



Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19

Alberto Aleta¹, David Martín-Corral^{2,3}, Ana Pastore y Piontti⁴, Marco Ajelli^{5,6}, Maria Litvinova¹, Matteo Chinazzi⁴, Natalie E. Dean⁷, M. Elizabeth Halloran^{8,9}, Ira M. Longini Jr⁷, Stefano Merler⁵, Alex Pentland¹⁰, Alessandro Vespignani^{1,4}✉, Esteban Moro^{1,2,10}✉ and Yamir Moreno^{1,11,12}✉

While severe social-distancing measures have proven effective in slowing the coronavirus disease 2019 (COVID-19) pandemic, second-wave scenarios are likely to emerge as restrictions are lifted. Here we integrate anonymized, geolocalized mobility data with census and demographic data to build a detailed agent-based model of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission in the Boston metropolitan area. We find that a period of strict social distancing followed by a robust level of testing, contact-tracing and household quarantine could keep the disease within the capacity of the healthcare system while enabling the reopening of economic activities. Our results show that a response system based on enhanced testing and contact tracing can have a major role in relaxing social-distancing interventions in the absence of herd immunity against SARS-CoV-2.

The first report of a new infectious disease, later called COVID-19, appeared on 31 December 2019¹. As of 15 July 2020, the disease has spread to 188 countries, with more than 13.3 million confirmed cases, and has killed more than 579,500 people². As the number of confirmed COVID-19 cases increased and the expansion of the disease entered into a global exponential growth phase, a large number of affected countries were forced to adopt non-pharmaceutical interventions at an unprecedented scale. Given the absence of specific antiviral prophylaxis, therapeutics or a vaccine, non-pharmaceutical interventions ranging from case isolation and quarantine of contacts, to the lockdown of entire populations have been implemented with the aim of suppressing and mitigating the epidemic before it could overwhelm the healthcare system. Although these aggressive measures appear to be successful in reducing the number of deaths and hospitalizations^{3,4}, and in reducing the transmission of the SARS-CoV-2 virus, the absence of herd immunity after the first wave of the epidemic points to a high risk of resurgence when interventions are relaxed and societies return to a ‘business as usual’ lifestyle^{5–7}. It is therefore of paramount importance to analyse different mitigation and containment strategies aimed at minimizing the risk of potential additional waves of the COVID-19 epidemic, while providing an acceptable trade-off between economic and public health objectives.

In this study, through the integration of anonymized and privacy-enhanced data from mobile devices and census data, we build a detailed sample of the synthetic population of the Boston

metropolitan area (BMA) in the United States (see Fig. 1a,b). This synthetic population (Fig. 1a) is used to define a data-driven agent-based model of SARS-CoV-2 transmission and to provide a quantitative analysis of the evolution of the epidemic and the effectiveness of social-distancing interventions. The model allows us to explore strategies concerning the lifting of social-distancing interventions in conjunction with testing and isolation of cases and tracing and quarantine of exposed contacts. Our results indicate that, after the abatement of the epidemic through the ‘stay-at-home’ orders and halt to all non-essential activities, a proactive policy of testing, contact tracing and household quarantine of contacts enables the gradual reopening of economic activities and workplaces, with a low COVID-19 incidence in the population and a manageable impact on the healthcare system.

To provide a quantitative estimate of the contact patterns for the population of agents and to build the synthetic population of the BMA, we used detailed mobility and socio-demographic data and generated a network that encodes the contact patterns of around 85,000 agents in the area during a period of six months (see Supplementary Information). Agents are chosen to be representative of the different census areas in the Boston area following the methodology used in ref. ⁸. This defines a weighted multilayer network consisting of three layers representing the network of social interactions at (1) workplace and community level, (2) households, and (3) schools, as shown in Fig. 1a. Connections between two agents in the workplace and community layer are estimated from the data by the

¹Institute for Scientific Interchange Foundation, Turin, Italy. ²Department of Mathematics and GISC, Universidad Carlos III de Madrid, Leganés, Spain.

³Zensei Technologies S.L., Madrid, Spain. ⁴Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA, USA. ⁵Bruno Kessler Foundation, Trento, Italy. ⁶Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, IN, USA. ⁷Department of Biostatistics, College of Public Health and Health Professions, University of Florida, Gainesville, FL, USA. ⁸Fred Hutchinson Cancer Research Center, Seattle, WA, USA. ⁹Department of Biostatistics, University of Washington, Seattle, WA, USA. ¹⁰Connection Science, Institute for Data Science and Society, Massachusetts Institute of Technology, Cambridge, MA, USA. ¹¹Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza, Spain. ¹²Department of Theoretical Physics, Faculty of Sciences, University of Zaragoza, Zaragoza, Spain.

✉e-mail: alexves@gmail.com; esteban.moroegido@gmail.com; yamir.moreno@gmail.com

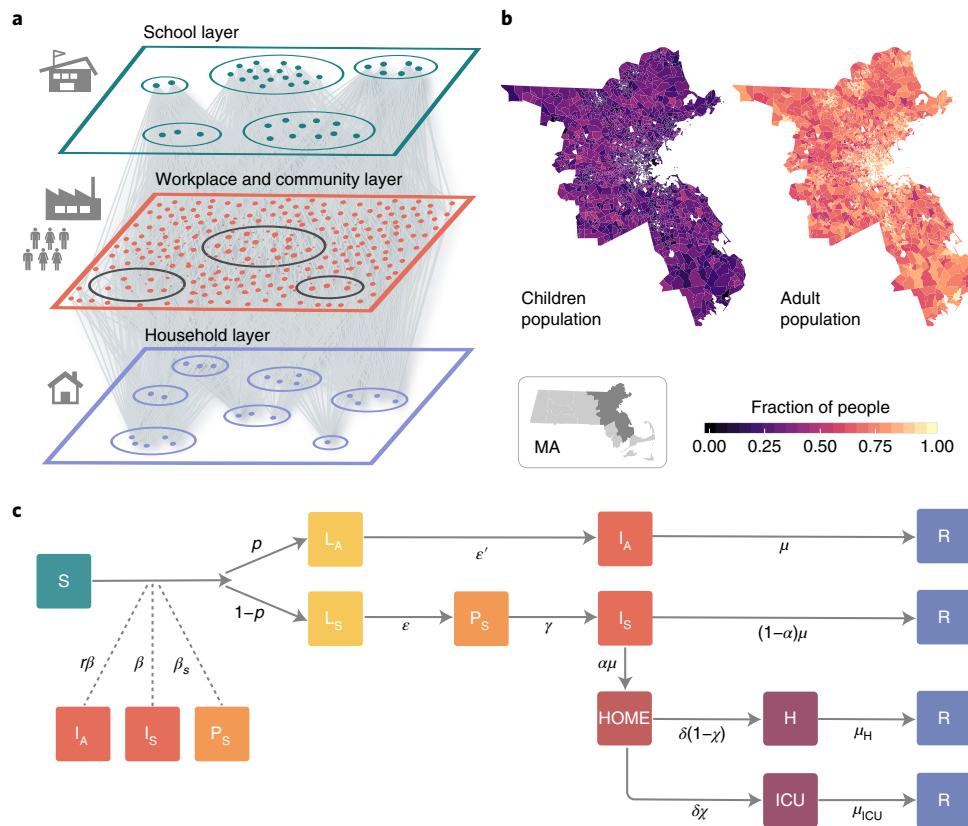


Fig. 1 | Model components. **a**, Schematic illustration of the weighted multilayer synthetic population built from mobility data in the metropolitan area of Boston. The agent-based system comprises around 64,000 adults and 21,000 children. **b**, Geographical distribution of the agents in MA, Massachusetts. Nodes in **a** are connected by more than five million weighted edges. Community layers (which include workplaces), are further classified into categories according to Foursquare’s taxonomy of places. **c**, The compartmental model used to describe the natural history of the disease and the transition rates (represented alongside the connecting arrows, and described in Methods, “Stochastic simulations of COVID-19 dynamics”) between the different states. Specifically, we consider susceptible (S), latent asymptomatic (L_A), latent symptomatic (L_S), presymptomatic (P_S), infectious asymptomatic (I_A), infectious symptomatic (I_S), hospitalized (H), hospitalized in intensive care (ICU) and recovered (R) individuals. More details of the model and the transitions between compartments are provided in Methods and the Supplementary Information.

probability of both being present in a specific place (for example, a restaurant, workplace or shop) weighted according to the time they have spent in the same place. A second layer represents the households of each anonymous individual. Using the home census block group of each anonymous user, we associate each individual with a specific household profile on the basis of socio-demographic data at US census block group level⁹. Families are generated by randomly mixing nodes from the community living in the same census block group, following the statistical features of family types and sizes. Finally, a third layer represents the contacts in schools (that is, every node represents one synthetic student and has contacts only with other individuals attending the same school). To study the evolving dynamics of the infection, we implemented a stochastic, discrete-time compartmental model in which individuals transition from one state to the other according to key time-to-event intervals (for example, incubation period, serial interval and time from symptom onset to hospital admission), as from available data on SARS-CoV-2 transmission. The natural history of the disease is captured by the epidemiological model represented in Fig. 1c, where we also show the transition rates among compartments^{8,10–12}. The model considers that susceptible individuals (S) become infected through contact with any of the infectious categories (infectious symptomatic (I_S), infectious asymptomatic (I_A) and pre-symptomatic (P_S)), transitioning to latent compartments (L_S and L_A), where they are infected but not yet infectious. 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 pre-symptomatic phase and that once they develop symptoms, they can experience diverse degrees of illness severity, from mild symptoms to hospitalization (H) or in need of an intensive care unit (ICU)¹³. Finally, individuals transition in the removed compartment (identifying recovered or dead individuals). The model assumes a basic reproductive number, $R_0 = 2.5$, which together with the other parameters (Supplementary Table 1) yields a generation time $T_g = 6.6$ d. We consider a 25% fraction of asymptomatic individuals. We report the full set of parameters used in the model and an extensive sensitivity analysis in the Supplementary Information. The model is not calibrated to account for the specific evolution of the COVID-19 epidemic in Boston, as it is aimed at showing the effect of different non-pharmaceutical interventions rather than providing a forensic analysis of the outbreak in the BMA. Furthermore, current data on the detailed reopening policies in the BMA during the summer are still under evaluation pending the future evaluation of the epidemic trajectory. Details on the generation of the synthetic population network and the infection-transmission model are provided in the Supplementary Information.

Results

To provide a baseline of the impact of COVID-19 in the BMA, we have first investigated an unmitigated scenario in which no

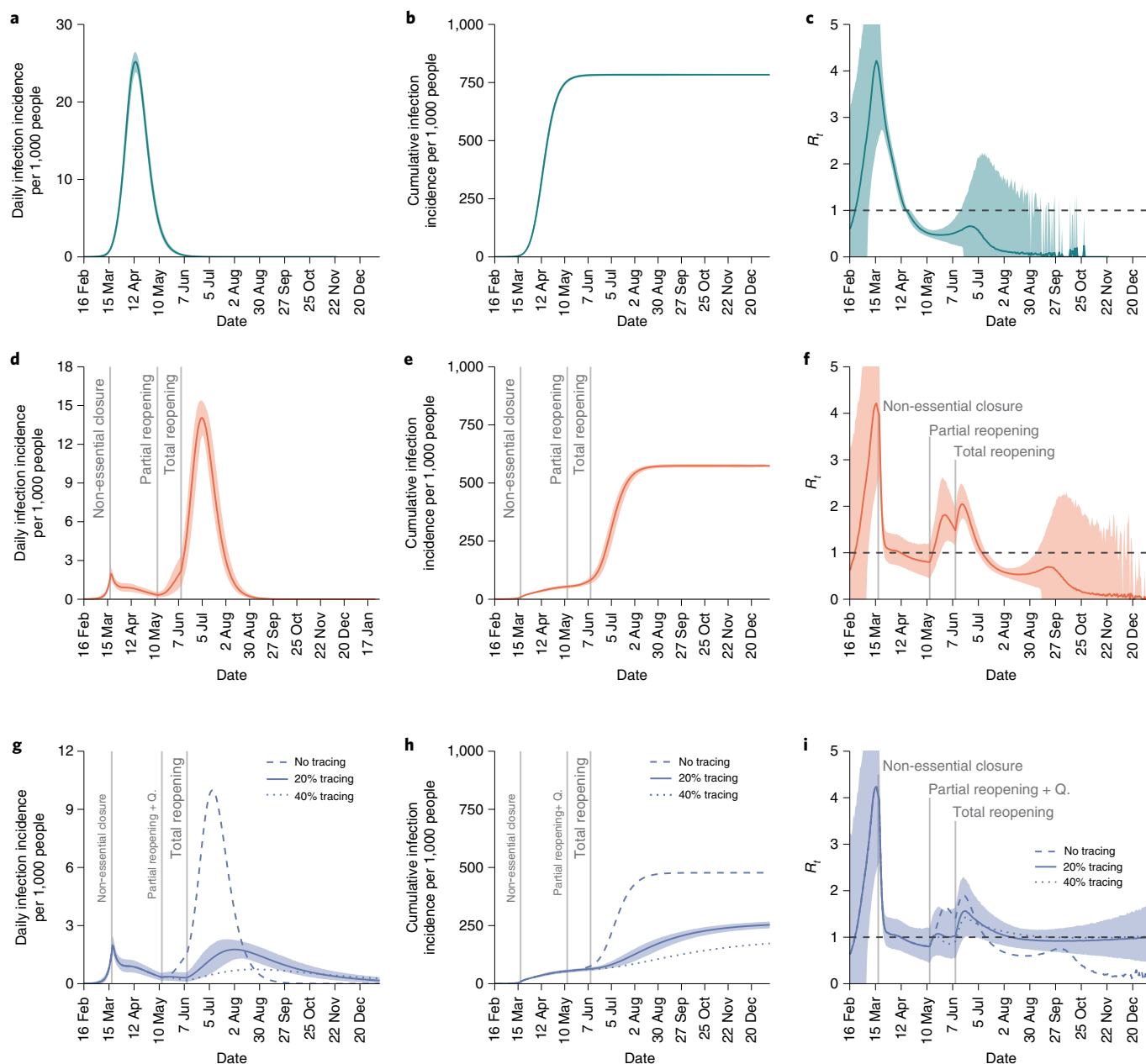


Fig. 2 | Impact of COVID-19 under different scenarios. **a–i**, Evolution of the number of new cases (**a,d,g**), the outbreak size (**b,e,h**) and the effective reproductive number (**c,f,i**) as a function of time in each situation studied. Results for the SARS-CoV-2 transmission dynamics are shown for the unmitigated scenario (**a–c**) and the two social-distancing interventions considered: LIFT (**d–f**) and LET (**g–i**) scenarios. In both LIFT and LET scenarios, we considered the closure of schools and non-essential places for eight weeks. This is the strictest lockdown period, which is followed by a partial lifting of the stay-at-home policy, the duration of which is set to four weeks. During the partial lifting, all places in the community layer are open except mass-gathering locations (such as restaurants, theatres, and others; see Supplementary Information). Finally, a full reopening takes place after the period of partial lifting ends (relevant events are marked with vertical lines). In **d–f**, no other measures are adopted concurrently to the lifting of the restrictions, whereas in **g–i**, the reopening is accompanied by an active policy consisting of testing the symptomatic individuals, home isolating them and quarantining (Q.) their household and the households of a fraction of their contacts, as indicated. Note that the vertical scales of graphs in **a,d,g** are not the same and that both the number of new cases and total cases are per 1,000 inhabitants. x-axes show dates in 2020. Feb, February; Mar, March; Apr, April; Jun, June; Jul, July; Aug, August; Sep, September; Oct, October; Nov, November; Dec, December. In all panels, the solid line represents the average over 10,000 simulations and the shaded region represents the 95% CI.

interventions are implemented. Results for the unmitigated scenario are shown in Fig. 2a–c. An unmitigated COVID-19 would have a peak of daily incidence of 25.2 (95% confidence interval (CI): 23.8–26.4) newly infected individuals per 1,000 people. The epidemic follows a typical trajectory; namely, when the effective reproduction number R_t as a function of time (Fig. 2c) becomes smaller

than 1, the transmission dynamics slow down and eventually vanish after having infected about 75% of the population (Fig. 2b). Figure 3a shows the evolution of the estimated number of new severely affected individuals who require hospitalization and admission into ICUs. At the peak of the unmitigated epidemic, the number of ICU beds needed exceeds the available capacity (dashed horizontal line

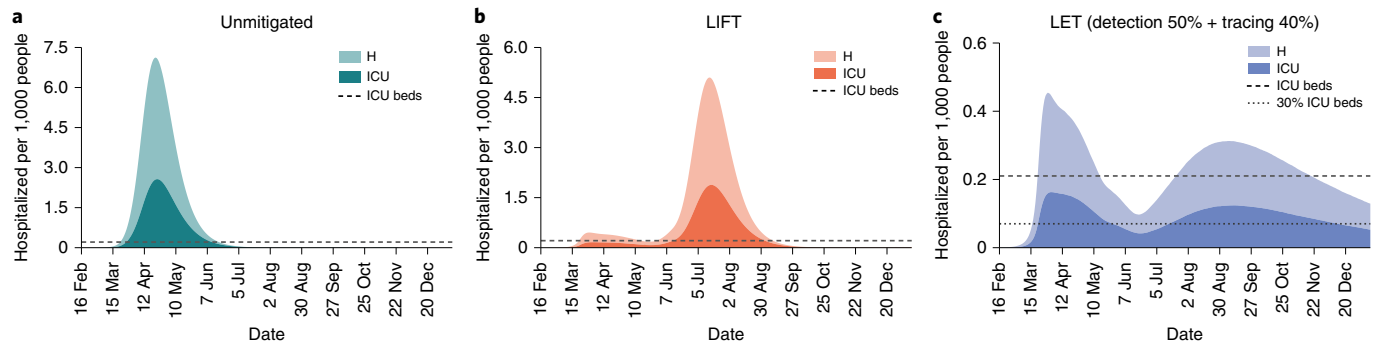


Fig. 3 | Impact on the Boston healthcare system. **a–c**, Estimated number of individuals per 1,000 inhabitants who would need hospitalization (H) or intensive care (ICU) for the unmitigated situation **(a)**, the LIFT scenario **(b)** and the LET scenario **(c)**. The horizontal dashed lines represent the ICU basal capacity of the Boston healthcare system; the dotted line in **c** indicates 30% of the ICU basal capacity. The x-axes show dates in 2020.

in Fig. 3a) by a factor of more than 10, indicating that the healthcare system would suffer large service disruptions, resulting in additional deaths due to hospitals overcrowded with patients with COVID-19 (https://www.ahd.com/data_sources.html). It is worth noting that current estimated fatality rates consider the general availability of ICU beds and critical care capacity. If this is not possible, the fatality rate may increase substantially. We do not report fatality estimates here, as this goes beyond the scope of our analysis and should consider specific data for the BMA, as well as changing medical treatment and therapeutics in future months.

To avoid the harmful effects of an unmitigated COVID-19 epidemic, governments and policy makers across the world have relied on the introduction of aggressive social-distancing measures. In the United States, as of April 15 2020, it was estimated that more than 95% of the population was under a stay-at-home or ‘shelter-in-place’ order^{14,15}. To model the social-distancing policies implemented in the whole BMA, we have considered March 17 2020 as the average starting date of social-distancing policies that include school closures, the shut down of all non-essential work activities as well as mobility restrictions (details in Supplementary Information). This scenario mimics the social-distancing intervention implemented in most of the high income countries, in Europe and across states in the United States. Such extreme social-distancing policies come with very large economic costs and social disruption effects¹⁶, prompting the question of what exit strategy can be devised to restart economic activities and normal societal functions¹⁷. For this reason, we explore two different scenarios for lifting social-distancing interventions:

- Lift scenario (LIFT): the stay-at-home order is lifted after eight weeks by reopening all work and community places, except for mass-gathering locations such as restaurants, theatres and similar locations (see Supplementary Information). The latter partial reopening is enforced for another four weeks, which is followed by a full lifting of all the remaining restrictions. We consider that schools will remain closed, given the impending summer break in July and August 2020. Indeed, some school systems, such as the Boston public schools, have announced that they will remain closed for the rest of the 2019–2020 school year.
- Lift and enhanced tracing (LET) scenario: The stay-at-home order is lifted as in the previous scenario. Once partial reopening is implemented, we assume that 50% of symptomatic COVID-19 cases can be diagnosed with SARS-CoV-2 infection, on average, within 2d after the onset of symptoms and that they are isolated at home and their household members are quarantined successfully for 2 weeks (a sensitivity analysis for lower rates of isolation and quarantine is presented in the Supplementary Information). Although COVID-19 tests are highly specific, our

50% detection rate also accounts for imperfect testing. We also assume that a fraction of the non-household contacts (we show results for 20% and 40%) of the symptomatic infections can be traced and quarantined along with their household—note that we consider that the contacts are identified with a rate proportional to the duration of the interaction with the symptomatic individual.

The above scenarios are mechanistically simulated on the multi-layer network of Fig. 1a by allowing different interactions (between effective contacts) according to the simulated strategy. As a result, the average number of interactions in the workplace and community layer changes from 10.86 (95% CI: 1.51–42.39) under the unmitigated scenario, to 4.10 (95% CI: 0–23.79) for the partial lockdown and 0.89 (95% CI: 0–8.39) contacts for the stay-at-home policy (further details in Methods and Supplementary Information). This result is in agreement with previously published work¹⁸ and recent reports in the New York City area¹⁹. It is worth remarking that the fluctuations in the number of contacts under the stay-at-home policy is due to a large extent to contacts that take place in grocery stores and other public venues.

The numerical results show that the LIFT scenario, while able to temporarily abate the epidemic incidence, does not prevent the resurgence of the epidemic and a second COVID-19 wave when the social-distancing measures are relaxed. In Fig. 2d, we show that following the lifting of social-distancing, the infection incidence starts to increase again, and the effective reproductive number, which dropped by around 75% and reached values below 1 with the intervention, increases to values up to 2.05 (95%CI: 1.73–2.47) (Fig. 2f). Indeed, at the time of lifting the social-distancing intervention, the population has not achieved the level of herd immunity that would protect it from a resurgence of the epidemic. It is important to stress that here we do not consider additional mitigation measures such as behavioural changes in the population, such as mask wearing and similar measures (see the Supplementary Information). We also estimate that a second wave of the epidemic still has the potential to infect a large fraction of the population (Fig. 2e) and to overwhelm the healthcare system, as shown in Fig. 3b. The number of ICU beds needed, although it half of the number of the unmitigated scenario, still far exceeds the estimated availability, as pointed out in similar scenario analysis^{5–7,20}. This suggests that lifting social distancing without the support of additional containment strategies is not a viable option.

In the case of the LET scenario, the lifting of the social-distancing intervention is accompanied by a substantial amount of contact tracing and precautionary quarantine of potentially exposed individuals. The quarantine is not limited to the contacts of the identified

Table 1 | Mean (and 95% CI) of the number of normal hospitalizations, ICU hospitalizations and symptomatic individuals identified or traced (when applicable) at the peak of the epidemic per 1,000 people. The estimated availability of ICU beds is 0.21 beds per 1,000 people

Scenario	Hospitalization	ICU	Individuals traced		
Unmitigated	4.57 (4.10–5.03)	2.56 (2.21–2.91)	-		
LIFT	3.22 (2.80–3.67)	1.87 (1.55–2.20)	-		
	No Tracing	2.70 (2.29–3.12)	1.58 (1.27–1.88)		
LET	Detection 30%	Tracing 20%	0.86 (0.65–1.10)	0.55 (0.39–0.72)	0.52 (0.36–0.69)
		Tracing 60%	0.35 (0.21–0.50)	0.22 (0.12–0.34)	0.17 (0.08–0.27)
	No Tracing	2.35 (1.97–2.75)	1.39 (1.11–1.68)	-	
LET	Detection 50%	Tracing 20%	0.44 (0.28–0.62)	0.28 (0.16–0.42)	0.39 (0.23–0.55)
		Tracing 40%	0.29 (0.18–0.43)	0.15 (0.08–0.26)	0.14 (0.05–0.23)

symptomatic COVID-19 sufferer, but extends to their households. This strategy amounts to a simplified tracing of contacts of contacts, which would not require extensive investigations. In other words, this strategy does not require the tracking of a large number of single contacts, but instead leverages the contacts' households as the basic unit²¹. Households could, however, be monitored with daily calls or messages to determine the onset of symptomatic infections and provide medical support as needed.

Figure 2g shows results obtained for different levels of tracing (no tracing, and 20% and 40% tracing) of the contacts of the symptomatic isolated COVID-19 cases. By comparing Fig. 2d with Fig. 2g (for no tracing), we find that quarantining households of symptomatic individuals alone is not sufficient to substantially change the course of the epidemic and the conclusions reached for the first of these scenarios. When 40% or more of the contacts of the detected symptomatic infections are traced and they and their households are quarantined, the ensuing reduction in transmission leads to a noticeable flattening of the epidemic curve and appears to effectively limit the possible resurgence of a second epidemic wave. It is also worth noticing that we assume the absence of other additional and minimally disruptive social-distancing policies such as crowd control, smart working, wearing of masks, and others that could lead to a further reduction of the transmissibility of the virus with respect to our estimates. In the Supplementary Information, we investigate the positive synergistic effect of the combination of the LET strategy in the presence of different levels of transmissibility reduction due to additional mitigation policies. It is important to stress that the contact tracing proposed here works at the level of household unit, simplifying the monitoring and follow up process, by contacting only one member of the household to monitor the onset of symptoms among all members (we further explore other isolation and quarantine strategies in the Supplementary Information). Figure 3c and Table 1 show the burden on hospitalization and ICU demand in the unmitigated situation and the two mitigation scenarios. The LET scenario allows relaxation of the social-distancing interventions while maintaining the hospital and ICU demand at levels close to the healthcare availability and surge capacity. For the sake of completeness, the SM file includes analysis of a LIFT scenario including schools and universities reopening in the autumn. The results show that in the absence of additional containment policies, the tracing effort would need to be raised to about 50% to cope with the increased number of infections.

Discussion

The efforts in the suppression and mitigation of COVID-19 are pursuing the objectives of protecting the healthcare system from disruptive failures due to overwhelming stress imposed by the large number of severe cases and of minimizing the morbidity and mortality

related to the epidemic. The aggressive social-distancing interventions implemented by many countries in response to the COVID-19 pandemic appear to have achieved the interruption of transmission and the abatement of the epidemic, but this has been at the price of huge societal disruption and economic costs. In such a context, the identification of 'exit strategies' that allow restarting economic and social activities while still protecting healthcare systems and minimizing the burden of the epidemic is of primary importance. Several modelling studies have already pointed out that resuming economic activities and social life is likely to lead to a resurgence of the COVID-19 epidemic, and combined social-distancing interventions of different degrees and intensity have been proposed to substantially delay and mitigate the epidemic^{16,20}. These interventions still generate economic loss and widespread disruption to social life. Here we show how testing, contact-tracing strategies at scale, based on home isolation of symptomatic individuals with COVID-19, and the quarantine of a fraction of their households' contacts, have the potential to provide a viable course of action to manage and mitigate the epidemic when social-distancing interventions are progressively lifted^{22,23}. These strategies present us with logistic challenges that include large-scale and rapid diagnostic capacity, and a large surge in the number of contact tracers. We have investigated what fraction of the population would be isolated or quarantined under the proposed contact-tracing and isolation strategy. Figure 4a shows the fraction of households that needs to be quarantined. Assuming the identification of 50% of the symptomatic infections, and tracing of 40% of their contacts and households, only about 9% of the population would be quarantined at any time. While this is certainly a relevant fraction of the population, it is a much better option compared with massive social-distancing policies affecting the entire population that last for many months.

In Table 1, we report the number of symptomatic infections for which the contact-tracing investigation should be performed in the basic scenarios. This number provides an estimate of the number of contact tracers per 1,000 people. It is important to note that the more effective the contact tracing starting from each individual is, the smaller the number of generally traced households will be, because the epidemic has lower incidence rates. In addition, as illustrated in Fig. 4b, the health status of the vast majority of quarantined individuals is unknown, as contact tracing does not imply testing. The curves in Fig. 4a constitute the upper bounds for each simulated case. If we assume that the capacity to do massive testing will probably ramp up in the near future, then it is expected that the actual number of people in quarantine could be significantly lowered by testing the quarantined household. This would also alleviate the burden on household members that were unable to go to work and increase compliance of quarantine for the positive cases. It is also worth remarking that many of the logistic challenges faced

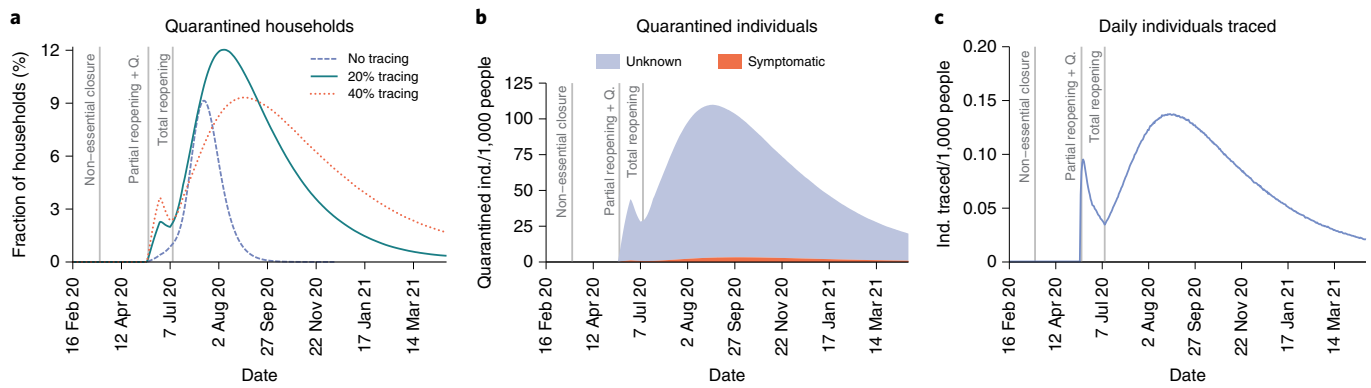


Fig. 4 | Affordability of the best way-out scenario. **a–c**, LET strategy with 50% detection and 40% tracing. **a**, Fraction of the population that needs to be put under quarantine as a function of time and percentage of contact tracing. **b**, Health status of the individuals that are quarantined for a contact-tracing level of 40%. Note that only symptomatic individuals are tested, which implies that a large fraction of the quarantined population is of unknown status. This fraction of individuals quarantined with unknown health conditions could be reduced if the capacity to test increases. The pandemic might span over several months depending on the level of contact tracing. **c**, Number of individuals whose contacts are traced each day, per 1,000 people. Relevant intervention actions are signalled by vertical dashed lines in all panels. x-axes show dates in 2020 (20) and 2021 (21). Jan, January.

with massive contact tracing might be eased by digital technologies that are currently being investigated across the world, following the examples of COVID-19 response in Asian countries²³. Also, it may be difficult to quarantine the entire household of individuals who are potentially exposed, since this is a hardship suffered with high uncertainty about their risk of infection. Offering other logistic quarantine solutions (quarantine centres or hotel rooms) might substantially increase the rate of compliance.

These results were obtained under several assumptions. There are very large uncertainties around the transmission of SARS-CoV-2; in particular, about the fraction of subclinical and asymptomatic cases and their transmission. Estimates of age-specific severity are informed by the analysis on individual-level data from China and other countries, and subject to change as more US data become available. We also do not include specific co-morbidities or pre-existing conditions of the specific BMA population. For this reason, we perform an extensive sensitivity analysis showing that the modelling results discussed here are robust to the plausible range of parameter values for the key time-to-event intervals of COVID-19 (for example, incubation period, serial interval, time from symptom onset to hospital admission, among other) as well as the fraction of pre-symptomatic and asymptomatic transmission (Supplementary Information). We are also not considering here potential changes to the virus transmissibility due to environmental factors; in particular, seasonal drivers such as temperature and humidity. The modelling does not consider possible reintroduction of SARS-CoV-2 in the population from infected travellers. Strategies based on testing, isolation and contact tracing might be hampered by the importation through travel of a large number of infections, thus travel restrictions and screening may need to be introduced to and from places that show sustained local transmission. Finally, we report in the Supplementary Information the effect of the widespread use in the population of masks or other personal protective equipment that lead to a reduction of the transmissibility of SARS-Cov-2. These active protection measures improve the effectiveness of the exit strategies modelled here.

The modelling of the impact of testing, contact tracing and isolation on second-wave scenarios of the COVID-19 epidemic could be instrumental for national and international agencies for public health response planning. While we show that contact tracing and household quarantine at scale may be effective, even assuming a complete lifting of the social-distancing measures, future decisions on when and for how long to relax policies will need to be informed

by ongoing surveillance. For instance, smart working from home for people who can adhere to it without serious disruptions should be encouraged. This, as well as other minimally disturbing policies, together with efficient and large-scale testing, contact tracing and monitoring of the epidemic, should be considered in the definition of exit strategies from large-scale social-distancing policies.

Methods

Weighted synthetic population. Our synthetic population constitutes around 85,000 nodes, of which 64,000 are adults and 21,000 are children (defined as individuals aged 17 or less), as shown in Fig. 1b. The total number of interactions among these individuals before social distancing is given by more than five million edges, see Supplementary Information for a more detailed description.

Community-weighted contact network. The community network is approximated using six months of data observation in the Boston area from anonymized users who have opted-in to provide access to their location data anonymously through a GDPR-compliant framework provided by Cuebiq. Individuals performing the analysis were legally required to never single out identifiable individuals and not make attempts to link these data to third-party data about an individual. In this layer, each agent in our synthetic population represents an anonymous individual of the real population. The data allow us to understand how infection can propagate in each layer by estimating colocation of two individuals in the same setting. We use a large database of 83,000 places from Foursquare application programming interface in the BMA. Specifically, the weight, $\omega_{C_{ij}}$, of a link between individuals i and j within the workplace plus community layer is computed according to the expression:

$$\omega_{C_{ij}} = \sum_p \frac{T_{ip} T_{jp}}{T_i T_j}, \quad \forall i, j$$

where T_{ip} is the total time that individual i was observed at place p and T_i is the total time that individual i has been observed at any place set within the workplace plus community layer. Note that while the mobility dataset we use is large, colocation events between individuals are still quite sparse. Because of this sparsity, and to protect individual privacy in our analysis, we have adopted this probabilistic approach to measure co-presence in all locations mapped in the dataset. Since agents are representative of the different census areas and groups of the Boston area, our probabilistic approach is a good proxy for the real probability of co-presence between those groups and areas when networks are scaled up to the total population of the Boston area, which is approximately 4,628,910 inhabitants. Finally, for robustness and computational reasons, we have included only links for which $\omega_{C_{ij}} > 0.01$.

Household-weighted contact network. We first localize individuals' approximate home place according to the US census block group. Then we assign a type of household based on Table B11016: Household Type by Household Size from US Census 2018 (ref. ³⁰), and mix individuals that live in the same block according to statistics of household type and size. Finally, children are assigned to households (see the Supplementary Information for a more detailed description). We also

assign individuals to an age group, on the basis of Table B01001: Sex and age from the US Census 2018. To assign weights, we assume that the probability of interaction within a household is proportional to the number of people living in the same household (well-mixing). Therefore, the weight, $\omega_{H_{ij}}$, of a link between individuals i and j within the same household is given by:

$$\omega_{H_{ij}} = \frac{1}{(n_h - 1)}$$

where n_h is the number of household members. This fraction is assumed to be the same for all individuals in the population.

School-weighted contact network. To calculate the weights of the links at the school layer, we mix together all children that live in the same school catchment area. Interactions are considered well-mixed; hence, the probability of interaction at a school is proportional to the number of children at the same school. Therefore, the weight, $\omega_{S_{ij}}$, of a link between children i and j within the same school is given by:

$$\omega_{S_{ij}} = \frac{1}{(n_s - 1)}$$

where n_s is the number of school members.

Calibration of intra-layer links. Within each connected component of the network in each layer (for example, a household or a school), the links between nodes are weighted to account for the effective daily number of contacts. For example, if we consider a school, while a student can potentially contact all her or his schoolmates, she or he only meet a relatively small fraction of them on a daily basis, as estimated in empirical studies on mixing patterns^{24,25}. To account for this, we calibrate the weight of the links in each layer of the synthetic network²⁶ so that the mean number of daily contacts matches the estimation provided in Mistry et al.²⁷ (details in Supplementary Information). On the basis of the analysis of contact survey data from 9 countries^{24,25,28,29}, this study estimated the mean number of daily contacts at 10.86, 4.11 and 11.41 in the community and workplace, household and school layers, respectively.

Stochastic simulations of COVID-19 dynamics. We describe the SARS-CoV-2 transmission process using a discrete-time susceptible–latent–infected–removed (SLIR) stochastic model, with some additional compartments to incorporate the special characteristics of SARS-CoV-2 infection (Fig. 1c). In particular, at each time-step t (1 d), the infectious asymptomatic (I_A), infectious symptomatic (I_S) and pre-symptomatic (P_S) individuals can transmit the disease to susceptible (S) individuals with probability $r\beta$, β and β_S , respectively. If the transmission is successful, the susceptible node will move to the latent asymptomatic state (L_A) with probability p or to the latent symptomatic state (L_S) with probability $(1-p)$. A latent asymptomatic individual becomes infectious asymptomatic after a period $(e)^{-1}$, whereas latent symptomatic individuals transition, after a period e^{-1} , to the pre-symptomatic (P_S) compartment. The average period to develop the disease and move to the infectious symptomatic state is γ^{-1} . Infectious asymptomatic nodes are removed (R) after an average of μ steps. Conversely, infectious symptomatic nodes can either recover after that period with probability $(1-\alpha)$ or, with probability α , these nodes will need hospitalization. It is considered that due to their symptoms they will self-isolate at home after an average period of μ^{-1} . Then, depending on the severity of the symptoms, after a period δ^{-1} the individual will end in hospitalization with probability $(1-\chi)$ or require hospitalization and ICU care with probability χ . Finally, individuals that are either hospitalized or in ICU become removed with probability μ_H or μ_{ICU} , respectively. We initialize the model in the city of Boston by selecting an attack rate on 17 March 2020 of 1.5% (a sensitivity analysis of this quantity is provided in the Supplementary Information).

Social-distancing strategies. To simulate social-distancing measures, we modify the synthetic population such that:

- School closures are simulated by removing all the schools from the system simultaneously.
- Partial ‘stay-at-home’ assumes that all places are open except restaurants, nightlife and cultural places. Closures of these places are simulated by removing the interactions that occur in any place that falls into that category according to Foursquare’s taxonomy of places. This is the situation after the first reopening.
- Full lockdown and confinement assume that schools and all non-essential workplaces are closed. Here we close all workplaces except essential ones and remove interactions that occur there. Essential workplaces are: hospitals, salons, barbershops, grocery stores, dispensaries, supermarkets, pet stores, pharmacies, urgent care centres, dry cleaners, drugstores, maternity clinics, medical supplies and petrol stations.

The connectivity distributions for each of the scenarios simulated as well as other statistics related to the effects of the lockdown are shown in the Supplementary Information.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available from Cuebiq through their Data for Good programme, but restrictions apply to the availability of these data, which were used under licence for the current study, and so are not publicly available. Aggregated data used in the models are however available from the authors upon reasonable request and permission of Cuebiq. Other data used comes from the American Community Survey (5-year) from the Census, which is publicly available.

Code availability

Custom code that supports the findings of this study is available from the corresponding author upon request.

Received: 22 June 2020; Accepted: 20 July 2020;

Published online: 05 August 2020

References

1. World Health Organization. *Novel Coronavirus—China* (2020); <https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/>
2. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
3. Kraemer, M. U. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
4. Flaxman, S., Mishra, S., Gandy, A., Bhatt, S., et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* <https://doi.org/10.1038/s41586-020-2405-7> (2020).
5. Walker, P.G.T. et al. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* **369**, 413–422 (2020).
6. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860–868 (2020).
7. Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P.-Y. & Colizza, V. Expected impact of lockdown in Île-de-France and possible exit strategies. Preprint at *medRxiv* <https://www.medrxiv.org/content/10.1101/2020.04.13.20063933v1> (2020).
8. Liu, Q.-H. et al. Measurability of the epidemic reproduction number in data-driven contact networks. *Proc. Natl Acad. Sci.* **115**, 12680–12685 (2018).
9. US Census Bureau. *2018 American Community Survey 5-Year Data* (2019); <https://www.census.gov/programs-surveys/acs>
10. Poletto, C., Meloni, S., Colizza, V., Moreno, Y. & Vespignani, A. Host mobility drives pathogen competition in spatially structured populations. *PLoS Comput. Biol.* **9**, e1003169 (2013).
11. Zhang, J. et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect. Dis.* **20**, 793–802 (2020).
12. Li, Q. et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).
13. Verity, R. et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* **20**, 669–677 (2020).
14. The White House. *15 Days to Slow the Spread* (2020); https://www.whitehouse.gov/wp-content/uploads/2020/03/03.16.20_coronavirus-guidance-8.5x11_315PM.pdf
15. Lee, A. These states have implemented stay-at-home orders. Here’s what that means for you. *CNN* <https://www.cnn.com/2020/03/23/us/coronavirus-which-states-stay-at-home-order-trnd/index.html> (23 March 2020).
16. Ahir, H., Bloom, M. & Furceri, D. Global uncertainty related to coronavirus at record high. *IMF Blog* <https://blogs.imf.org/2020/04/04/global-uncertainty-related-to-coronavirus-at-record-high/> (2020).
17. Gottlieb, S., Rivers, C., McClellan, M. B., Silvis, L. & Watson, C. National coronavirus response: a road map to reopening (American Enterprise Institute, 2020); <https://www.aei.org/research-products/report/national-coronavirus-response-a-road-map-to-reopening/>
18. Zhang, J. et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368**, 1481–1486 (2020).
19. Bakker, M., Berke, A., Groh, M., Pentland, A. & Moro, E. *Effect of Social Distancing Measures in the New York City Metropolitan Area* (MIT IDSS and MIT Media Lab, 2020); <http://curveflattening.media.mit.edu/posts/social-distancing-new-york-city/>
20. Leung, K., Wu, J. T., Liu, D. & Leung, G. M. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *Lancet* **395**, 1382–1393 (2020).

21. Kucharski, A. J. et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *Lancet Infect. Dis.* [https://doi.org/10.1016/S1473-3099\(20\)30457-6](https://doi.org/10.1016/S1473-3099(20)30457-6) (2020).
22. Hellewell, J. et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* **8**, e488–e496 (2020).
23. Ferretti, L. et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).
24. Mossong, J. et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74 (2008).
25. Ajelli, M. & Litvinova, M. Estimating contact patterns relevant to the spread of infectious diseases in Russia. *J. Theor. Biol.* **419**, 1–7 (2017).
26. Litvinova, M., Liu, Q.-H., Kulikov, E. S. & Ajelli, M. Reactive school closure weakens the network of social interactions and reduces the spread of influenza. *Proc. Natl Acad. Sci. USA* **116**, 13174–13181 (2019).
27. Mistry, D. et al., Inferring high-resolution human mixing patterns for disease modeling. Preprint at *arXiv* <https://arxiv.org/abs/2003.01214> (2020).
28. Béraud, G. et al. The French connection: the first large population-based contact survey in France relevant for the spread of infectious diseases. *PLoS ONE* **10**, 0133203 (2015).
29. Zhang, J. et al. patterns of human social contact and contact with animals in Shanghai, China. *Sci. Rep.* **9**, 1–11 (2019).

Acknowledgements

N.E.D., I.M.L.Jr, M.E.H., A.Py.P. and A.V. acknowledge the support of NIH/NIAID AI139761. M.C. and A.V. acknowledge support from Google Cloud Healthcare and Life Sciences Solutions via the GCP research credits program. M.E.H. acknowledges support from the NIH/NIGMS U54 GM111274. M.C., A.Py.P. and A.V. acknowledge the support of the McGovern Foundation. E.M. acknowledges partial support by MINECO (FIS2016-78904-C3-3-P). Y.M. acknowledges partial support from the Government of Aragon and

FEDER funds, Spain, through grant E36-17R (FENOL), and by MINECO and FEDER funds (FIS2017-87519-P). A.A. and Y.M. acknowledge support from Intesa Sanpaolo Innovation Center. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

A.A., D.M.-C., M.A., A.V., E.M. and Y.M. designed the research; A.A. performed research with contributions from D.M.-C.; A.A., D.M.-C., M.A., A.V., E.M. and Y.M. analysed the results. A.V. and Y.M. wrote the first draft of the manuscript; A.A., D.M.-C., A.Py.P., M.A., M.L., M.C., N.E.D., M.E.H., I.M.L.Jr, S.M., A.P., A.V., E.M. and Y.M. discussed results and edited the manuscript. All authors approved the final version.

Competing interests

M.E.H. reports grants from the National Institute of General Medical Sciences during the conduct of the study; A.V. reports grants and personal fees from Metabiota Inc. outside of the submitted work; M.C. and A.Py.P. report grants from Metabiota Inc. outside of the submitted work. The authors declare no other relationships or activities that could appear to have influenced the submitted work.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41562-020-0931-9>.

Correspondence and requests for materials should be addressed to A.V., E.M. or Y.M.

Peer review information Primary handling editor: Stavroula Kousta.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study is available from Cuebiq through their Data for Good program, but restrictions apply to the availability of these data, which were used under licenses for the current study, and so are not publicly available. Aggregated data used in the models are however available from the authors upon reasonable request and permission of Cuebiq. Other data used comes from the American Community Survey (5-year) from the Census, which is publicly available.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The mobility data was obtained from Cuebiq, a location intelligence and measurement company. The dataset consists of anonymized records of GPS locations from users that opted-in to share the data anonymously in the Boston metropolitan area over a period of 6 months, from October 2016 to March 2017.
Research sample	It consists of 65k agents (2% of the population in the Boston Metropolitan Area), 64k (75%) of them are adults and 21k (25%) corresponds to children.
Sampling strategy	The dataset consists of anonymized records of GPS locations from users that opted-in to share the data anonymously in the Boston metropolitan area over a period of 6 months, from October 2016 to March 2017.
Data collection	do not apply
Timing	From October 2016 to March 2017
Data exclusions	no data was excluded
Non-participation	do not apply
Randomization	do not apply

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging