.html. Accessed 3 December 2020.
14. A. Lardieri, New York City Mayor de Blasio considering shelter in place. *U.S. News & World Report*, 2020. https://www.usnews.com/news/health-news/articles/2020-03-17/new-york-city-mayor-bill-de-blasio-considering-shelter-in-place. Accessed 3 December 2020.
15. J. Calfas, T. D. Hobbs, Schools shut in Seattle area as coronavirus spreads. *The Wall Street Journal*, 2020. https://www.wsj.com/articles/coronavirus-spreads-world-wide-containment-is-an-unlikely-outcome-11583403706. Accessed 3 December 2020.
16. J. A. Durkan, Mayoral proclamation of civil emergency. City of Seattle. 2020. https://durkan.seattle.gov/wp-content/uploads/sites/9/2020/03/COVID-19-Mayoral-Proclamation-of-Civil-Emergency.pdf. Accessed 3 December 2020.
17. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
18. Commercial laboratory seroprevalence survey data. *CDC*, 2020. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/commercial-lab-surveys.html. Accessed 11 September 2020.
19. S. Pei, S. Kandula, J. Shaman, Differential effects of intervention timing on COVID-19 spread in the United States. *Sci. Adv.* **6**, eabd6370 (2020).
20. A. Endo, S. Abbott, A. Kucharski, S. Funk, Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res.* **5**, 67 (2020).
21. K. Sun et al., Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science* **371**, 6526 (2021).
22. J. Zhang et al., Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368**, 1481–1486 (2020).
23. C. I. Jarvis et al.; CMMID COVID-19 working group, Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Med.* **18**, 124 (2020).
24. J. Lu et al., COVID-19 outbreak associated with air conditioning in restaurant, Guangzhou, China, 2020. *Emerg. Infect. Dis.* **26**, 1628–1631 (2020).
25. K. A. Fisher et al.; IVY Network Investigators; CDC COVID-19 Response Team, Community and close contact exposures associated with COVID-19 among symptomatic adults ≥18 years in 11 outpatient

health care facilities - United States, July 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 1258–1264 (2020).

26. F. Y. Lan, C. Suharlim, S. N. Kales, J. Yang, Association between SARS-CoV-2 infection, exposure risk and mental health among a cohort of essential retail workers in the USA. *Occup. Environ. Med.* **78**, 237–243 (2020).

27. R. A. Shumsky, L. Debo, R. M. Lebeaux, Q. P. Nguyen, A. G. Hoen, Retail store customer flow and COVID-19 transmission. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2019225118 (2021).

28. Z. Susswein, S. Bansal, Characterizing superspreading of SARS-CoV-2: From mechanism to measurement. *medRxiv* [Preprint] (2020). https://doi.org/10.1101/2020.12.08.20246082. Accessed 17 December 2020.

29. M. Weed, A. Foad, Rapid scoping review of evidence of outdoor transmission of COVID-19. *medRxiv* [Preprint] (2020). https://doi.org/10.1101/2020.09.04.20188417. Accessed 17 December 2020.

30. S. Hu *et al.*, Infectivity, susceptibility, and risk factors associated with SARS-CoV-2 transmission under intensive contact tracing in Hunan, China. *Nat. Commun.* **12**, 1533 (2021).

31. C. Rhee *et al.*; CDC Prevention Epicenters Program, Incidence of nosocomial COVID-19 in patients hospitalized at a large US academic medical center. *JAMA Netw. Open* **3**, e2020498 (2020).

32. A. Richterman, E. A. Meyerowitz, M. Cevik, Hospital-acquired SARS-CoV-2 infection: Lessons for public health. *JAMA* **324**, 2155–2156 (2020).

33. E. Moro, D. Calacci, X. Dong, A. Pentland, Mobility patterns are associated with experienced income segregation in large US cities. *Nat. Commun.* **12**, 4633 (2021).