

# Mobility network models of COVID-19 explain inequities and inform reopening

Serina Chang<sup>\*,1</sup>, Emma Pierson<sup>\*,1</sup>, Pang Wei Koh<sup>\*,1</sup>,  
Jaline Gerardin<sup>2</sup>, Beth Redbird<sup>3,4</sup>, David Grusky<sup>5,6</sup>, Jure Leskovec<sup>†,1,7</sup>

<sup>1</sup> Department of Computer Science, Stanford University, Stanford, CA 94305, USA

<sup>2</sup> Department of Preventive Medicine, Northwestern University, Chicago, IL 60611, USA

<sup>3</sup> Department of Sociology, Northwestern University, Evanston, IL, 60208, USA

<sup>4</sup> Institute for Policy Research, Northwestern University, Evanston, IL, 60208, USA

<sup>5</sup> Department of Sociology, Stanford University, Stanford, CA 94305, USA

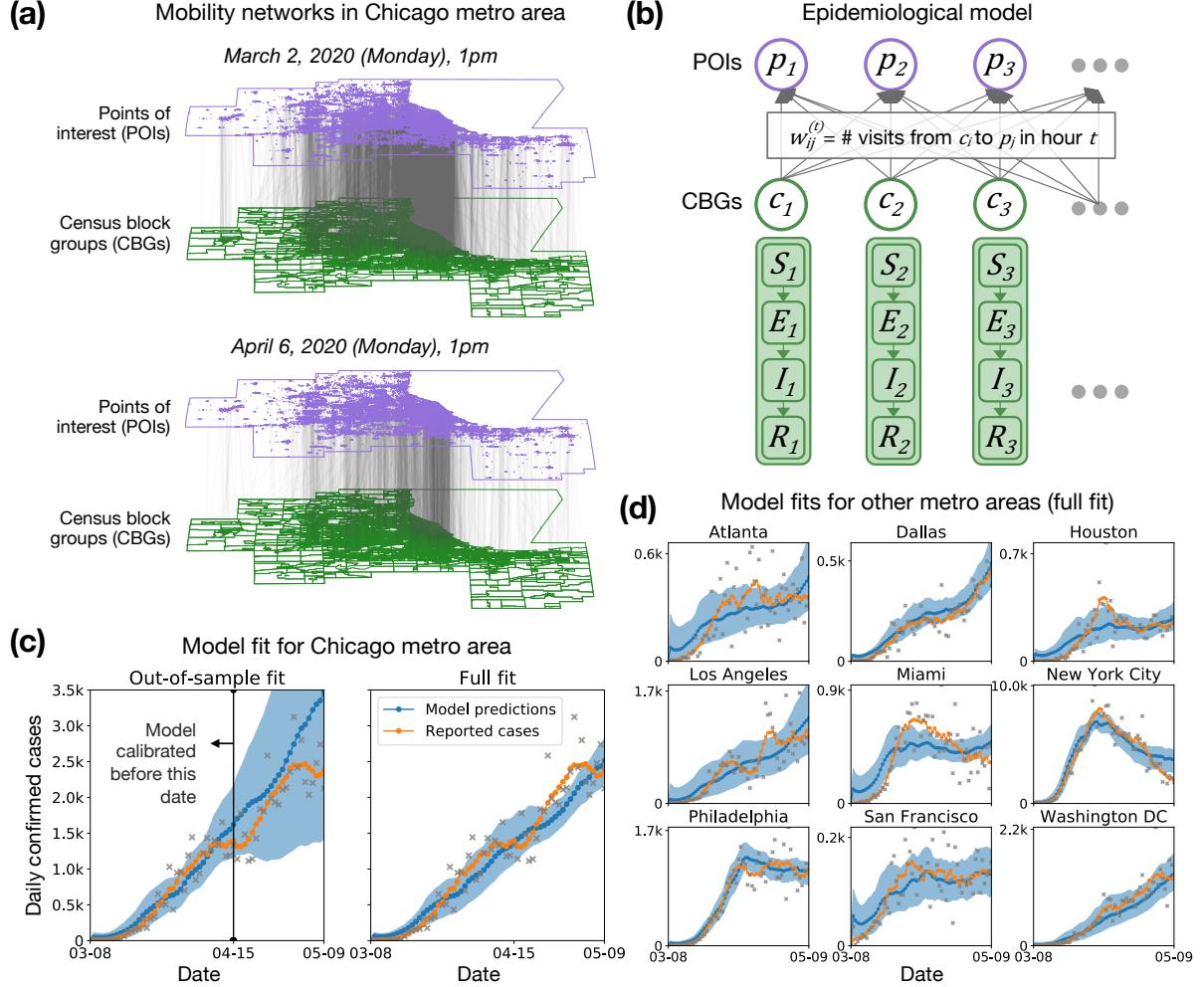
<sup>6</sup> Center on Poverty and Inequality, Stanford University, Stanford, CA 94305, USA

<sup>7</sup> Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

\* These authors contributed equally to this work

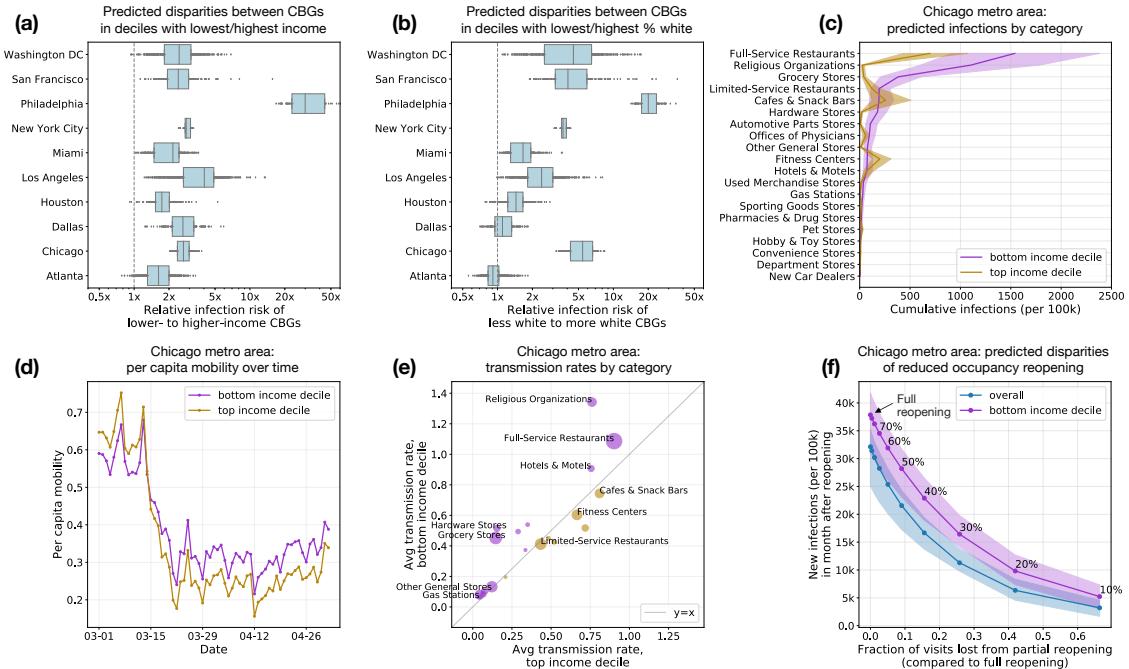
† Corresponding author. Email: jure@cs.stanford.edu

The COVID-19 pandemic dramatically changed human mobility patterns, necessitating epidemiological models which capture the effects of changes in mobility on virus spread.<sup>1</sup> We introduce a metapopulation SEIR model that integrates fine-grained, dynamic mobility networks to simulate the spread of SARS-CoV-2 in 10 of the largest US metropolitan statistical areas. Derived from cell phone data, our mobility networks map the hourly movements of 98 million people from neighborhoods (census block groups, or CBGs) to points of interest (POIs) such as restaurants and religious establishments, connecting 57k CBGs to 553k POIs with 5.4 billion hourly edges. We show that by integrating these networks, a relatively simple SEIR model can accurately fit the real case trajectory, despite substantial changes in population behavior over time. Our model predicts that a small minority of “superspread” POIs account for a large majority of infections and that restricting maximum occupancy at each POI is more effective than uniformly reducing mobility. Our model also correctly predicts higher infection rates among disadvantaged racial and socioeconomic groups<sup>2–8</sup> solely from differences in mobility: we find that disadvantaged groups have not been able to reduce mobility as sharply, and that the POIs they visit are more crowded and therefore higher-risk. By capturing who is infected at which locations, our model supports detailed analyses that can inform more effective and equitable policy responses to COVID-19.



**Figure 1:** Model description and fit. **(a)** The mobility network captures hourly visits from each census block group (CBG) to each point of interest (POI). The vertical lines indicate that most visits are between nearby POIs and CBGs. Visits dropped dramatically from March (top) to April (bottom), as indicated by the lower density of grey lines. **(b)** We overlaid a disease spread model on the mobility network, with each CBG having its own set of SEIR compartments. New infections occur at both POIs and CBGs, with the mobility network governing how subpopulations from different CBGs interact as they visit POIs. **(c)** Left: To test out-of-sample prediction, we calibrated the model on data before April 15, 2020 (vertical black line). Even though its parameters remain fixed over time, the model accurately predicts the case trajectory in the Chicago metro area after April 15 by using mobility data (RMSE on daily cases, Apr 15–May 2 = 406). Right: Model fit further improved when we calibrated the model on the full range of data (RMSE on daily cases, Apr 15–May 2 = 387). **(d)** We fit separate models to 10 of the largest US metropolitan statistical areas, modeling a total population of 98 million people; here, we show full model fits, as in (c)-Right. In (c) and (d), the blue line represents model predictions and grey x's represent the daily reported cases; since they tend to have great variability, we also show the smoothed weekly average (orange line). Shaded regions denote 2.5th and 97.5th percentiles across parameter sets and stochastic realizations. Across metro areas, we sample 97 parameter sets, with 30 stochastic realizations each ( $N = 2,910$ ); see Table S6 for the number of sets per metro area.





**Figure 3:** Mobility patterns give rise to infection disparities. For (c–f), the Chicago metro area is used as an example, but references to results for all metro areas are provided for each subfigure. **(a)** In every metro area, our model predicts that people in lower-income CBGs are likelier to be infected. **(b)** People in non-white CBGs area are also likelier to be infected, although results are more variable across metro areas. **(c)** The overall predicted disparity is driven by a few POI categories like full-service restaurants (Figure S2). **(d)** One reason for the predicted disparities is that higher-income CBGs were able to reduce their mobility levels below those of lower-income CBGs (Extended Data Figure 6). **(e)** Within each POI category, people from lower-income CBGs tend to visit POIs that have higher predicted transmission rates (Extended Data Table 3). The size of each dot represents the average number of visits per capita made to the category. The top 10 out of 20 categories with the most visits are labeled, covering 0.48–2.88 visits per capita (Hardware Stores–Full-Service Restaurants). **(f)** Reopening (at different levels of reduced maximum occupancy) leads to more predicted infections in lower-income CBGs than the overall population (Extended Data Figure 3). In (c–f), purple denotes lower-income CBGs, yellow denotes higher-income CBGs, and blue represents the overall population. Aside from (d) and (e), which were directly extracted from mobility data, all results in this figure represent predictions aggregated over model realizations. Across metro areas, we sample 97 parameter sets, with 30 stochastic realizations each ( $N = 2,910$ ); see Table S6 for the number of sets per metro area. Shaded regions in (c) and (f) denote the 2.5th–97.5th percentiles; boxes in (a–b) denote the interquartile range, with data points outside the range individually shown.

## **1 Introduction**

2 In response to the COVID-19 crisis, stay-at-home orders were enacted in many countries to reduce  
3 contact between individuals and slow the spread of the virus.<sup>9,10</sup> Since then, public officials have  
4 continued to deliberate over when to reopen, which places are safe to return to, and how much  
5 activity to allow.<sup>11</sup> Answering these questions requires epidemiological models that can capture  
6 the effects of changes in mobility on virus spread. In particular, findings of COVID-19 “super-  
7 spreader” events<sup>12–15</sup> motivate models that can reflect the heterogeneous risks of visiting different  
8 locations, while well-reported disparities in infection rates<sup>2–8</sup> require models that can explain the  
9 disproportionate impact of the virus on disadvantaged groups.

10 To address these needs, we construct fine-grained dynamic mobility networks from cell  
11 phone geolocation data, and use these networks to model the spread of SARS-CoV-2 within 10  
12 of the largest metropolitan statistical areas (referred to below as metro areas) in the United States.  
13 These networks map the hourly movements of 98 million people from census block groups (CBGs),  
14 which are geographical units that typically contain 600–3,000 people, to specific points of inter-  
15 est (POIs). As shown in Table S1, POIs are non-residential locations that people visit such as  
16 restaurants, grocery stores, and religious establishments. On top of each network, we overlay a  
17 metapopulation SEIR model that tracks the infection trajectories of each CBG as well as the POIs  
18 at which these infections are likely to have occurred. This builds upon prior work that models dis-  
19 ease spread using aggregate<sup>16–19</sup>, historical<sup>20–22</sup>, or synthetic mobility data<sup>23–25</sup>; separately, other  
20 work has analyzed mobility data in the context of COVID-19, but without an underlying model of  
21 disease spread.<sup>26–30</sup>

22 Combining our epidemiological model with these mobility networks allows us to not only  
23 accurately fit observed case counts, but also to conduct detailed analyses that can inform more  
24 effective and equitable policy responses to COVID-19. By capturing information about individual  
25 POIs (e.g., hourly number of visitors, median visit duration), our model can estimate the impacts  
26 of specific reopening strategies, such as only reopening certain POI categories or restricting maxi-  
27 mum occupancy at each POI. By modeling movement from CBGs, our model can identify at-risk  
28 populations and correctly predicts, solely from mobility patterns, that disadvantaged racial and  
29 socioeconomic groups face higher rates of infection. Our model thus enables analysis of urgent  
30 health inequities; we use it to illuminate two mobility-related mechanisms driving these disparities  
31 and to evaluate the disparate impact of reopening on disadvantaged groups.

## 32 Mobility network model

33 We use data from SafeGraph, a company that aggregates anonymized location data from mobile  
34 applications, to study mobility patterns from March 1–May 2, 2020. For each metro area, we  
35 represent the movement of individuals between CBGs and POIs as a bipartite network with time-  
36 varying edges, where the weight of an edge between a CBG and POI represents the number of  
37 visitors from that CBG to that POI at a given hour (Figure 1a). SafeGraph also provides the area  
38 in square feet of each POI, as well as its North American Industry Classification System (NAICS)  
39 category (e.g., fitness center, full-service restaurant) and median visit duration in minutes. We  
40 validated the SafeGraph mobility data by comparing to Google mobility data (Figure S1, Tables  
41 S2–S3), and used iterative proportional fitting<sup>31</sup> to derive POI-CBG networks from the raw Safe-  
42 Graph data. Overall, these networks comprise 5.4 billion hourly edges between 56,945 CBGs and  
43 552,758 POIs (Extended Data Table 1).

44 We overlay a SEIR model on each mobility network,<sup>16,20</sup> where each CBG maintains its own  
45 susceptible (S), exposed (E), infectious (I), and removed (R) states (Figure 1b). New infections  
46 occur at both POIs and CBGs, with the mobility network governing how subpopulations from  
47 different CBGs interact as they visit POIs. We use each POI’s area, median visit duration, and  
48 time-varying density of infectious individuals to determine the POI’s hourly infection rate. The  
49 model has only three free parameters, which scale (1) transmission rates at POIs, (2) transmission  
50 rates at CBGs, and (3) the initial proportion of exposed individuals (Extended Data Table 2); all  
51 three parameters remain constant over time. We calibrate a separate model for each metro area  
52 using confirmed case counts from *The New York Times* by minimizing root mean square error  
53 (RMSE) to daily incident cases.<sup>32</sup> Our model accurately fits observed daily case counts in all 10  
54 metro areas from March 8–May 9, 2020 (Figure 1c,d). Additionally, when only calibrated on case  
55 counts up to April 14, the model predicted case counts reasonably well on the held-out time period  
56 from April 15–May 9, 2020 (Figure 1c and Extended Data Figure 1a). Our key technical finding  
57 is that the dynamic mobility network allows even our SEIR model with just three static parameters  
58 to accurately fit observed cases, despite changing policies and behaviors during that period.

## 59 Mobility reduction and reopening plans

60 **The magnitude of mobility reduction is as important as its timing.** Mobility in the US dropped  
61 sharply in March 2020: e.g., overall POI visits in the Chicago metro area fell by 54.7% between  
62 the first week of March and the first week of April 2020. We constructed counterfactual mobility

networks by scaling the magnitude of mobility reduction down and by shifting the timeline earlier and later, and applied our model to the counterfactual networks to simulate the resulting infection trajectories. Across metro areas, we found that the magnitude of mobility reduction was at least as important as its timing (Figure 2a, Tables S4–S5): e.g., if the mobility reduction in the Chicago metro area had been only one quarter as large, predicted infections would have increased by  $3.3 \times$  (95% CI, 2.8–3.8), compared to a  $1.5 \times$  (95% CI, 1.4–1.6) increase had people begun reducing their mobility one full week later. Furthermore, if no mobility reduction had occurred at all, predicted infections in the Chicago metro area would have increased by a striking  $6.2 \times$  (95% CI, 5.2–7.1). Our results concord with earlier findings that mobility reductions can dramatically reduce infections.<sup>19,33,34</sup>

**A minority of POIs account for a majority of predicted infections.** We next investigated if how we reduce mobility—i.e., to which POIs—matters. We computed the number of infections that occurred at each POI in our simulations from March 1–May 2, 2020, and found that a majority of predicted infections occurred at a small fraction of “superspread” POIs; e.g., in the Chicago metro area, 10% of POIs accounted for 85% (95% CI, 83%–87%) of the predicted infections at POIs (Figure 2b, Figure S10). Certain categories of POIs also contributed far more to infections (e.g., full-service restaurants, hotels), although our model predicted time-dependent variation in how much each category contributed (Extended Data Figure 2). For example, restaurants and fitness centers contributed less to predicted infections over time, likely due to lockdown orders closing these POIs, while grocery stores remained steady or even grew in their contribution, which concords with their status as essential businesses.

**Reopening with reduced maximum occupancy.** If a minority of POIs produce the majority of infections, then reopening strategies that specifically target high-risk POIs should be especially effective. To test one such strategy, we simulated reopening on May 1, and modeled the effects of reducing maximum occupancy in which the numbers of hourly visits to each POI returned to their “normal” levels from the first week of March but were capped if they exceeded a fraction of the POI’s maximum occupancy.<sup>35</sup> Full reopening without reducing maximum occupancy produced a spike in predicted infections: in the Chicago metro area, our models projected that an additional 32% (95% CI, 25%–35%) of the population would be infected by the end of May (Figure 2c). However, reducing maximum occupancy substantially reduced risk without sharply reducing over-

93 all mobility: capping at 20% maximum occupancy in the Chicago metro area cut down predicted  
94 new infections by more than 80% but only lost 42% of overall visits, and we observed similar  
95 trends across other metro areas (Extended Data Figure 3). This highlights the non-linearity of pre-  
96 dicted infections as a function of visits: one can achieve a disproportionately large reduction in  
97 infections with a small reduction in visits. Furthermore, compared to another reopening strategy  
98 that uniformly reduced visits to each POI from their levels in early March, reducing maximum oc-  
99 cupancy always resulted in fewer predicted infections for the same number of total visits (Figure 2c  
100 and Extended Data Figure 4). This is because reduced maximum occupancy takes advantage of  
101 the time-varying visit density within each POI, disproportionately reducing visits to the POI during  
102 the most risky high-density periods, but leaving visit counts unchanged during less risky periods.  
103 These results support earlier findings that targeted closures may be more effective than region-wide  
104 measures, while incurring substantially lower economic costs.<sup>36</sup>

105 **Relative risk of reopening different categories of POIs.** Because we found that certain POI  
106 categories contributed far more to predicted infections in March (Extended Data Figure 2), we also  
107 expected that reopening some POI categories would be riskier than reopening others. To assess this,  
108 we simulated reopening each category in turn on May 1 (by returning its mobility patterns to early  
109 March levels, as before), while keeping all other POIs at their reduced mobility levels from the end  
110 of April. We found large variation in predicted reopening risks: on average across metro areas, full-  
111 service restaurants, gyms, hotels, cafes, religious organizations, and limited-service restaurants  
112 produced the largest predicted increases in infections when reopened (Extended Data Figure 5d).  
113 Reopening full-service restaurants was particularly risky: in the Chicago metro area, we predicted  
114 an additional 596k (95% CI, 434k-686k) infections by the end of May, more than triple the next  
115 riskiest POI category (Figure 2d). These risks are summed over all POIs in the category, but  
116 the relative risks after normalizing by the number of POIs were broadly similar (Extended Data  
117 Figure 5c). These categories were predicted to more be dangerous because, in the mobility data,  
118 their POIs tended to have higher visit densities and/or visitors stayed there longer (Figures S15–  
119 S24).

## 120 **Demographic disparities in infections**

121 We characterize the differential spread of SARS-CoV-2 along demographic lines by using US  
122 Census data to annotate each CBG with its racial composition and median income, then tracking

123 predicted infection rates in CBGs with different demographic compositions: for example, within  
124 each metro area, comparing CBGs in the top and bottom deciles for income. We use this approach  
125 to study the mobility mechanisms behind disparities and to quantify how different reopening strate-  
126 gies impact disadvantaged groups.

127 **Predicting disparities from mobility data.** Despite only having access to mobility data and no  
128 other demographic information, our models correctly predicted higher risks of infection among  
129 disadvantaged racial and socioeconomic groups.<sup>2–8</sup> Across all metro areas, individuals from CBGs  
130 in the bottom decile for income were substantially likelier to have been infected by the end of  
131 the simulation, even though all individuals began with equal likelihoods of infection (Figure 3a).  
132 This predicted disparity was driven primarily by a few POI categories (e.g., full-service restau-  
133 rants), which infected far larger proportions of lower-income CBGs than higher-income CBGs  
134 (Figure 3c and Figure S2). We similarly found that CBGs with fewer white residents had higher  
135 predicted risks of infection, although results were more variable across metro areas (Figure 3b).  
136 In SI Discussion, we confirm that the magnitude of the disparities our model predicts are gen-  
137 erally consistent with real-world disparities and further explore the large predicted disparities in  
138 Philadelphia, which stem from substantial differences in the POIs that are frequented by higher-  
139 versus lower-income CBGs. In the analysis below, we discuss two mechanisms producing higher  
140 predicted infection rates among lower-income CBGs, and we show in Extended Data Figure 6b  
141 and Extended Data Table 4 that similar results hold for racial disparities as well.

142 **Lower-income CBGs saw smaller reductions in mobility.** First, across all metro areas, lower-  
143 income CBGs did not reduce their mobility as sharply in the first few weeks of March 2020,  
144 and had higher mobility than higher-income CBGs for most of March through May (Figure 3d,  
145 Extended Data Figure 6a). For example, in April, lower-income CBGs in the Chicago metro  
146 area had 27% more POI visits per capita than higher-income CBGs. Category-level differences in  
147 visit patterns partially explained the infection disparities within each category: e.g., lower-income  
148 CBGs made substantially more visits per capita to grocery stores than did higher-income CBGs  
149 (Figure S3), and consequently experienced more predicted infections at that category (Figure S2).

150 **POIs visited by lower-income CBGs have higher transmission rates.** Differences in visits per  
151 capita do not fully explain the infection disparities: for example, Cafes & Snack Bars were visited

more frequently by higher-income CBGs in every metro area (Figure S3), but our model predicted that Cafes & Snack Bars infected a larger proportion of lower-income CBGs in the majority of metro areas (Figure S2). We found that even within a POI category, the predicted transmission rates at POIs frequented by lower-income CBGs tended to be higher than the corresponding rates for higher-income CBGs (Figure 3e; Extended Data Table 3), because POIs frequented by lower-income CBGs tended to be smaller and more crowded in the mobility data. As a case study, we examined grocery stores in further detail. In 8 of the 10 metro areas, visitors from lower-income CBGs encountered higher predicted transmission rates at grocery stores than those from higher-income CBGs (median transmission rate ratio of 2.19, Extended Data Table 3). Why was one visit to the grocery store predicted to be twice as dangerous for a lower-income individual? SafeGraph data showed that the average grocery store visited by lower-income individuals had 59% more hourly visitors per square foot, and their visitors stayed 17% longer on average (medians across metro areas). These findings highlight how fine-grained differences in mobility patterns—how often people go out and which POIs they go to—can ultimately contribute to dramatic disparities in predicted infection outcomes.

**Reopening plans must account for disparate impact.** Because disadvantaged groups suffer a larger burden of infection, it is critical to not just consider the overall impact of reopening plans but also their disparate impact on disadvantaged groups specifically. For example, our model predicted that full reopening in the Chicago metro area would result in an additional 39% (95% CI, 31%-42%) of the population of CBGs in the bottom income decile being infected within a month, compared to 32% (95% CI, 25%-35%) of the overall population (Figure 3f; results for all metro areas in Extended Data Figure 3). Similarly, Figure S4 illustrates that reopening individual POI categories tends to have a larger predicted impact on lower-income CBGs. More conservative reopening plans produce smaller absolute disparities in predicted infections—e.g., we predict that reopening at 20% of maximum occupancy in Chicago would result in additional infections for 6% (95% CI, 4%-8%) of the overall population and 10% (95% CI, 7%-13%) of the population in CBGs in the bottom income decile (Figure 3f)—though the relative disparity remains.

## Discussion

The mobility dataset we use has limitations: it does not cover all populations, does not contain all POIs, and cannot capture sub-CBG heterogeneity. Our model itself is also parsimonious, and does

not include all real-world features relevant to disease transmission. We discuss these limitations in more detail in SI Discussion. However, the predictive accuracy of our model suggests that it broadly captures the relationship between mobility and transmission, and we thus expect our broad conclusions—e.g., that people from lower-income CBGs have higher infection rates in part because they tend to visit denser POIs and because they have not reduced mobility by as much (likely because they cannot as easily work from home<sup>4</sup>)—to hold robustly. Our fine-grained network modeling approach naturally extends to other mobility datasets and models which capture more aspects of real-world transmission, and this remains an interesting direction for future work.

Our results can guide policymakers seeking to assess competing approaches to reopening. Despite growing concern about racial and socioeconomic disparities in infections and deaths, it has been difficult for policymakers to act on those concerns; they are currently operating without much evidence on the disparate impacts of reopening policies, prompting calls for research that both identifies the causes of observed disparities and suggests policy approaches to mitigate them.<sup>5,8,37,38</sup> Our fine-grained mobility modeling addresses both these needs. Our results suggest that infection disparities are not the unavoidable consequence of factors that are difficult to address in the short term, like differences in preexisting conditions; on the contrary, short-term policy decisions can substantially affect infection outcomes by altering the overall amount of mobility allowed, the types of POIs reopened, and the extent to which POI occupancies are capped. Considering the disparate impact of reopening plans may lead policymakers to, e.g., (1) favor more conservative reopening plans, (2) increase testing in disadvantaged neighborhoods predicted to be high risk (especially given known disparities in access to tests<sup>2</sup>), and (3) prioritize distributing masks and other personal protective equipment to disadvantaged populations. As reopening policies continue to be debated, it is critical to build models that can assess the effectiveness and equity of competing policies. We hope that our approach, by capturing heterogeneity across POIs, demographic groups, and cities, helps address this need.

## References

1. Buckee, C. O. *et al.* Aggregated mobility data could help fight COVID-19. *Science* **368**, 145 (2020).
2. Wilson, C. These Graphs Show How COVID-19 Is Ravaging New York City's Low-Income Neighborhoods. *Time* (2020). Available at <https://time.com/5821212/coronavirus-low-income-communities/>.
3. Garg, S. *et al.* Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 — COVID-NET, 14 States, March 1—30, 2020 (CDC Morbidity and Mortality Weekly Report (MMWR), 2020). Available at <https://www.cdc.gov/mmwr/volumes/69/wr/mm6915e3.htm>.
4. Reeves, R. V. & Rothwell, J. Class and COVID: How the less affluent face double risks. *The Brookings Institution* (2020). Available at <https://www.brookings.edu/blog/up-front/2020/03/27/class-and-covid-how-the-less-affluent-face-double-risks/>.
5. Pareek, M. *et al.* Ethnicity and COVID-19: an urgent public health research priority. *The Lancet* **395**, 1421–1422 (2020).
6. van Dorn, A., Cooney, R. E. & Sabin, M. L. COVID-19 exacerbating inequalities in the US. *The Lancet* **395**, 1243–1244 (2020).
7. Yancy, C. W. COVID-19 and African Americans. *JAMA* **323**, 1891–1892 (2020).
8. Chowkwanyun, M. & Reed Jr, A. L. Racial Health Disparities and Covid-19—Caution and Context. *New England Journal of Medicine* (2020).
9. Wu, J. W., Smith, S., Khurana, M., Siemaszko, C. & DeJesus-Banos, B. Stay-at-home orders across the country. *NBC News* (2020). Available at <https://www.nbcnews.com/health/health-news/heres-what-stay-home-orders-across-country-n1168736>.
10. Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* (2020).
11. Rojas, R. & Delkic, M. As states reopen, governors balance existing risks with new ones. *The New York Times* (2020). Available at <https://www.nytimes.com/2020/05/17/us/coronavirus-states-reopen.html>.
12. Endo, A., Abbott, S., Kucharski, A. J., Funk, S. *et al.* Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research* **5**, 67 (2020).
13. Adam, D. *et al.* Clustering and superspreading potential of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections in Hong Kong. Available at <https://europepmc.org/article/ppr/ppr165671>.
14. Park, S. Y. *et al.* Coronavirus Disease Outbreak in Call Center, South Korea. *Emerging Infectious Diseases* **26** (2020).
15. Bi, Q. *et al.* Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases* (2020).
16. Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).

17. Jia, J. S. *et al.* Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* (2020).
18. Pei, S., Kandula, S. & Shaman, J. Differential Effects of Intervention Timing on COVID-19 Spread in the United States. *medRxiv* (2020). Available at doi.org/10.1101/2020.05.15.20103655.
19. Lai, S. *et al.* Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* (2020).
20. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* **368**, 489–493 (2020).
21. Pei, S. & Shaman, J. Initial Simulation of SARS-CoV2 Spread and Intervention Effects in the Continental US. *medRxiv* (2020). Available at doi.org/10.1101/2020.03.21.20040303.
22. Aleta, A. *et al.* Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic. *medRxiv* (2020). Available at doi.org/10.1101/2020.05.06.20092841.
23. Duque, D. *et al.* COVID-19: How to Relax Social Distancing If You Must. *medRxiv* (2020). Available at doi.org/10.1101/2020.04.29.20085134.
24. Block, P. *et al.* Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nature Human Behaviour* (2020).
25. Karin, O. *et al.* Adaptive cyclic exit strategies from lockdown to suppress COVID-19 and allow economic activity. *medRxiv* (2020). Available at doi.org/10.1101/2020.04.04.20053579.
26. Gao, S., Rao, J., Kang, Y., Liang, Y. & Kruse, J. Mapping county-level mobility pattern changes in the united states in response to covid-19. *SIGSPATIAL Special* **12**, 16–26 (2020).
27. Klein, B. *et al.* Assessing changes in commuting and individual mobility in major metropolitan areas in the United States during the COVID-19 outbreak (2020). Available at networkscienceinstitute.org/publications/assessing-changes-in-commuting-and-individual-mobility-in-major-metropolitan-areas-in-the-united-states-during-the-covid-19-outbreak.
28. Benzell, S. G., Collis, A. & Nicolaides, C. Rationing social contact during the COVID-19 pandemic: Transmission risk and social benefits of US locations. *Proceedings of the National Academy of Sciences* (2020).
29. Baicker, K., Dube, O., Mullainathan, S., Devin, P. & Wezerek, G. Is It Safer to Visit a Coffee Shop or a Gym? *The New York Times* (2020). Available at <https://nytimes.com/interactive/2020/05/06/opinion/coronavirus-us-reopen.html>.
30. Hsiang, S. *et al.* The effect of large-scale anti-contagion policies on the coronavirus (COVID-19) pandemic. *Nature* (2020).
31. Deming, W. E. & Stephan, F. F. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics* **11**, 427–444 (1940).
32. The New York Times. Coronavirus (Covid-19) Data in the United States (2020). Available at <https://github.com/nytimes/covid-19-data>.

33. Tian, H. *et al.* An investigation of transmission control measures during the first 50 days of the covid-19 epidemic in china. *Science* **368**, 638–642 (2020).
34. Watts, D. J., Muhamad, R., Medina, D. C. & Dodds, P. S. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences* **102**, 11157–11162 (2005).
35. California Department of Public Health. COVID-19 Industry Guidance: Retail (2020). Available at <https://covid19.ca.gov/pdf/guidance-retail.pdf>.
36. Birge, J., Candogan, O. & Feng, Y. Controlling epidemic spread: Reducing economic losses with targeted closures. *Working paper, Becker Friedman Institute* (2020). Available at [https://bfi.uchicago.edu/wp-content/uploads/BFI\\_WP\\_202057-1.pdf](https://bfi.uchicago.edu/wp-content/uploads/BFI_WP_202057-1.pdf).
37. Webb Hooper, M., Nápoles, A. M. & Pérez-Stable, E. J. COVID-19 and Racial/Ethnic Disparities. *JAMA* (2020).
38. Laurencin, C. T. & McClinton, A. The COVID-19 Pandemic: a Call to Action to Identify and Address Racial and Ethnic Disparities. *Journal of Racial and Ethnic Health Disparities* **7**, 398–402 (2020).

207 **Methods**

208 The Methods section is structured as follows. We describe the datasets we use in Methods M1  
209 and the mobility network that we derive from these datasets in Methods M2. In Methods M3,  
210 we discuss the SEIR model we overlay on the mobility network; in Methods M4, we describe  
211 how we calibrate this model and quantify uncertainty in its predictions. Finally, in Methods M5,  
212 we provide details on the experimental procedures used for our analyses of mobility reduction,  
213 reopening plans, and demographic disparities.

214 **M1 Datasets**

215 **SafeGraph.** We use data provided by SafeGraph, a company that aggregates anonymized loca-  
216 tion data from numerous mobile applications. SafeGraph data captures the movement of people  
217 between census block groups (CBGs), which are geographical units that typically contain a pop-  
218 ulation of between 600 and 3,000 people, and points of interest (POIs) like restaurants, grocery  
219 stores, or religious establishments. Specifically, we use the following SafeGraph datasets:

- 220 1. Places Patterns<sup>39</sup> and Weekly Patterns (v1)<sup>40</sup>. These datasets contain, for each POI, hourly  
221 counts of the number of visitors, estimates of median visit duration in minutes (the “dwell  
222 time”), and aggregated weekly and monthly estimates of visitors’ home CBGs. We use visi-  
223 tor home CBG data from the Places Patterns dataset, as described below: for privacy reasons,  
224 SafeGraph excludes a home CBG from this dataset if fewer than 5 devices were recorded at  
225 the POI from that CBG over the course of the month. For each POI, SafeGraph also provides  
226 their North American Industry Classification System (NAICS) category, as well as estimates  
227 of its median visit duration in minutes and physical area in square feet. (Area is computed  
228 using the footprint polygon SafeGraph assigns to the POI.<sup>41,42</sup>) We analyze Places Patterns  
229 data from January 1, 2019 to February 29, 2020 and Weekly Patterns data from March 1,  
230 2020 to May 2, 2020.
- 231 2. Social Distancing Metrics,<sup>43</sup> which contains daily estimates of the proportion of people stay-  
232 ing home in each CBG. We analyze Social Distancing Metrics data from March 1, 2020 to  
233 May 2, 2020.

234 We focus on 10 of the largest metropolitan statistical areas in the US (Extended Data Table 1). We  
235 chose these metro areas by taking a random subset of the SafeGraph Patterns data and picking the  
236 10 metro areas with the most POIs in the data. Our methods in this paper can be straightforwardly

237 applied, in principle, to the other metro areas in the original SafeGraph data. For each metro area,  
238 we include all POIs that meet all of the following requirements: (1) the POI is located in the  
239 metro area; (2) SafeGraph has visit data for this POI for every hour that we model, from 12am  
240 on March 1, 2020 to 11pm on May 2, 2020; (3) SafeGraph has recorded the home CBGs of this  
241 POI’s visitors for at least one month from January 2019 to February 2020; (4) the POI is not a  
242 “parent” POI. “Parent” POIs comprise a small fraction of POIs in the dataset which overlap and  
243 include the visits from their “child” POIs: for example, many malls in the dataset are parent POIs  
244 which include the visits from stores which are their child POIs. To avoid double-counting visits,  
245 we remove all parent POIs from the dataset. After applying these POI filters, we include all CBGs  
246 that have at least 1 recorded visit to at least 10 of the remaining POIs; this means that CBGs from  
247 outside the metro area may be included if they visit this metro area frequently enough. Summary  
248 statistics of the post-processed data are in Extended Data Table 1. Overall, we analyze 57k CBGs  
249 from the 10 metro areas, and over 310M visits from these CBGs to nearly 553k POIs.

250 SafeGraph data has been used to study consumer preferences<sup>44</sup> and political polarization.<sup>45</sup>  
251 More recently, it has been used as one of the primary sources of mobility data in the US for  
252 tracking the effects of the SARS-CoV-2 pandemic.<sup>26,28,46–48</sup> In SI Methods Section 1, we show  
253 that aggregate trends in SafeGraph mobility data match up to aggregate trends in Google mobility  
254 data in the US,<sup>49</sup> before and after the imposition of stay-at-home measures. Previous analyses  
255 of SafeGraph data have shown that it is geographically representative: for example, it does not  
256 systematically over-represent individuals from CBGs in different counties or with different racial  
257 compositions, income levels, or educational levels.<sup>50,51</sup>

258 **US Census.** Our data on the demographics of census block groups (CBGs) comes from the US  
259 Census Bureau’s American Community Survey (ACS).<sup>52</sup> We use the 5-year ACS (2013-2017)  
260 to extract the median household income, proportion of white residents, and proportion of black  
261 residents of each CBG. For the total population of each CBG, we use the most recent one-year  
262 estimates (2018); one-year estimates are noisier but we wish to minimize systematic downward  
263 bias in our total population counts (due to population growth) by making them as recent as possible.

264 **New York Times.** We calibrate our models using the COVID-19 dataset published by the *The*  
265 *New York Times*.<sup>32</sup> Their dataset consists of cumulative counts of cases and deaths in the United  
266 States over time, at the state and county level. For each metro area that we model, we sum over

267 the county-level counts to produce overall counts for the entire metro area. We convert the cumu-  
268 lative case and death counts to daily counts for the purposes of model calibration, as described in  
269 Methods M4.

270 **Data ethics.** The dataset from *The New York Times* consists of aggregated COVID-19 confirmed  
271 case and death counts collected by journalists from public news conferences and public data re-  
272 leases. For the mobility data, consent was obtained by the third-party sources collecting the data.  
273 SafeGraph aggregates data from mobile applications that obtain opt-in consent from their users  
274 to collect anonymous location data. Google’s mobility data consists of aggregated, anonymized  
275 sets of data from users who have chosen to turn on the Location History setting. Additionally, we  
276 obtained IRB exemption for SafeGraph data from the Northwestern University IRB office.

## 277 M2 Mobility network

278 **Definition.** We consider a complete undirected bipartite graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  with time-varying  
279 edges. The vertices  $\mathcal{V}$  are partitioned into two disjoint sets  $\mathcal{C} = \{c_1, \dots, c_m\}$ , representing  $m$   
280 census block groups (CBGs), and  $\mathcal{P} = \{p_1, \dots, p_n\}$ , representing  $n$  points of interest (POIs).  
281 From US Census data, each CBG  $c_i$  is labeled with its population  $N_{c_i}$ , income distribution, and  
282 racial and age demographics. From SafeGraph data, each POI  $p_j$  is similarly labeled with its  
283 category (e.g., restaurant, grocery store, or religious organization), its physical size in square feet  
284  $a_{p_j}$ , and the median dwell time  $d_{p_j}$  of visitors to  $p_j$ . The weight  $w_{ij}^{(t)}$  on an edge  $(c_i, p_j)$  at time  $t$   
285 represents our estimate of the number of individuals from CBG  $c_i$  visiting POI  $p_j$  at the  $t$ -th hour  
286 of simulation. We record the number of edges (with non-zero weights) in each metro area and over  
287 all hours from March 1, 2020 to May 2, 2020 in Extended Data Table 1. Across all 10 metro areas,  
288 we study 5.4 billion edges between 56,945 CBGs and 552,758 POIs.

289 **Network estimation (overview).** The central technical challenge in constructing this network is  
290 estimating the network weights  $W^{(t)} = \{w_{ij}^{(t)}\}$  from SafeGraph data, since this visit matrix is not  
291 directly available from the data. Our general methodology for network estimation is as follows:

- 292 1. From SafeGraph data, we can derive a time-independent estimate  $\bar{W}$  of the visit matrix  
293 that captures the aggregate distribution of visits from CBGs to POIs from January 2019 to  
294 February 2020.

295     2. However, visit patterns differ substantially from hour to hour (e.g., day versus night) and  
296       day to day (e.g., pre- versus post-lockdown). To capture these variations, we use current  
297       SafeGraph data to estimate the CBG marginals  $U^{(t)}$ , i.e., the number of people in each CBG  
298       who are out visiting POIs at hour  $t$ , as well as the POI marginals  $V^{(t)}$ , i.e., the total number  
299       of visitors present at each POI  $p_j$  at hour  $t$ .

300     3. We then apply the iterative proportional fitting procedure (IPFP) to estimate an hourly visit  
301       matrix  $W^{(t)}$  that is consistent with the hourly marginals  $U^{(t)}$  and  $V^{(t)}$  but otherwise “as  
302       similar as possible” to the distribution of visits in the aggregate visit matrix  $\bar{W}$ , in terms of  
303       Kullback-Leibler divergence.

304     IPFP is a classic statistical method<sup>31</sup> for adjusting joint distributions to match pre-specified marginal  
305       distributions, and it is also known in the literature as biproportional fitting, the RAS algorithm, or  
306       raking<sup>53</sup>. In the social sciences, it has been widely used to infer the characteristics of local subpop-  
307       ulations (e.g., within each CBG) from aggregate data<sup>54-56</sup>. IPFP estimates the joint distribution of  
308       visits from CBGs to POIs by alternating between scaling each row to match the hourly row (CBG)  
309       marginals  $U^{(t)}$  and scaling each column to match the hourly column (POI) marginals  $V^{(t)}$ . For  
310       further details about the estimation procedure, we refer the reader to SI Methods Section 3.

### 311     **M3 Model dynamics**

312     To model the spread of SARS-CoV-2, we overlay a metapopulation disease transmission model on  
313       the mobility network defined in Methods M2. The transmission model structure follows prior work  
314       on epidemiological models of SARS-CoV-2<sup>16,20</sup> but incorporates a fine-grained mobility network  
315       into the calculations of the transmission rate. We construct separate mobility networks and models  
316       for each metropolitan statistical area.

317       We use a SEIR model with susceptible ( $S$ ), exposed ( $E$ ), infectious ( $I$ ), and removed ( $R$ )  
318       compartments. Susceptible individuals have never been infected, but can acquire the virus through  
319       contact with infectious individuals, which may happen at POIs or in their home CBG. They then  
320       enter the exposed state, during which they have been infected but are not infectious yet. Individuals  
321       transition from exposed to infectious at a rate inversely proportional to the mean latency period.  
322       Finally, they transition into the removed state at a rate inversely proportional to the mean infectious  
323       period. The removed state represents individuals who can no longer be infected or infect others,  
324       e.g., because they have recovered, self-isolated, or died.

325      Each CBG  $c_i$  maintains its own SEIR instantiation, with  $S_{c_i}^{(t)}$ ,  $E_{c_i}^{(t)}$ ,  $I_{c_i}^{(t)}$ , and  $R_{c_i}^{(t)}$  representing  
 326      how many individuals in CBG  $c_i$  are in each disease state at hour  $t$ , and  $N_{c_i} = S_{c_i}^{(t)} + E_{c_i}^{(t)} + I_{c_i}^{(t)} +$   
 327       $R_{c_i}^{(t)}$ . At each hour  $t$ , we sample the transitions between states as follows:

$$N_{S_{c_i} \rightarrow E_{c_i}}^{(t)} \sim \text{Pois}\left(\frac{S_{c_i}^{(t)}}{N_{c_i}} \sum_{j=1}^n \lambda_{p_j}^{(t)} w_{ij}^{(t)}\right) + \text{Binom}\left(S_{c_i}^{(t)}, \lambda_{c_i}^{(t)}\right) \quad (1)$$

$$N_{E_{c_i} \rightarrow I_{c_i}}^{(t)} \sim \text{Binom}\left(E_{c_i}^{(t)}, 1/\delta_E\right) \quad (2)$$

$$N_{I_{c_i} \rightarrow R_{c_i}}^{(t)} \sim \text{Binom}\left(I_{c_i}^{(t)}, 1/\delta_I\right), \quad (3)$$

328      where  $\lambda_{p_j}^{(t)}$  is the rate of infection at POI  $p_j$  at time  $t$ ;  $w_{ij}^{(t)}$ , the  $ij$ -th entry of the visit matrix from  
 329      the mobility network (Methods M2), is the number of visitors from CBG  $c_i$  to POI  $p_j$  at time  $t$ ;  
 330       $\lambda_{c_i}^{(t)}$  is the base rate of infection that is independent of visiting POIs;  $\delta_E$  is the mean latency period;  
 331      and  $\delta_I$  is the mean infectious period.

332      We then update each state to reflect these transitions. Let  $\Delta S_{c_i}^{(t)} := S_{c_i}^{(t+1)} - S_{c_i}^{(t)}$ , and likewise  
 333      for  $\Delta E_{c_i}^{(t)}$ ,  $\Delta I_{c_i}^{(t)}$ , and  $\Delta R_{c_i}^{(t)}$ . Then,

$$\Delta S_{c_i}^{(t)} := -N_{S_{c_i} \rightarrow E_{c_i}}^{(t)} \quad (4)$$

$$\Delta E_{c_i}^{(t)} := N_{S_{c_i} \rightarrow E_{c_i}}^{(t)} - N_{E_{c_i} \rightarrow I_{c_i}}^{(t)} \quad (5)$$

$$\Delta I_{c_i}^{(t)} := N_{E_{c_i} \rightarrow I_{c_i}}^{(t)} - N_{I_{c_i} \rightarrow R_{c_i}}^{(t)} \quad (6)$$

$$\Delta R_{c_i}^{(t)} := N_{I_{c_i} \rightarrow R_{c_i}}^{(t)}. \quad (7)$$

### 334      M3.1    The number of new exposures $N_{S_{c_i} \rightarrow E_{c_i}}^{(t)}$

335      We separate the number of new exposures  $N_{S_{c_i} \rightarrow E_{c_i}}^{(t)}$  in CBG  $c_i$  at time  $t$  into two parts: cases  
 336      from visiting POIs, which are sampled from  $\text{Pois}\left((S_{c_i}^{(t)}/N_{c_i}) \sum_{j=1}^n \lambda_{p_j}^{(t)} w_{ij}^{(t)}\right)$ , and other cases not  
 337      captured by visiting POIs, which are sampled from  $\text{Binom}\left(S_{c_i}^{(t)}, \lambda_{c_i}^{(t)}\right)$ .

338      **New exposures from visiting POIs.** We assume that any susceptible visitor to POI  $p_j$  at time  $t$   
 339      has the same independent probability  $\lambda_{p_j}^{(t)}$  of being infected and transitioning from the susceptible  
 340      ( $S$ ) to the exposed ( $E$ ) state. Since there are  $w_{ij}^{(t)}$  visitors from CBG  $c_i$  to POI  $p_j$  at time  $t$ , and  
 341      we assume that a  $S_{c_i}^{(t)}/N_{c_i}$  fraction of them are susceptible, the number of new exposures among  
 342      these visitors is distributed as  $\text{Binom}(w_{ij}^{(t)} S_{c_i}^{(t)}/N_{c_i}, \lambda_{p_j}^{(t)}) \approx \text{Pois}(\lambda_{p_j}^{(t)} w_{ij}^{(t)} S_{c_i}^{(t)}/N_{c_i})$ . The number of

343 new exposures among all outgoing visitors from CBG  $c_i$  is therefore distributed as the sum of the  
 344 above expression over all POIs,  $\text{Pois}\left(\left(S_{c_i}^{(t)} / N_{c_i}\right) \sum_{j=1}^n \lambda_{p_j}^{(t)} w_{ij}^{(t)}\right)$ .

345 We model the infection rate at POI  $p_j$  at time  $t$ ,  $\lambda_{p_j}^{(t)} := \beta_{p_j}^{(t)} \cdot I_{p_j}^{(t)} / V_{p_j}^{(t)}$ , as the product of its  
 346 transmission rate  $\beta_{p_j}^{(t)}$  and proportion of infectious individuals  $I_{p_j}^{(t)} / V_{p_j}^{(t)}$ , where  $V_{p_j}^{(t)} := \sum_{i=1}^m w_{ij}^{(t)}$   
 347 is the total number of visitors to  $p_j$  at time  $t$ . We model the transmission rate at POI  $p_j$  at time  $t$  as

$$\beta_{p_j}^{(t)} := \psi \cdot d_{p_j}^2 \cdot \frac{V_{p_j}^{(t)}}{a_{p_j}}, \quad (8)$$

348 where  $a_{p_j}$  is the physical area of  $p_j$ , and  $\psi$  is a transmission constant (shared across all POIs) that  
 349 we fit to data. The inverse scaling of transmission rate with area  $a_{p_j}$  is a standard simplifying  
 350 assumption.<sup>57</sup> The dwell time fraction  $d_{p_j} \in [0, 1]$  is what fraction of an hour an average visitor  
 351 to  $p_j$  at any hour will spend there (SI Methods Section 3); it has a quadratic effect on the POI  
 352 transmission rate  $\beta_{p_j}^{(t)}$  because it reduces both (1) the time that a susceptible visitor spends at  $p_j$   
 353 and (2) the density of visitors at  $p_j$ . With this expression for the transmission rate  $\beta_{p_j}^{(t)}$ , we can  
 354 calculate the infection rate at POI  $p_j$  at time  $t$  as

$$\begin{aligned} \lambda_{p_j}^{(t)} &= \beta_{p_j}^{(t)} \cdot \frac{I_{p_j}^{(t)}}{V_{p_j}^{(t)}} \\ &= \psi \cdot d_{p_j}^2 \cdot \frac{I_{p_j}^{(t)}}{a_{p_j}}. \end{aligned} \quad (9)$$

355 For sufficiently large values of  $\psi$  and a sufficiently large proportion of infected individuals, the  
 356 expression above can sometimes exceed 1. To address this, we simply clip the infection rate to 1.  
 357 However, this occurs very rarely for the parameter settings and simulation duration that we use.

358 Finally, to compute the number of infectious individuals at  $p_j$  at time  $t$ ,  $I_{p_j}^{(t)}$ , we assume that  
 359 the proportion of infectious individuals among the  $w_{kj}^{(t)}$  visitors to  $p_j$  from a CBG  $c_k$  mirrors the  
 360 overall density of infections  $I_{c_k}^{(t)} / N_{c_k}$  in that CBG, although we note that the scaling factor  $\psi$  can  
 361 account for differences in the ratio of infectious individuals who visit POIs. This gives

$$I_{p_j}^{(t)} := \sum_{k=1}^m \frac{I_{c_k}^{(t)}}{N_{c_k}} w_{kj}^{(t)}. \quad (10)$$

362 **Base rate of new exposures not captured by visiting POIs.** In addition to the new exposures  
 363 from infections at POIs, we model a CBG-specific base rate of new exposures that is independent of

<sup>364</sup> POI visit activity. This captures other sources of infections, e.g., household infections or infections  
<sup>365</sup> at POIs that are absent from the SafeGraph data. We assume that at each hour, every susceptible  
<sup>366</sup> individual in CBG  $c_i$  has a base probability  $\lambda_{c_i}^{(t)}$  of becoming infected and transitioning to the  
<sup>367</sup> exposed state, where

$$\lambda_{c_i}^{(t)} := \beta_{\text{base}} \cdot \frac{I_{c_i}^{(t)}}{N_{c_i}} \quad (11)$$

<sup>368</sup> is the product of the base transmission rate  $\beta_{\text{base}}$  and the proportion of infectious individuals in  
<sup>369</sup> CBG  $c_i$ .  $\beta_{\text{base}}$  is a constant (shared across all CBGs) that we fit to data.

<sup>370</sup> **Overall number of new exposures.** Putting all of the above together yields the expression for  
<sup>371</sup> the distribution of new exposures in CBG  $c_i$  at time  $t$ ,

$$\begin{aligned} N_{S_{c_i} \rightarrow E_{c_i}}^{(t)} &\sim \text{Pois}\left(\frac{S_{c_i}^{(t)}}{N_{c_i}} \sum_{j=1}^n \lambda_{p_j}^{(t)} w_{ij}^{(t)}\right) + \text{Binom}\left(S_{c_i}^{(t)}, \lambda_{c_i}^{(t)}\right) \\ &= \underbrace{\text{Pois}\left(\psi \cdot \frac{S_{c_i}^{(t)}}{N_{c_i}} \cdot \sum_{j=1}^n \frac{d_{p_j}^2}{a_{p_j}} \left(\sum_{k=1}^m \frac{I_{c_k}^{(t)}}{N_{c_k}} w_{kj}^{(t)}\right) w_{ij}^{(t)}\right)}_{\text{new infections from visiting POIs}} + \underbrace{\text{Binom}\left(S_{c_i}^{(t)}, \beta_{\text{base}} \cdot \frac{I_{c_i}^{(t)}}{N_{c_i}}\right)}_{\text{base rate of new CBG infections}}. \end{aligned} \quad (12)$$

### <sup>372</sup> M3.2 The number of new infectious and removed cases

<sup>373</sup> We model exposed individuals as becoming infectious at a rate inversely proportional to the mean  
<sup>374</sup> latency period  $\delta_E$ . At each time step  $t$ , we assume that each exposed individual has a constant,  
<sup>375</sup> time-independent probability of becoming infectious, with

$$N_{E_{c_i} \rightarrow I_{c_i}}^{(t)} \sim \text{Binom}\left(E_{c_i}^{(t)}, 1/\delta_E\right). \quad (13)$$

<sup>376</sup> Similarly, we model infectious individuals as transitioning to the removed state at a rate inversely  
<sup>377</sup> proportional to the mean infectious period  $\delta_I$ , with

$$N_{I_{c_i} \rightarrow R_{c_i}}^{(t)} \sim \text{Binom}\left(I_{c_i}^{(t)}, 1/\delta_I\right), \quad (14)$$

<sup>378</sup> We estimate  $\delta_E = 96$  hours<sup>20,58</sup> and  $\delta_I = 84$  hours<sup>20</sup> from prior literature.

379 **M3.3 Model initialization**

380 In our experiments,  $t = 0$  is the first hour of March 1, 2020. We approximate the infectious  $I$  and  
 381 removed  $R$  compartments at  $t = 0$  as initially empty, with all infected individuals in the exposed  
 382  $E$  compartment. We further assume the same expected initial prevalence  $p_0$  in every CBG  $c_i$ . At  
 383  $t = 0$ , every individual in the metro area has the same independent probability  $p_0$  of being exposed  
 384  $E$  instead of susceptible  $S$ . We thus initialize the model state by setting

$$S_{c_i}^{(0)} = N_{c_i} - E_{c_i}^{(0)} \quad (15)$$

$$E_{c_i}^{(0)} \sim \text{Binom}(N_{c_i}, p_0) \quad (16)$$

$$I_{c_i}^{(0)} = 0 \quad (17)$$

$$R_{c_i}^{(0)} = 0. \quad (18)$$

385 **M3.4 Aggregate mobility and no-mobility baseline models**

386 **Comparison to aggregate mobility model.** Our model uses a detailed mobility network to simu-  
 387 late disease spread. To test if this detailed model is necessary, or if our model is simply making use  
 388 of aggregate mobility patterns, we tested an alternate SEIR model that uses the aggregate number  
 389 of visits made to any POI in the metro area in each hour, but not the breakdown of visits between  
 390 specific CBGs to specific POIs. Like our model, the aggregate mobility model also captures new  
 391 cases from visiting POIs and a base rate of infection that is independent of POI visit activity; thus,  
 392 the two models have the same three free parameters ( $\psi$ , scaling transmission rates at POIs;  $\beta_{\text{base}}$ ,  
 393 the base transmission rate; and  $p_0$ , the initial fraction of infected individuals). However, instead of  
 394 having POI-specific rates of infection, the aggregate mobility model only captures a single prob-  
 395 ability that a susceptible person from any CBG will become infected due to a visit to any POI at  
 396 time  $t$ ; we make this simplification because the aggregate mobility model no longer has access to  
 397 the breakdown of visits between CBGs and POIs. This probability  $\lambda_{POI}^{(t)}$  is defined as

$$\lambda_{POI}^{(t)} = \psi \cdot \underbrace{\frac{\sum_{i=1}^m \sum_{j=1}^n w_{ij}^{(t)}}{nm}}_{\text{average mobility at time } t} \cdot \frac{I^{(t)}}{N}, \quad (19)$$

398 where  $m$  is the number of CBGs,  $n$  is the number of POIs,  $I^{(t)}$  is the total number of infectious  
 399 individuals at time  $t$ , and  $N$  is the total population size of the metro area. For the base rate of

400 infections in CBGs, we assume the same process as in our network model: the probability  $\lambda_{c_i}^{(t)}$  that  
 401 a susceptible person in CBG  $c_i$  will become infected in their CBG at time  $t$  is equal to  $\beta_{\text{base}}$  times  
 402 the current infectious fraction of  $c_i$  (Equation 11). Putting it together, the aggregate mobility model  
 403 defines the number of new exposures in CBG  $c_i$  at time  $t$  as

$$N_{S_{c_i} \rightarrow E_{c_i}}^{(t)} \sim \underbrace{\text{Binom}\left(S_{c_i}^{(t)}, \lambda_{POI}^{(t)}\right)}_{\text{new infections from visiting POIs}} + \underbrace{\text{Binom}\left(S_{c_i}^{(t)}, \lambda_{c_i}^{(t)}\right)}_{\text{base rate of new CBG infections}}. \quad (20)$$

404 All other dynamics remain the same between the aggregate mobility model and our network model,  
 405 and we calibrated the models in the same way, which we will describe in Methods M4. We found  
 406 that our network model substantially outperformed the aggregate mobility model in out-of-sample  
 407 cases prediction: on average across metro areas, our best-fit network model's out-of-sample RMSE  
 408 was only 58% that of the best-fit aggregate mobility model (Extended Data Figure 1). This demon-  
 409 strates that it is not only general mobility patterns, but specifically the mobility *network* that allows  
 410 our model to accurately fit observed cases.

411 **Comparison to baseline that does not use mobility data.** To determine the extent to which  
 412 mobility data might aid in modeling the case trajectory, we also compared our model to a baseline  
 413 SEIR model that does not use mobility data and simply assumes that all individuals within an  
 414 metro area mix uniformly. In this no-mobility baseline, an individual's risk of being infected and  
 415 transitioning to the exposed state at time  $t$  is

$$\lambda^{(t)} := \beta_{\text{base}} \cdot \frac{I^{(t)}}{N}, \quad (21)$$

416 where  $I^{(t)}$  is the total number of infectious individuals at time  $t$ , and  $N$  is the total population size  
 417 of the metro area. As above, the other model dynamics are identical, and for model calibration we  
 418 performed a similar grid search over  $\beta_{\text{base}}$  and  $p_0$ . As expected, we found both the network and  
 419 aggregate mobility models outperformed the no-mobility model on out-of-sample case predictions  
 420 (Extended Data Figure 1).

## 421 M4 Model calibration and validation

422 Most of our model parameters can either be estimated from SafeGraph and US Census data, or  
 423 taken from prior work (see Extended Data Table 2 for a summary). This leaves 3 model parameters

424 that do not have direct analogues in the literature, and that we therefore need to calibrate with data:

425 1. The transmission constant in POIs,  $\psi$  (Equation (9))

426 2. The base transmission rate,  $\beta_{\text{base}}$  (Equation (11))

427 3. The initial proportion of exposed individuals at time  $t = 0$ ,  $p_0$  (Equation (16)).

428 In this section, we describe how we fit these parameters to published numbers of confirmed cases,  
429 as reported by *The New York Times*. We fit models for each metro area separately.

#### 430 M4.1 Selecting parameter ranges

431 **Transmission rate factors  $\psi$  and  $\beta_{\text{base}}$ .** We select parameter ranges for the transmission rate fac-  
432 tors  $\psi$  and  $\beta_{\text{base}}$  by checking if the model outputs match plausible ranges of the basic reproduction  
433 number  $R_0$  pre-lockdown, since  $R_0$  has been the study of substantial prior work on SARS-CoV-  
434 2.<sup>59</sup> Under our model, we can decompose  $R_0 = R_{\text{base}} + R_{\text{POI}}$ , where  $R_{\text{POI}}$  describes transmission  
435 due to POIs and  $R_{\text{base}}$  describes the remaining transmission (as in Equation (12)). We first establish  
436 plausible ranges for  $R_{\text{base}}$  and  $R_{\text{POI}}$  before translating these into plausible ranges for  $\beta_{\text{base}}$  and  $\psi$ .

437 We assume that  $R_{\text{base}}$  ranges from 0.1–2.  $R_{\text{base}}$  models transmission that is not correlated  
438 with activity at POIs in the SafeGraph dataset, including within-household transmission and trans-  
439 mission at POI categories which are not well-captured in the SafeGraph dataset. We chose the  
440 lower limit of 0.1 because beyond that point, base transmission would only contribute minimally  
441 to overall  $R$ , whereas previous work suggests that within-household transmission is a substantial  
442 contributor to overall transmission.<sup>60–62</sup> Household transmission alone is not estimated to be suf-  
443 ficient to tip overall  $R_0$  above 1; for example, a single infected individual has been estimated to  
444 cause an average of 0.32 (0.22, 0.42) secondary within-household infections.<sup>60</sup> However, because  
445  $R_{\text{base}}$  may also capture transmission at POIs not captured in the SafeGraph dataset, to be conser-  
446 vative, we chose an upper limit of  $R_{\text{base}} = 2$ ; as we describe below, the best-fit models for all 10  
447 metro areas have  $R_{\text{base}} < 2$ , and 9 out of 10 have  $R_{\text{base}} < 1$ . We allow  $R_{\text{POI}}$  to range from 1–3,  
448 which corresponds to allowing  $R_0 = R_{\text{POI}} + R_{\text{base}}$  to range from 1.1–5. This is a conservatively  
449 wide range, since prior work estimates a pre-lockdown  $R_0$  of 2–3.<sup>59</sup>

450 To determine the values of  $R_{\text{base}}$  and  $R_{\text{POI}}$  that a given pair of  $\beta_{\text{base}}$  and  $\psi$  imply, we seeded a  
451 fraction of index cases and then ran the model on looped mobility data from the first week of March  
452 to capture pre-lockdown conditions. We initialized the model by setting  $p_0$ , the initial proportion

453 of exposed individuals at time  $t = 0$ , to  $p_0 = 10^{-4}$ , and then sampling in accordance with Equation  
 454 (16). Let  $N_0$  be the number of initial exposed individuals sampled. We computed the number of  
 455 individuals that these  $N_0$  index cases went on to infect through base transmission,  $N_{\text{base}}$ , and POI  
 456 transmission,  $N_{\text{POI}}$ , which gives

$$R_{\text{POI}} = \frac{N_{\text{POI}}}{N_0} \quad (22)$$

$$R_{\text{base}} = \frac{N_{\text{base}}}{N_0}. \quad (23)$$

457 We averaged these quantities over stochastic realizations for each metro area. Figure S6 shows  
 458 that, as expected,  $R_{\text{base}}$  is linear in  $\beta_{\text{base}}$  and  $R_{\text{POI}}$  is linear in  $\psi$ .  $R_{\text{base}}$  lies in the plausible range  
 459 when  $\beta_{\text{base}}$  ranges from 0.0012–0.024, and  $R_{\text{POI}}$  lies in the plausible range (for at least one metro  
 460 area) when  $\psi$  ranges from 515–4,886, so these are the parameter ranges we consider when fitting  
 461 the model. As described in Methods M4.2, we verified that case count data for all metro areas can  
 462 be fit using parameter settings for  $\beta_{\text{base}}$  and  $\psi$  within these ranges.

463 **Initial prevalence of exposures,  $p_0$ .** The extent to which SARS-CoV-2 infections had spread  
 464 in the US by the start of our simulation (March 1, 2020) is currently unclear.<sup>63</sup> To account for  
 465 this uncertainty, we allow  $p_0$  to vary across a large range between  $10^{-5}$  and  $10^{-2}$ . As described  
 466 in Methods M4.2, we verified that case count data for all metro areas can be fit using parameter  
 467 settings for  $p_0$  within this range.

## 468 M4.2 Fitting to the number of confirmed cases

469 Using the parameter ranges above, we grid searched over  $\psi$ ,  $\beta_{\text{base}}$ , and  $p_0$  to find the models that  
 470 best fit the number of confirmed cases reported by *The New York Times* (NYT).<sup>32</sup> For each metro  
 471 area, we tested 1,500 different combinations of  $\psi$ ,  $\beta_{\text{base}}$ , and  $p_0$  in the parameter ranges specified  
 472 above, with parameters linearly spaced for  $\psi$  and  $\beta_{\text{base}}$  and logarithmically spread for  $p_0$ .

473 In Methods M3, we directly model the number of infections but not the number of confirmed  
 474 cases. To estimate the number of confirmed cases, we assume that an  $r_c = 0.1^{20,58,64-66}$  proportion  
 475 of infections will be confirmed, and moreover that they will be confirmed exactly  $\delta_c = 168$  hours  
 476 (7 days)<sup>20,66</sup> after becoming infectious. From these assumptions, we can calculate the predicted

477 number of newly confirmed cases across all CBGs in the metro area on day  $d$ ,

$$N_{\text{cases}}^{(\text{day } d)} = r_c \cdot \sum_{i=1}^m \sum_{\tau=24(d-1)+1-\delta_c}^{24d-\delta_c} N_{E_{c_i} \rightarrow I_{c_i}}^{(\tau)}, \quad (24)$$

478 where  $m$  indicates the total number of CBGs in the metro area and for convenience we define  
 479  $N_{E_{c_i} \rightarrow I_{c_i}}^{(\tau)}$ , the number of newly infectious people at hour  $\tau$ , to be 0 when  $\tau < 1$ .

480 From NYT data, we have the reported number of new cases  $\hat{N}_{\text{cases}}^{(\text{day } d)}$  for each day  $d$ , summed  
 481 over each county in the metro area. We compare the reported number of cases and the number of  
 482 cases that our model predicts by computing the root-mean-squared-error (RMSE) between each of  
 483 the  $D = \lfloor T/24 \rfloor$  days of our simulations,

$$\text{RMSE} = \sqrt{\frac{1}{D} \sum_{d=1}^D \left( N_{\text{cases}}^{(\text{day } d)} - \hat{N}_{\text{cases}}^{(\text{day } d)} \right)^2}. \quad (25)$$

484 For each combination of model parameters and for each metro area, we quantify model fit with the  
 485 NYT data by running 30 stochastic realizations and averaging their RMSE. Note that we measure  
 486 model fit based on the daily number of new reported cases (as opposed to the cumulative number  
 487 of reported cases).<sup>67</sup>

488 Our simulation spans March 1 to May 2, 2020, and we use mobility data from that period.  
 489 However, because we assume that cases will be confirmed  $\delta_c = 7$  days after individuals become  
 490 infectious, we predict the number of cases with a 7 day offset, from March 8 to May 9, 2020.

### 491 M4.3 Parameter selection and uncertainty quantification

492 Throughout this paper, we report aggregate predictions from different parameter sets of  $\psi, \beta_{\text{base}}$ ,  
 493 and  $p_0$  and multiple stochastic realizations. For each metro area, we:

494 1. Find the best-fit parameter set, i.e., with the lowest average RMSE on daily incident cases  
 495 over stochastic realizations.

496 2. Select all parameter sets that achieve an RMSE (averaged over stochastic realizations) within  
 497 20% of the RMSE of the best-fit parameter set.

498 3. Pool together all predictions across those parameter sets and all of their stochastic realiza-  
 499 tions, and report their mean and 2.5th/97.5th percentiles.

500 On average, each metro area has 9.7 parameter sets that achieve an RMSE within 20% of the  
501 best-fitting parameter set (Table S6). For each parameter set, we have results for 30 stochastic  
502 realizations.

503 This procedure corresponds to rejection sampling in an Approximate Bayesian Computation  
504 framework,<sup>16</sup> where we assume an error model that is Gaussian with constant variance; we pick an  
505 acceptance threshold based on what the best-fit model achieves; and we use a uniform parameter  
506 grid instead of sampling from a uniform prior. It quantifies uncertainty from two sources. First, the  
507 multiple realizations capture stochastic variability between model runs with the same parameters.  
508 Second, simulating with all parameter sets that are within 20% of the RMSE of the best fit captures  
509 uncertainty in the model parameters  $\psi$ ,  $\beta_{\text{base}}$ , and  $p_0$ . The latter is equivalent to assuming that the  
510 posterior probability over the true parameters is uniformly spread among all parameter sets within  
511 the 20% threshold.

#### 512 **M4.4 Model validation on out-of-sample cases**

513 We validate our models by showing that they predict the number of confirmed cases on out-of-  
514 sample data when we have access to corresponding mobility data. For each metro area, we split  
515 the available NYT dataset into a training set (spanning March 8, 2020 to April 14, 2020) and  
516 a test set (spanning April 15, 2020 to May 9, 2020). We fit the model parameters  $\psi$ ,  $\beta_{\text{base}}$ , and  
517  $p_0$ , as described in Methods M4.2, but only using the training set. We then evaluate the predictive  
518 accuracy of the resulting model on the test set. When running our models on the test set, we still use  
519 mobility data from the test period. Thus, this is an evaluation of whether the models can accurately  
520 predict the number of cases, given mobility data, in a time period that was not used for model  
521 calibration. Extended Data Figure 1 shows that our network model fits the out-of-sample case  
522 data fairly well, and that our model substantially outperforms alternate models that use aggregated  
523 mobility data (without a network) or do not use mobility data at all (Methods M3.4). Note that  
524 we only use this train/test split to evaluate out-of-sample model accuracy. All other results are  
525 generated using parameter sets that best fit the *entire* dataset, as described above.

### 526 **M5 Analysis details**

527 In this section, we include additional details about the experiments underlying the figures in the  
528 paper. We omit explanations for figures that are completely described in the main text.

529 **Counterfactuals of mobility reduction (Figure 2a, Tables S4–S5).** To simulate what would  
 530 have happened if we changed the magnitude or timing of mobility reduction, we modify the real  
 531 mobility networks from March 1–May 2, 2020, and then run our models on the hypothetical data.  
 532 In Figure 2a, we report the total number of people per 100k population ever infected (i.e., in the  
 533 exposed, infectious, and removed states) by the end of the simulation.

534 To simulate a smaller magnitude of mobility reduction, we interpolate between the mobility  
 535 network from the first week of simulation (March 1–7, 2020), which we use to represent typical  
 536 mobility levels, and the actual observed mobility network for each week. Let  $W^{(t)}$  represent the  
 537 observed visit matrix at the  $t$ -th hour of simulation, and let  $f(t) = t \bmod 168$  map  $t$  to its corre-  
 538 sponding hour in the first week of simulation, since there are 168 hours in a week. To represent the  
 539 scenario where people had committed to  $\alpha \in [0, 1]$  times the actual observed reduction in mobility,  
 540 we construct a visit matrix  $\tilde{W}_\alpha^{(t)}$  that is an  $\alpha$ -convex combination of  $W^{(t)}$  and  $W^{f(t)}$ ,

$$\tilde{W}_\alpha^{(t)} := \alpha W^{(t)} + (1 - \alpha) W^{f(t)}. \quad (26)$$

541 If  $\alpha$  is 1, then  $\tilde{W}_\alpha^{(t)} = W^{(t)}$ , and we use the actual observed mobility network for the simulation.  
 542 On the other hand, if  $\alpha = 0$ , then  $\tilde{W}_\alpha^{(t)} = W^{f(t)}$ , and we assume that people did not reduce  
 543 their mobility levels at all by looping the visit matrix for the first week of March throughout the  
 544 simulation. Any other  $\alpha \in [0, 1]$  interpolates between these two extremes.

545 To simulate changing the timing of mobility reduction, we shift the mobility network by  
 546  $d \in [-7, 7]$  days. Let  $T$  represent the last hour in our simulation (May 2, 2020, 11PM), let  
 547  $f(t) = t \bmod 168$  map  $t$  to its corresponding hour in the first week of simulation as above, and  
 548 similarly let  $g(t)$  map  $t$  to its corresponding hour in the last week of simulation (April 27–May 2,  
 549 2020). We construct the time-shifted visit matrix  $\tilde{W}_d^{(t)}$

$$\tilde{W}_d^{(t)} := \begin{cases} W^{(t-24d)} & \text{if } 0 \leq t - 24d \leq T, \\ W^{f(t-24d)} & \text{if } t - 24d < 0, \\ W^{g(t-24d)} & \text{otherwise.} \end{cases} \quad (27)$$

550 If  $d$  is positive, this corresponds to starting mobility reduction  $d$  days later; if we imagine time on  
 551 a horizontal line, this shifts the time series to the right by  $24d$  hours. However, doing so leaves  
 552 the first  $24d$  hours without visit data, so we fill it in by reusing visit data from the first week of

553 simulation. Likewise, if  $d$  is negative, this corresponds to starting mobility reduction  $d$  days earlier,  
 554 and we fill in the last  $24d$  hours with visit data from the last week of simulation.

555 **Distribution of predicted infections over POIs (Figure 2b, Extended Data Figure 2, Figure  
 556 S10).** We run our models on the observed mobility data from March 1–May 2, 2020 and record  
 557 the number of predicted infections that occur at each POI. Specifically, for each hour  $t$ , we compute  
 558 the number of expected infections that occur at each POI  $p_j$  by taking the number of susceptible  
 559 people who visit  $p_j$  in that hour multiplied by the POI infection rate  $\lambda_{p_j}^{(t)}$  (Equation (9)). In Figures  
 560 2b and S10, we sort the POIs by their total predicted number of infections (summed over hours)  
 561 and plot the cumulative distribution of infections over this ordering of POIs. In Extended Data  
 562 Figure 2, we select the POI categories that contribute the most to predicted infections and plot  
 563 the daily proportion of POI infections each category accounted for (summed over POIs within the  
 564 category) over time.

565 **Reducing mobility by capping maximum occupancy (Figure 2c, Extended Data Figure 3).**  
 566 We implemented two partial reopening strategies: one that uniformly reduced visits at POIs to a  
 567 fraction of full activity, and the other that “capped” each POI’s hourly visits to a fraction of the  
 568 POI’s maximum occupancy. For each reopening strategy, we started the simulation at March 1,  
 569 2020 and ran it until May 31, 2020, using the observed mobility network from March 1–April  
 570 30, 2020, and then using a hypothetical post-reopening mobility network from May 1–31, 2020,  
 571 corresponding to the projected impact of that reopening strategy. Because we only have observed  
 572 mobility data from March 1–May 2, 2020, we impute the missing mobility data up to May 31,  
 573 2020 by looping mobility data from the first week of March, as in the above analysis on the effect  
 574 of past reductions in mobility. Let  $T$  represent the last hour for which we have observed mobility  
 575 data (May 2, 2020, 11PM). To simplify notation, we define

$$h(t) := \begin{cases} t & \text{if } t < T, \\ f(t) & \text{otherwise,} \end{cases} \quad (28)$$

576 where, as above,  $f(t) = t \bmod 168$ . This function leaves  $t$  unchanged if there is observed mo-  
 577 bility data at time  $t$ , and otherwise maps  $t$  to the corresponding hour in the first week of our  
 578 simulation.

579 To simulate a reopening strategy that uniformly reduced visits to an  $\gamma$ -fraction of their origi-  
 580 nal level, where  $\gamma \in [0, 1]$ , we constructed the visit matrix

$$\tilde{W}_\gamma^{(t)} := \begin{cases} W^{h(t)} & \text{if } t < \tau, \\ \gamma W^{h(t)} & \text{otherwise,} \end{cases} \quad (29)$$

581 where  $\tau$  represents the first hour of reopening (May 1, 2020, 12AM). In other words, we use the  
 582 actual observed mobility network up until hour  $\tau$ , and then subsequently simulate an  $\gamma$ -fraction of  
 583 full mobility levels.

584 To simulate the reduced occupancy strategy, we first estimated the maximum occupancy  $M_{p_j}$   
 585 of each POI  $p_j$  as the maximum number of visits that it ever had in one hour, across all of March  
 586 1 to May 2, 2020. As in previous sections, let  $w_{ij}^{(t)}$  represent the  $i, j$ -th entry in the observed visit  
 587 matrix  $W^{(t)}$ , i.e., the number of people from CBG  $c_i$  who visited  $p_j$  in hour  $t$ , and let  $V_{p_j}^{(t)}$  represent  
 588 the total number of visitors to  $p_j$  in that hour, i.e.,  $\sum_i w_{ij}^{(t)}$ . We simulated capping at a  $\beta$ -fraction of  
 589 maximum occupancy, where  $\beta \in [0, 1]$ , by constructing the visit matrix  $\tilde{W}_\beta^{(t)}$  whose  $i, j$ -th entry is

$$\tilde{w}_{ij\beta}^{(t)} := \begin{cases} w_{ij}^{h(t)} & \text{if } t < \tau \text{ or } V_{p_j}^{(t)} \leq \beta M_{p_j}, \\ \frac{\beta M_{p_j}}{V_{p_j}^{(t)}} w_{ij}^{h(t)} & \text{otherwise.} \end{cases} \quad (30)$$

590 This corresponds to the following procedure: for each POI  $p_j$  and time  $t$ , we first check if  $t <$   
 591  $\tau$  (reopening has not started) or if  $V_{p_j}^{(t)} \leq \beta M_{p_j}$  (the total number of visits to  $p_j$  at time  $t$  is  
 592 below the allowed maximum  $\beta M_{p_j}$ ). If so, we leave  $w_{ij}^{h(t)}$  unchanged. Otherwise, we compute the  
 593 scaling factor  $\frac{\beta M_{p_j}}{V_{p_j}^{(t)}}$  that would reduce the total visits to  $p_j$  at time  $t$  down to the allowed maximum  
 594  $\beta M_{p_j}$ , and then scale down all visits from each CBG  $c_i$  to  $p_j$  proportionately. For both reopening  
 595 strategies, we calculate the predicted increase in cumulative incidence at the end of the reopening  
 596 period (May 31, 2020), compared to the start of the reopening period (May 1, 2020).

597 **Relative risk of reopening different categories of POIs (Figure 2d, Extended Data Figure 5,  
 598 Figure S11, Figures S15–S24).** We study separately reopening the 20 POI categories with the  
 599 most visits in SafeGraph data. In this analysis, we follow prior work<sup>28</sup> and do not study four cat-  
 600 egories: “Child Day Care Services” and “Elementary and Secondary Schools” (because children  
 601 under 13 are not well-tracked by SafeGraph); “Drinking Places (Alcoholic Beverages)” (because

602 SafeGraph seems to undercount these locations<sup>28</sup>) and “Nature Parks and Other Similar Institu-  
 603 tions” (because boundaries and therefore areas are not well-defined by SafeGraph). We also ex-  
 604 clude “General Medical and Surgical Hospitals” and “Other Airport Operations” (because hospi-  
 605 tals and air travel both involve many additional risk factors our model is not designed to capture).  
 606 We do not filter out these POIs during model fitting (i.e., we assume that people visit these POIs,  
 607 and that transmissions occur there) because including them still increases the proportion of over-  
 608 all mobility our dataset captures; we simply do not analyze these categories, because we wish to  
 609 be conservative and only focus on categories where we are most confident that we are capturing  
 610 transmission faithfully.

611 This reopening analysis is similar to the previous experiments on reducing maximum occu-  
 612 pancy vs. uniform reopening. As above, we set the reopening time  $\tau$  to May 1, 2020, 12AM. To  
 613 simulate reopening a POI category, we take the set of POIs in that category,  $\mathcal{V}$ , and set their activity  
 614 levels after reopening to that of the first week of March. For POIs not in the category  $\mathcal{V}$ , we keep  
 615 their activity levels after reopening the same, i.e., we simply repeat the activity levels of the last  
 616 week of our data (April 27–May 2, 2020): This gives us the visit matrix  $\tilde{W}^{(t)}$  with entries

$$\tilde{w}_{ij}^{(t)} := \begin{cases} w_{ij}^{(t)} & \text{if } t < \tau, \\ w_{ij}^{f(t)} & \text{if } t \geq \tau, p_j \in \mathcal{V} \\ w_{ij}^{g(t)} & \text{if } t \geq \tau, p_j \notin \mathcal{V}. \end{cases} \quad (31)$$

617 As in the above reopening analysis,  $f(t)$  maps  $t$  to the corresponding hour in the first week of  
 618 March, and  $g(t)$  maps  $t$  to the corresponding hour in the last week of our data. For each category,  
 619 we calculate the predicted difference between (1) the cumulative fraction of people who have been  
 620 infected by the end of the reopening period (May 31, 2020) and (2) the cumulative fraction of  
 621 people infected by May 31 had we not reopened the POI category (i.e., if we simply repeated  
 622 the activity levels of the last week of our data). This seeks to model the increase in cumulative  
 623 incidence by end of May from reopening the POI category. In Extended Data Figure 5 and Figures  
 624 S15–S24, the bottom right panel shows the predicted increase for the category as a whole, and  
 625 the bottom left panel shows the predicted increase *per POI* (i.e., the total increase divided by the  
 626 number of POIs in the category).

627 **Per-capita mobility (Figure 3d, Extended Data Figure 6, Figure S3).** Each group of CBGs  
 628 (e.g., the bottom income decile) comprises a set  $\mathcal{U}$  of CBGs that fit the corresponding crite-  
 629 ria. In Figure 3d and Extended Data Figure 6, we show the daily per-capita mobilities of dif-  
 630 ferent pairs of groups (broken down by income and by race). To measure the per-capita mo-  
 631 bility of a group on day  $d$ , we take the total number of visits made from those CBGs to any  
 632 POI,  $\sum_{c_i \in \mathcal{U}} \sum_{p_j \in \mathcal{P}} \sum_{t=24d}^{24d+23} w_{ij}^{(t)}$ , and divide it by the total population of the CBGs in the group,  
 633  $\sum_{c_i \in \mathcal{U}} N_{c_i}$ . In Figure S3, we show the total number of visits made by each group to each POI  
 634 category, accumulated over the entire data period (March 1–May 2, 2020) and then divided by the  
 635 total population of the group.

636 **Average predicted transmission rate of a POI category (Figure 3e, Tables 3–4).** We compute  
 637 the predicted average hourly transmission rate experienced by a group of CBGs  $\mathcal{U}$  at a POI category  
 638  $\mathcal{V}$  as

$$\bar{\beta}_{\mathcal{U}\mathcal{V}} := \frac{\sum_{c_i \in \mathcal{U}} \sum_{p_j \in \mathcal{V}} \sum_{t=1}^T w_{ij}^{(t)} \beta_{p_j}^{(t)}}{\sum_{c_i \in \mathcal{U}} \sum_{p_j \in \mathcal{V}} \sum_{t=1}^T w_{ij}^{(t)}}, \quad (32)$$

639 where, as above,  $\beta_{p_j}^{(t)}$  is the transmission rate at POI  $p_j$  in hour  $t$  (Equation (8)),  $w_{ij}^{(t)}$  is the number  
 640 of visitors from CBG  $c_i$  at POI  $p_j$  in hour  $t$ , and  $T$  is the last hour in our simulation. This represents  
 641 the expected transmission rate encountered during a visit by someone from a CBG in group  $\mathcal{U}$  to a  
 642 POI in category  $\mathcal{V}$ .

## Methods references

39. SafeGraph. Places Schema (2020). Available at <https://docs.safegraph.com/docs/places-schema>.
40. SafeGraph. Weekly Patterns (2020). Available at <https://docs.safegraph.com/docs/weekly-patterns>.
41. SafeGraph. Using SafeGraph Polygons to Estimate Point-Of-Interest Square Footage (2019). Available at <https://www.safegraph.com/blog/using-safegraph-polygons-to-estimate-point-of-interest-square-footage>.
42. SafeGraph. Guide to Points-of-Interest Data: POI Data FAQ (2020). Available at <https://www.safegraph.com/points-of-interest-poi-data-guide>.
43. SafeGraph. Social Distancing Metrics (2020). Available at <https://docs.safegraph.com/docs/social-distancing-metrics>.
44. Athey, S., Blei, D., Donnelly, R., Ruiz, F. & Schmidt, T. Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data. In *AEA Papers and Proceedings*, vol. 108, 64–67 (2018).
45. Chen, M. K. & Rohla, R. The effect of partisanship and political advertising on close family ties. *Science* **360**, 1020–1024 (2018).
46. Farboodi, M., Jarosch, G. & Shimer, R. Internal and external effects of social distancing in a pandemic (2020). Available at <https://nber.org/papers/w27059>.
47. Killeen, B. D. *et al.* A County-level Dataset for Informing the United States' Response to COVID-19 (2020). Available at <https://arxiv.org/abs/2004.00756>.
48. Allcott, H. *et al.* Polarization and public health: Partisan differences in social distancing during the Coronavirus pandemic (2020). Available at [nber.org/papers/w26946](https://nber.org/papers/w26946).
49. Google. COVID-19 community mobility reports (2020). Available at <https://google.com/covid19/mobility/>.
50. Athey, S., Ferguson, B., Gentzkow, M. & Schmidt, T. Experienced Segregation (2019). Available at <https://gsb.stanford.edu/faculty-research/working-papers/experienced-segregation>.
51. Squire, R. F. What about bias in the SafeGraph dataset? (2019). Available at <https://safegraph.com/blog/what-about-bias-in-the-safegraph-dataset>.
52. US Census. American Community Survey. Available at <https://census.gov/programs-surveys/acs>.
53. Bishop, Y. M., Fienberg, S. E. & Holland, P. W. Discrete multivariate analysis (1975).
54. Birkin, M. & Clarke, M. Synthesis—a synthetic spatial information system for urban and regional analysis: methods and examples. *Environment and planning A* **20**, 1645–1671 (1988).
55. Wong, D. W. The reliability of using the iterative proportional fitting procedure. *The Professional Geographer* **44**, 340–348 (1992).
56. Simpson, L. & Tranmer, M. Combining sample and census data in small area estimates: Iterative proportional fitting with standard software. *The Professional Geographer* **57**, 222–234 (2005).

57. Hu, H., Nigmatulina, K. & Eckhoff, P. The scaling of contact rates with population density for the infectious disease models. *Mathematical biosciences* **244**, 125–134 (2013).
58. Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases* **20**, 553 – 558 (2020).
59. Park, M., Cook, A. R., Lim, J. T., Sun, Y. & Dickens, B. L. A systematic review of COVID-19 epidemiology based on current evidence. *Journal of Clinical Medicine* **9**, 967 (2020).
60. Curmei, M., Ilyas, A., Evans, O. & Steinhardt, J. Estimating household transmission of sars-cov-2. *medRxiv* (2020). Available at <https://doi.org/10.1101/2020.05.23.20111559>.
61. Li, W. *et al.* The characteristics of household transmission of COVID-19. *Clinical Infectious Diseases* .
62. Gudbjartsson, D. F. *et al.* Spread of sars-cov-2 in the icelandic population. *New England Journal of Medicine* (2020).
63. Carey, B. & Glanz, J. Hidden Outbreaks Spread Through U.S. Cities Far Earlier Than Americans Knew, Estimates Say. *The New York Times* (2020). Available at <https://nytimes.com/2020/04/23/us/coronavirus-early-outbreaks-cities.html>.
64. Bommer, C. & Vollmer, S. Average detection rate of SARS-CoV-2 infections has improved since our last estimates but is still as low as nine percent on March 30th (2020). Available at <https://www.uni-goettingen.de/en/606540.html>.
65. Javan, E., Fox, S. J. & Meyers, L. A. The unseen and pervasive threat of COVID-19 throughout the US. *medRxiv* (2020). Available at <https://doi.org/10.1101/2020.04.06.20053561>.
66. Perkins, A. *et al.* Estimating unobserved SARS-CoV-2 infections in the United States. *medRxiv* (2020). Available at <https://doi.org/10.1101/2020.03.15.20036582>.
67. King, A. A., Domenech de Cellès, M., Magpantay, F. M. & Rohani, P. Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to ebola. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20150347 (2015).

**Acknowledgements.** The authors thank Yong-Yeol Ahn, Nic Fishman, Tatsunori Hashimoto, Percy Liang, Roni Rosenfeld, Jacob Steinhardt, Ryan Tibshirani, Johan Ugander, seminar participants, and our reviewers for support and helpful comments. We also thank Nick Singh, Ryan Fox Squire, Jessica Williams-Holt, Jonathan Wolf, Ruowei Yang, and others at SafeGraph for cell phone mobility data and helpful feedback. This research was supported by US National Science Foundation under OAC-1835598 (CINES), OAC-1934578 (HDR), CCF-1918940 (Expeditions), IIS-2030477 (RAPID), Stanford Data Science Initiative, Wu Tsai Neurosciences Institute, and Chan Zuckerberg Biohub. S.C. was supported by an NSF Fellowship. E.P. was supported by a Hertz Fellowship. P.W.K. was supported by the Facebook Fellowship Program. J.L. is a Chan Zuckerberg Biohub investigator.

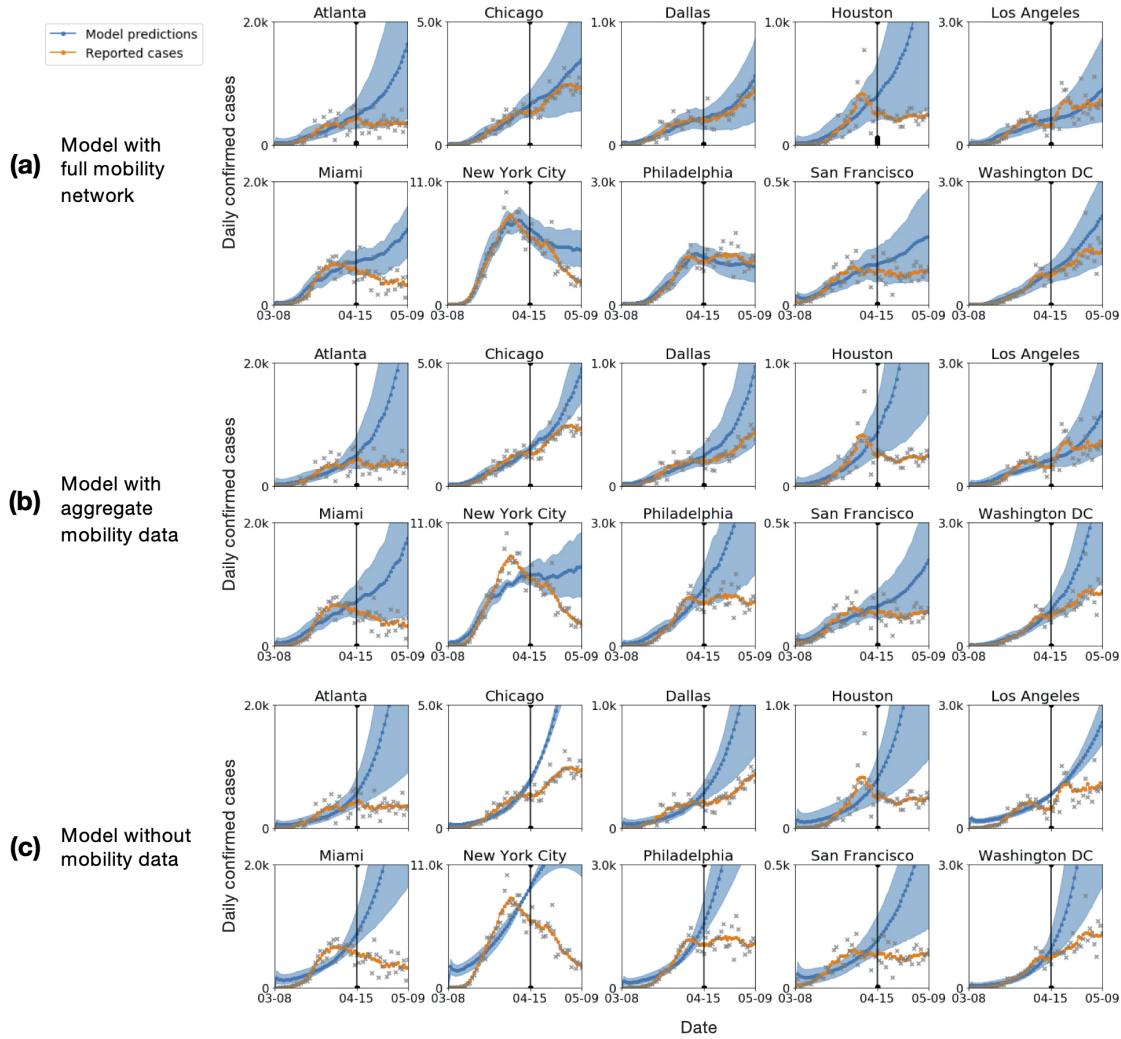
**Author Contributions.** S.C., E.P., and P.W.K. performed computational analysis. All authors jointly analyzed the results and wrote the paper.

**Author Information.** The authors declare no conflict of interest. Correspondence should be addressed to [jure@cs.stanford.edu](mailto:jure@cs.stanford.edu).

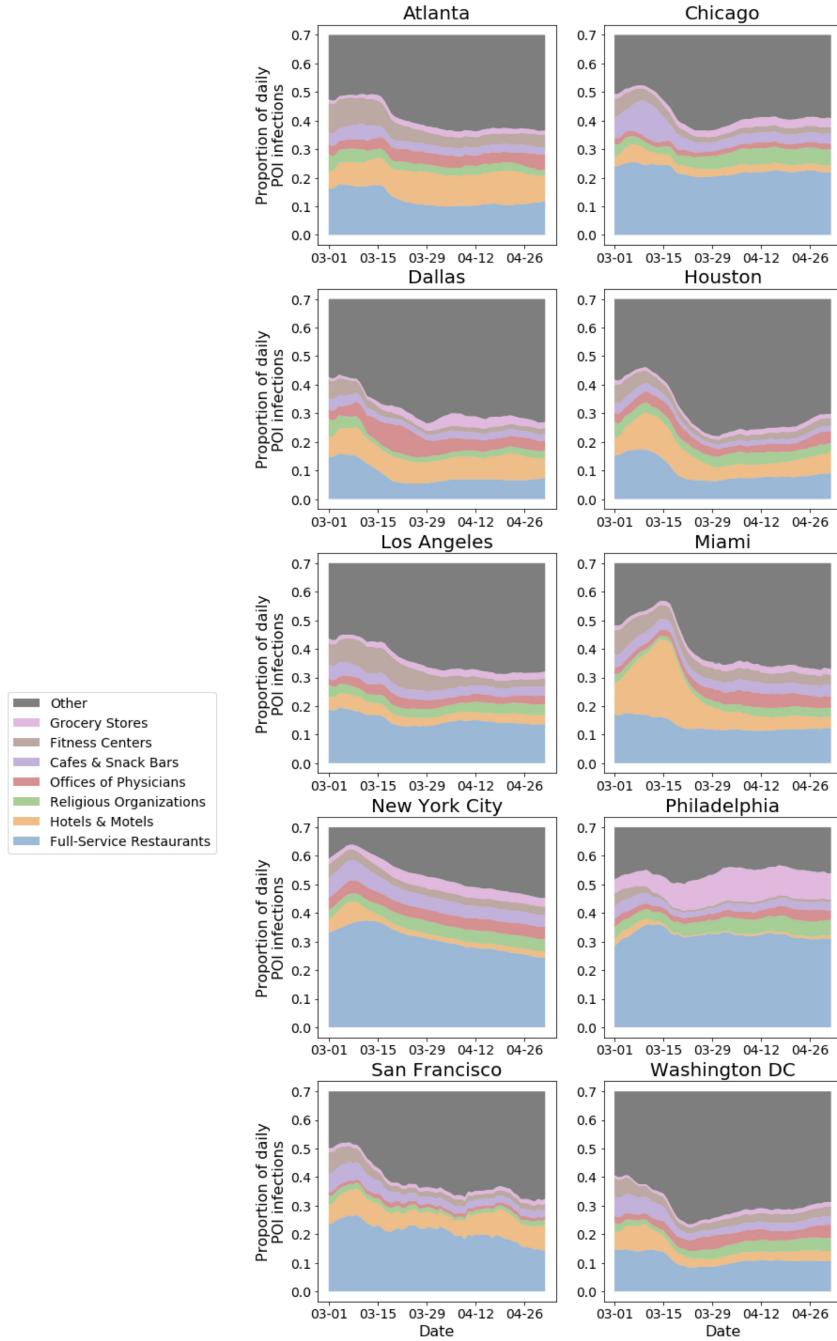
**Data Availability.** Census data, case and death counts from *The New York Times*, and Google mobility data are publicly available. Cell phone mobility data is freely available to researchers, non-profits, and governments through the SafeGraph COVID-19 Data Consortium.

**Code Availability.** Code is publicly available at <http://snap.stanford.edu/covid-mobility/>.

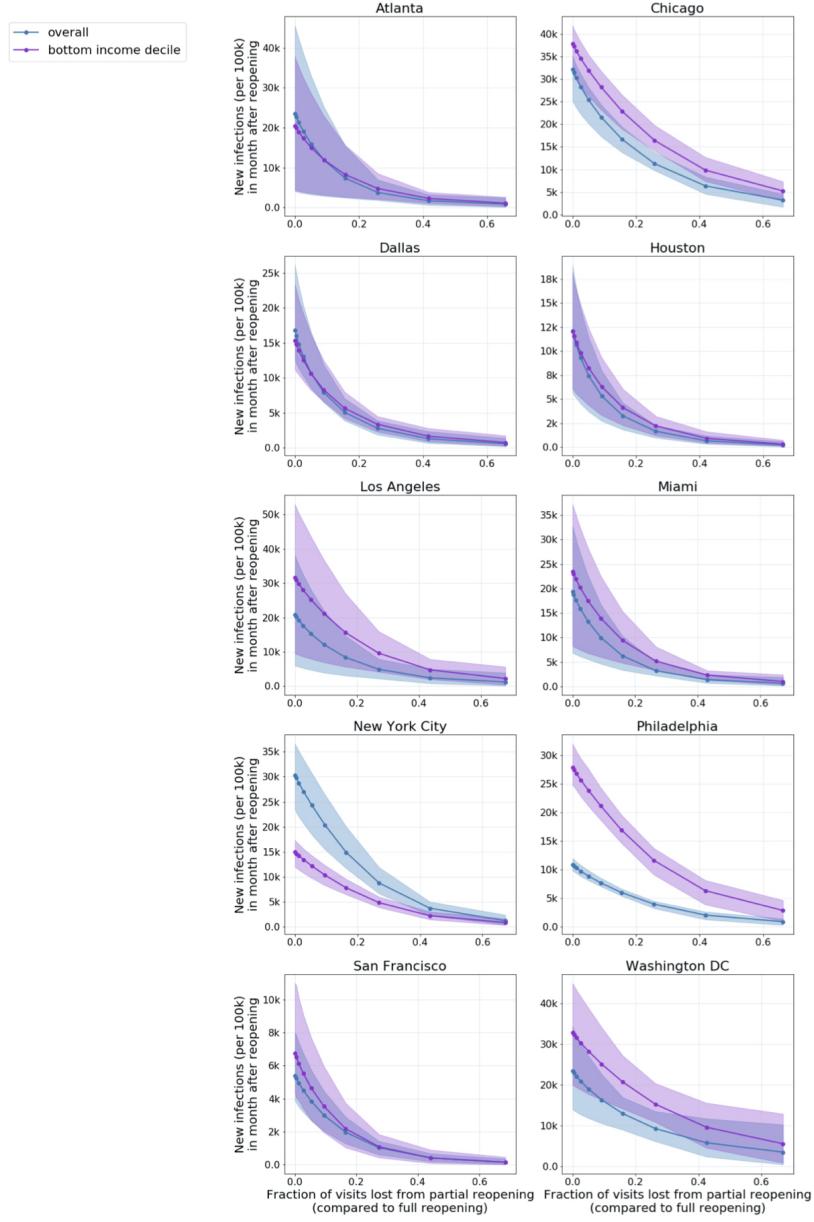
# Extended data



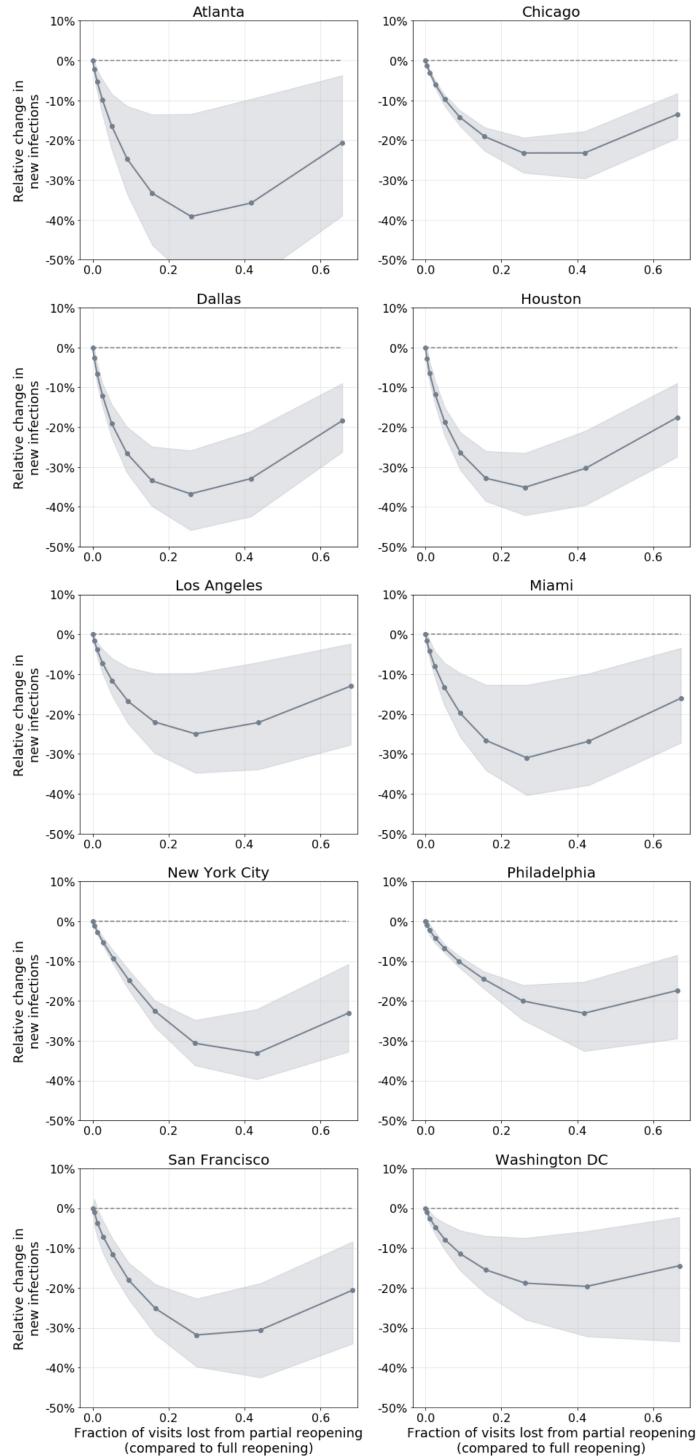
**Extended Data Figure 1:** Predicted (blue) and true (orange) daily case counts for **(a)** our model, which uses hourly mobility networks, **(b)** an SEIR model which uses hourly aggregated mobility data, and **(c)** a baseline SEIR model which does not use mobility data (see Methods M3.4 for details). Incorporating mobility information improves out-of-sample fit and having a network, instead of an aggregate measure, further improves fit: on average across metro areas, our best-fit network model's out-of-sample error (RMSE) was only 58% that of the best-fit aggregate mobility model. All three models are calibrated on observed case counts before April 15, 2020 (vertical black line). The grey x's represent the daily reported cases; since they tend to have great variability, we also show the smoothed weekly average (orange line). Shaded regions denote 2.5th and 97.5th percentiles across sampled parameters and stochastic realizations.



**Extended Data Figure 2:** Distribution of POI infections over time. We selected the POI categories that our models predicted contributed the most to infections, and plotted the predicted *proportion* of POI infections each category accounted for over time. Our model predicts time-dependent variation of where transmissions may have occurred. For example, Full-Service Restaurants (blue) and Fitness Centers (brown) contributed less to predicted infections over time, likely due to lockdown orders closing these POIs, while grocery stores remained steady or even grew in their predicted contribution, likely because they remained open as essential businesses. Hotels & Motels (yellow) also feature in these plots; most notably, the model predicts a peak in their contributed infections in Miami around mid-March — this would align with college spring break, with Miami as a popular vacation spot for students. The proportions are stacked in these plots, and the y-axes are truncated at 0.7 because every plot would only show “Other” from 0.7 to 1.0.

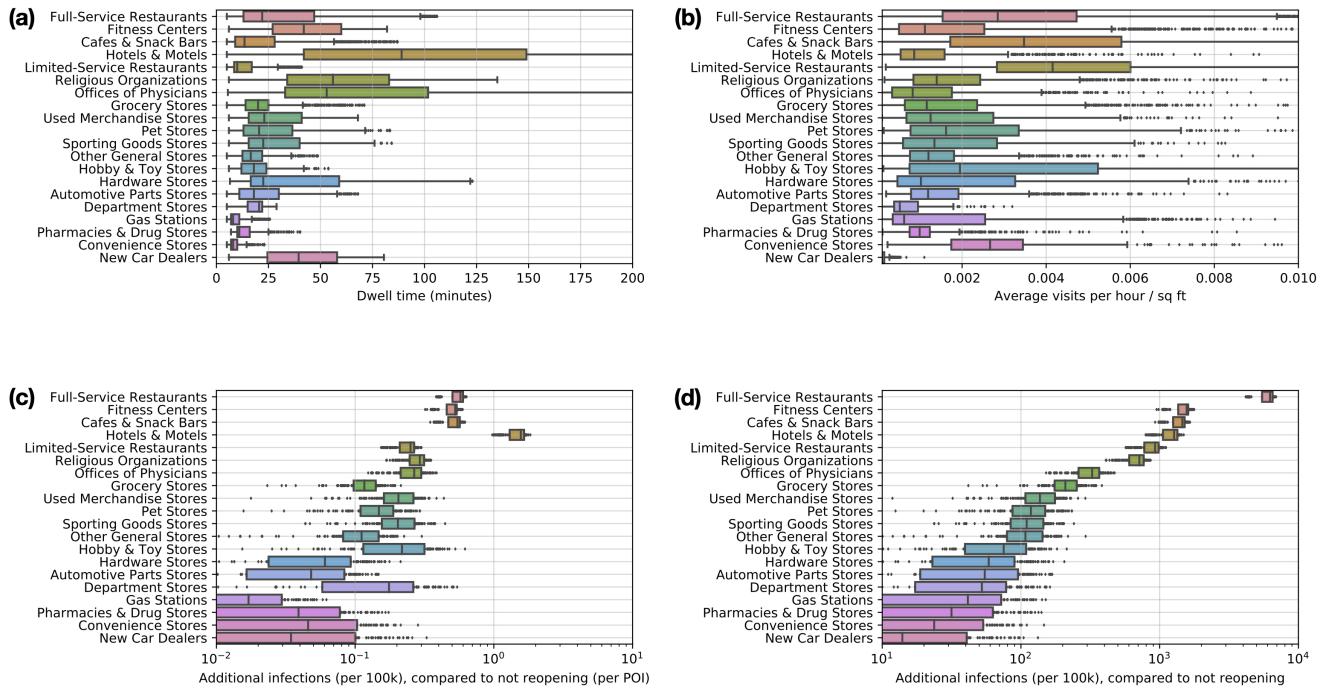


**Extended Data Figure 3:** Trade-off between number of new infections and visits lost, with different levels of reduced maximum occupancy reopening. We simulate reopening starting on May 1, 2020 and run the simulation until the end of the month. Each dot represents the level of occupancy reduction: e.g., capping visits at 50% of maximum occupancy. The y-coordinate represents the predicted number of new infections incurred after reopening (per 100k population) and the x-coordinate represents the fraction of visits lost from partial reopening compared to full reopening. Shaded regions denote 2.5th and 97.5th percentiles across parameter sets and stochastic realizations. In 4 metro areas, the predicted cost of new infections from reopening is roughly similar for lower-income CBGs and the overall population, but in 5 metro areas, the lower-income CBGs incur more predicted infections from reopening. Notably, New York City (NYC) is the only metro area where this trend is reversed; this is because the model predicts that such a high fraction—65% (95% CI, 62%-68%)—of lower-income CBGs in NYC had been infected before reopening that after reopening, only a minority of the lower-income population is still susceptible (in comparison, the second highest fraction infected before reopening was 31% (95% CI, 28%-35%) for Philadelphia, and the rest ranged from 1%-14%).

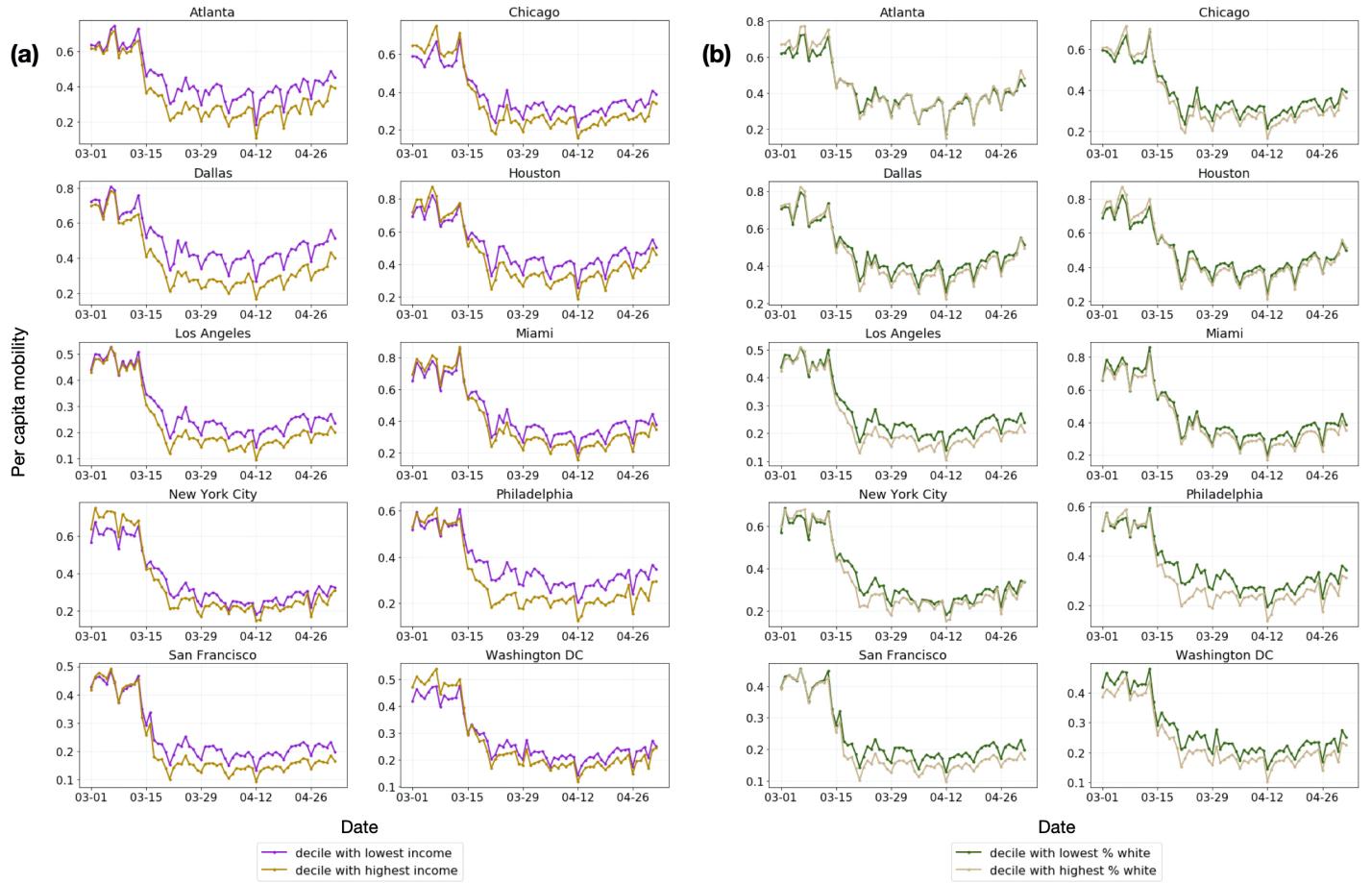


**Extended Data Figure 4:** Reduced occupancy versus uniform reduction reopening. Compared to partially reopening by uniformly reducing visits, the reduced occupancy strategy—which disproportionately targets high-risk POIs with sustained high occupancy—always results in a smaller predicted increase in infections for the same number of visits. The y-axis plots the relative difference between the predicted increase in cumulative infections (from May 1 to May 31) under the reduced occupancy strategy as compared to the uniform reduction strategy. The shaded regions denote the 2.5th and 97.5th percentiles over sampled parameters and stochastic realizations.

### All metro areas



**Extended Data Figure 5:** POI attributes in all 10 metro areas combined. The top two plots pool POIs from all metro areas and plot quantities from the mobility data, showing **(a)** the distribution of dwell time, and **(b)** the average number of hourly visitors divided by the area of the POI in square feet. Each point represents one POI; boxes depict the interquartile range across POIs, with data points outside the range individually shown. The bottom two plots pool across model realizations from all metro areas, and show model predictions for the increase in infections (per 100k population) from reopening a POI category: **(c)** per POI, and **(d)** for the category as a whole. Each point represents a model realization; boxes depict the interquartile range across realizations, with data points outside the range individually shown. Across MSAs, we model 552,758 POIs in total, and we sample 97 parameters and 30 stochastic realizations ( $N = 2,910$ ); see Table S6 for the number of sets per metro area. The boxes denote the interquartile range, with data points outside the range individually shown. Colors are used to distinguish the different POI categories, but do not have any additional meaning.



**Extended Data Figure 6:** Daily per-capita mobility over time, **(a)** comparing lower-income to higher-income CBGs and **(b)** comparing less white to more white CBGs. See Methods M5 for details.

Metro area	CBGs	POIs	Hourly edges	Total modeled pop	Total visits
Atlanta	3,130	39,411	540,166,727	7,455,619	27,669,692
Chicago	6,812	62,420	540,112,026	10,169,539	33,785,702
Dallas	4,877	52,999	752,998,455	9,353,561	37,298,053
Houston	3,345	49,622	609,766,288	7,621,541	32,943,613
Los Angeles	8,904	83,954	643,758,979	16,101,274	38,101,674
Miami	3,555	40,964	487,544,190	6,833,129	26,347,947
New York City	14,763	122,428	1,057,789,207	20,729,481	66,581,080
Philadelphia	4,565	37,951	304,697,220	6,759,058	19,551,138
San Francisco	2,943	28,713	161,575,167	5,137,800	10,728,090
Washington DC	4,051	34,296	312,620,619	7,740,276	17,898,324
<b>All metro areas combined</b>	<b>56,945</b>	<b>552,758</b>	<b>5,411,028,878</b>	<b>97,901,278</b>	<b>310,905,313</b>

**Extended Data Table 1:** Dataset summary statistics from March 1–May 2, 2020.

Param.	Description	Value (Source)
$\delta_E$	mean latency period	96 hours <sup>20,58</sup>
$\delta_I$	mean infectious period	84 hours <sup>20</sup>
$\delta_c$	period from infectious to confirmed	7 days <sup>20,66</sup>
$r_c$	percentage of cases which are detected	10% <sup>20,58,64–66</sup>
$\beta_{\text{base}}$	base CBG transmission rate	Variable (Estimated)
$N_{c_i}$	population size of CBG $c_i$	Variable (2018 US Census <sup>52</sup> )
$\psi$	scaling factor for POI transmission	Variable (Estimated)
$w_{ij}^{(t)}$	# visitors from CBG $c_i$ to POI $p_j$ at time $t$	Variable (SafeGraph)
$a_{p_j}$	area of POI $p_j$ in square feet	Variable (SafeGraph)
$p_0$	initial proportion of exposed population	Variable (Estimated)
$S_{c_i}^{(0)}$	initial susceptible population in CBG $c_i$	$(1 - p_0)N_{c_i}$
$E_{c_i}^{(0)}$	initial exposed population in CBG $c_i$	$p_0N_{c_i}$
$I_{c_i}^{(0)}$	initial infectious population in CBG $c_i$	0
$R_{c_i}^{(0)}$	initial removed population in CBG $c_i$	0

**Extended Data Table 2:** Model parameters. If the parameter has a fixed value, we specify it under **Value**; otherwise, we write “Variable” to indicate that it varies across CBG / POI / metro area.

Metro area	ATL	CHI	DAL	HOU	LA	MIA	NY	PHL	SF	DC	<b>Median</b>
Full-Service Restaurants	0.764	1.204	0.956	1.000	1.445	1.232	2.035	2.883	1.758	1.171	1.218
Limited-Service Restaurants	0.940	0.950	1.002	0.906	1.067	0.872	1.901	1.614	0.994	0.962	0.978
Other General Stores	0.782	1.083	0.957	0.729	0.760	0.894	1.218	1.312	1.045	0.950	0.954
Gas Stations	1.326	1.865	1.310	1.515	2.254	2.195	1.899	6.461	1.357	1.870	1.868
Fitness Centers	0.536	0.907	0.708	0.670	1.461	0.789	1.151	1.516	0.995	1.160	0.951
Grocery Stores	0.948	3.080	0.838	1.333	2.408	1.498	4.984	10.437	2.478	1.977	2.192
Cafes & Snack Bars	1.385	0.919	0.716	1.120	1.327	2.168	1.943	1.757	0.982	0.932	1.224
Hotels & Motels	1.228	1.200	0.814	0.804	1.229	1.134	1.260	1.993	1.199	1.346	1.214
Religious Organizations	1.546	1.763	0.956	0.919	1.746	1.464	1.756	1.736	1.515	1.852	1.641
Hardware Stores	3.938	3.340	1.575	2.111	1.333	0.939	3.553	6.716	4.202	13.560	3.446
Department Stores	1.132	1.230	0.978	0.911	1.083	1.431	1.667	0.976	0.867	1.042	1.062
Offices of Physicians	1.235	0.721	0.667	1.036	1.141	1.687	1.307	1.319	1.193	0.445	1.167
Pharmacies & Drug Stores	1.636	1.389	1.176	0.854	1.718	1.555	2.577	5.624	1.200	1.699	1.596
Sporting Goods Stores	0.936	1.540	1.129	0.812	1.168	0.700	1.253	1.161	0.826	2.777	1.145
Automotive Parts Stores	0.890	1.707	0.862	1.086	1.990	1.414	1.524	2.697	1.753	1.246	1.469
Used Merchandise Stores	0.993	0.931	1.000	1.315	1.017	1.074	1.352	1.668	1.587	0.814	1.046
Convenience Stores	1.208	0.932	1.613	0.647	0.838	0.824	1.736	2.322	1.086	1.428	1.147
Pet Stores	1.260	0.820	1.192	1.487	1.536	0.776	3.558	1.652	2.124	0.905	1.374
New Car Dealers	2.036	1.471	0.741	0.809	1.180	1.377	2.022	1.129	0.395	0.872	1.154
Hobby & Toy Stores	1.168	1.110	1.165	0.853	1.771	1.520	1.525	1.088	0.883	0.926	1.138
<b>Median</b>	1.188	1.202	0.968	0.915	1.330	1.305	1.746	1.702	1.196	1.166	

**Extended Data Table 3:** Predicted transmission rate disparities at each POI category between income groups. We report the ratio of the average predicted transmission rate encountered by visitors from CBGs in the bottom income decile to that for the top income decile. A ratio greater than 1 means that visitors from CBGs in the bottom income decile experienced higher (more dangerous) predicted transmission rates. See Methods M5 for details.

Metro area	ATL	CHI	DAL	HOU	LA	MIA	NY	PHL	SF	DC	Median
Full-Service Restaurants	0.802	1.354	0.981	0.965	1.065	1.167	2.418	2.661	1.223	1.013	1.116
Limited-Service Restaurants	0.940	1.144	1.028	0.940	0.820	0.919	2.136	1.523	0.799	1.346	0.984
Other General Stores	0.776	1.277	0.838	0.841	1.527	1.132	2.158	1.313	0.925	1.312	1.204
Gas Stations	1.402	1.891	1.389	1.190	1.336	1.857	1.818	2.286	2.321	1.316	1.610
Fitness Centers	0.607	1.167	0.670	0.831	0.780	1.066	1.447	1.977	1.103	1.205	1.084
Grocery Stores	0.589	3.664	0.613	1.195	2.386	0.950	5.864	13.705	2.243	2.262	2.252
Cafes & Snack Bars	1.308	1.104	0.845	0.840	0.976	2.619	1.767	2.456	1.045	0.867	1.074
Hotels & Motels	0.977	1.007	1.366	0.718	1.112	1.024	1.449	2.494	0.654	0.899	1.015
Religious Organizations	0.938	1.606	1.060	0.953	2.096	1.795	1.933	2.040	1.674	1.188	1.640
Hardware Stores	0.909	3.900	1.523	1.461	1.952	0.586	5.032	3.898	11.103	13.432	2.925
Department Stores	1.081	1.301	0.805	0.777	0.992	2.337	2.479	1.357	1.089	1.402	1.195
Offices of Physicians	0.894	1.323	1.006	1.415	0.898	1.117	1.652	2.073	0.694	1.911	1.220
Pharmacies & Drug Stores	0.888	1.376	0.930	0.732	1.538	1.674	3.315	3.366	1.135	1.715	1.457
Sporting Goods Stores	0.767	0.674	0.650	0.506	1.946	0.818	1.532	2.152	0.880	1.715	0.849
Automotive Parts Stores	1.049	1.479	1.010	1.353	2.998	2.657	1.740	3.387	1.646	0.601	1.562
Used Merchandise Stores	0.858	1.195	0.699	1.060	1.270	0.593	1.500	3.024	1.425	0.799	1.128
Convenience Stores	2.016	5.055	1.272	2.188	0.761	0.902	1.911	2.276	1.239	1.844	1.878
Pet Stores	0.925	1.624	0.724	1.465	1.506	0.881	2.715	10.182	1.568	2.408	1.537
New Car Dealers	1.008	1.398	0.812	0.736	0.942	0.998	1.977	0.866	0.772	0.383	0.904
Hobby & Toy Stores	2.569	0.853	0.628	0.979	1.373	1.388	2.237	0.825	0.864	1.286	1.132
<b>Median</b>	0.932	1.339	0.888	0.959	1.303	1.092	1.955	2.281	1.119	1.314	

**Extended Data Table 4:** Predicted transmission rate disparities at each POI category between racial groups. We report the ratio of the average predicted transmission rate encountered by visitors from CBGs with the lowest (bottom decile) proportion of white residents versus that for the top decile. A ratio greater than 1 means that visitors from CBGs in the bottom decile experienced higher (more dangerous) predicted transmission rates. See Methods M5 for details.