natureresearch
https://doi.org/10.1038/s41586-020-2923-3

Supplementary information

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

In the format provided by the authors and unedited

Contents

Supplementary Methods

1 Comparison of Google and SafeGraph mobility data

To assess the reliability of the SafeGraph datasets, we measured the correlation between mobility trends according to SafeGraph versus Google.[1] Google provides a high-level picture of mobility changes around the world for several categories of places, such as grocery stores or restaurants. We analyzed three of the categories defined by Google: *Retail & recreation* (e.g., restaurants, shopping centers, movie theaters), *Grocery & pharmacy* (e.g., grocery stores, farmers markets, pharmacies), and *Residential* (i.e. places of residence). We omitted *Transit stations* because they are not well-covered by SafeGraph POIs, *Parks* because SafeGraph informed us that parks are sometimes inaccurately classified in their data (e.g., other POIs are categorized as parks), and *Workplaces* because we do not model whether people are at work. To analyze the *Retail & recreation* and *Grocery & pharmacy* categories, we used POI visits in the SafeGraph Patterns datasets, identifying POIs in each category based on their 6-digit North American Industry Classification System (NAICS) codes (Table S2). For the *Residential* category, we used SafeGraph Social Distancing Metrics, which provides daily counts of the number of people in each CBG who stayed at home for the entire day.

For each US region and category, Google tracks how the number of visits to the category has changed over the last few months, compared to baseline levels of activity before SARS-CoV-2. To set this baseline, they compute the median number of visits to the category for each day of the week, over a 5-week span from January 3–February 6, 2020. For a given day of interest, they then compute the relative change in number of visits seen on this day compared to the baseline for the corresponding day of week. We replicated this procedure on SafeGraph data, and compared the results to Google's trends for Washington DC and 14 states that appear in the metro areas that we model. For each region and category, we measured the Pearson correlation between the relative change in number of visits according to Google versus Safegraph, from March 1–May 2, 2020. Across the 15 regions, we found that the median Pearson correlation was 0.96 for *Retail & recreation*, 0.79 for *Grocery & pharmacy*, and 0.88 for *Residential*. As an illustrative example, we visualize the results for New York state in Figure S1, and provide a full table of results for every state in Table S3. The high correlations demonstrate that the SafeGraph and Google mobility datasets agree well on the timing and directional changes of mobility over this time period, providing a validation of the reliability of SafeGraph data.

2 Sensitivity analyses and robustness checks

2.1 Time-varying base transmission rate

In our model, we assumed that $\lambda_{c_i}^{(t)}$, the base rate of infection in CBG c_i, was equal to a constant base transmission rate β_{base} multiplied by the infectious fraction of c_i at time t (Equation 11). We conducted a sensitivity analysis where we assumed, instead, a time-varying base transmission rate that incorporated an additional factor $\hat{p}_{c_i}^{(t)}$, the estimated proportion of people at home in c_i at time t. Under this modified model, $\lambda_{c_i}^{(t)}$ became

$$\lambda_{c_i}^{(t)} = \underbrace{\beta_{\text{base}} \cdot \hat{p}_{c_i}^{(t)}}_{\text{base transmission rate}} \cdot \frac{I_{c_i}^{(t)}}{N_{c_i}}. \tag{1}$$

During the time period we simulate, there was a dramatic increase in the number of people staying at home. As a result, we guessed that there might be a corresponding increase in the frequency and duration of interactions within households. By scaling the base transmission rate with the proportion of people staying home, this sensitivity analysis explored the possibility that the transmission rate outside of POIs might have increased together with the number of people staying home.

We estimated the daily proportion of people staying at home in each CBG by computing `completely_home_device_count/device_count` from SafeGraph's Social Distancing Metrics. We ran the same procedure (Methods M4) to calibrate this modified model, and evaluated its ability to fit incident daily cases. We found that the modified model did not yield significant improvement over our original model; for example, taking the median over metro areas, the fitted modified model's RMSE was only 2% smaller than that of the original model. When choosing parameter sets based on fit to the training set (March 1–April 14, 2020), the modified model's out-of-sample RMSE was 8% smaller than that of the original model, but it only outperformed the original model on 6 out of 10 metro areas, so the improvement was inconsistent. Finally, Figure S6 visualizes the predictions of the fitted modified model compared to the original model; again, they are very similar to each other and have approximately equal fit to the reported cases. Thus, because the modified model did not significantly improve model fit, we opted to use a fixed base transmission rate to keep the model simple.

2.2 Modifying the parametric form for POI transmission rates

In our model, the transmission rate at a POI p_j at hour t,

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

depends on two key ingredients: $d_{p_j}^2$, which reflects how much time visitors spend there, and $V_{p_j}^{(t)}/a_{p_j}$, which reflects the density (number of visitors per sq ft) of the POI in that hour. These assumptions are based on prior expectations that a visit is more dangerous for a susceptible individual if they spend more time there and/or if the place is more crowded. To assess empirically the role that each of these two terms play, we compared our transmission rate formula to two perturbed versions of it: one that removed the dwell time term, and another that removed the density term. For each of these formulas, we computed the risk of visiting a POI category as the average transmission rate of the category, with the rate of each POI weighted by the proportion of category visits that went to that POI. Then, we evaluated whether the relative risks predicted by each formula concorded with the rankings of POI categories proposed by independent epidemiological experts.[2,3] In our evaluations, we included all of the categories that we analyzed (i.e., the 20 categories with the most visits in SafeGraph data; see Methods M5) that overlapped with categories described in the external rankings. To compare against Emanuel et al.[2], we also converted their categorical groupings into numerical score, i.e., "Low" $\rightarrow 1$, "Low/Medium" $\rightarrow 2$, etc., up to "High" $\rightarrow 5$. Sims et al.[3] already provided numerical ratings so we did not have to perform a conversion.

As shown in Figure S7, we find that the predicted relative risks match external sources best when we use our original parametric form that accounts for both dwell time and density: restaurants, cafes, religious organizations, and gyms are among the most dangerous, while grocery stores and retail (e.g., clothing stores) are less dangerous. However, when we assume only dwell time matters and remove the density term, we see unrealistic changes in the ranking: e.g., restaurants drop close to grocery stores, despite both sets of experts deeming them far apart in terms of risk. When we assume only density matters and remove dwell time, we also see unrealistic changes: e.g., limited-service restaurants are predicted to be far riskier than full-service restaurants, and gyms and religious organizations are no longer predicted as risky, which contradicts both of our sources. These findings demonstrate that both factors — the dwell time and density — are important toward faithfully modeling transmission at POIs, since the predictions become less realistic when either factor is taken out.

2.3 Stochastic sampling of confirmed cases

To predict confirmed case counts from the SEIR trajectories, our model assumes a fixed proportion of infected people are confirmed after a fixed confirmation delay (Methods M4.2). Since these proportions and delays are quite variable in reality, we conducted a sensitivity analysis where instead we tried stochastically sampling the number of confirmed cases and the confirmation delay. For each day d, we first computed $N^{(d)}_{E_{c_i} \to I_{c_i}}$, the number of people who became infectious on this day; we then sampled from $\text{Binom}\big(N^{(d)}_{E_{c_i} \to I_{c_i}}, r_c\big)$ to get the number of confirmed cases that would result from this group of infections. For each case that was to be confirmed, we drew its confirmation delay (i.e., delay from becoming infectious to being confirmed) from distributions fitted on empirical line-list data: either Gamma$(1.85, 3.57)$[4] or Exp(6.1).[5]

We found that our model predictions barely changed when we sampled case trajectories stochastically using either delay distribution, as opposed to assuming a fixed confirmation rate and delay (Figure S8). However, an advantage of our fixed method is that it allows us to predict confirmed cases up to δ_c (i.e., 7) days after the last day of simulation, whereas we cannot do the same when we sample confirmed cases and delays stochastically. This is because, if delays are stochastic, predicting the number of confirmed cases on, for example, the 5th day after the simulation ends depends on the number of newly infectious individuals every day before and including that day, but since the simulation ended days before, the model would not have sufficient information to make the prediction. On the other hand, the fixed method simply translates and scales the newly infectious curve, so we can predict the number of confirmed cases 5 days after the simulation ends, since it only depends on the number of newly infectious individuals 2 days before the end of simulation. Due to this advantage, we opted to use the fixed method.

2.4 Model calibration metrics

For each metro area, our model fitting procedure selects all parameter sets that achieve an RMSE within 20% of the best-fit parameter set's RMSE (Methods M4.3). As a final sensitivity analysis, we tested three alternative model fitting procedures that used different metrics to decide when to accept or reject a parameter set. For each procedure, we evaluated the correlation between its ranking of parameter sets and our original ranking. We recomputed our downstream analyses using the fitted models outputted by each procedure and verified that our key results on superspreader POIs (Figure S10), the effects of reopening (Figure S11), and predicted group disparities (Figure S12) all remained similar.

Poisson likelihood model. Our model calibration procedure, which uses RMSE to assess fit, implicitly assumes that error in the number of observed cases is drawn from a normal (Gaussian) distribution. As a sensitivity analysis, we tested a Poisson error model instead, using negative log-likelihood as a measure of fit, and using the same 20% threshold. We note that the homoscedastic Gaussian model will likely prioritize fitting parts of the case trajectory that have higher case counts, whereas a Poisson model will comparatively prioritize fitting parts of the case trajectory with lower case counts. We found that ranking models via Poisson likelihood was consistent with ranking models using RMSE (both computed on daily incident cases, as described above): the median Spearman correlation over metro areas between models ranked by Poisson likelihood vs. RMSE was 0.97.

Model acceptance threshold. As described above, we set the acceptance threshold for model calibration (i.e., the threshold for rejection sampling in the Approximate Bayesian Computation framework) to 20% of the RMSE of the best-fit model. We selected this threshold because beyond that point, model fit qualitatively deteriorated based on inspection of the case trajectories. As a sensitivity analysis, we selected a different threshold (10%), while still using RMSE as the error metric. We do not report correlations here, since the reduced threshold simply corresponded to selecting a subset of the originally chosen parameters.

Fitting to deaths. In addition to the number of confirmed cases, the NYT data also contains the daily reported number of deaths due to COVID-19 by county. As an additional test, we calibrated our models to fit this death data instead of case data. To estimate the number of deaths N_{deaths}, we use a similar process as for the number of cases N_{cases}, except that we replace r_c with $r_d = 0.66\%$, the estimated infection fatality rate for COVID-19,[6] and δ_c with $\delta_d = 432$ hours (18 days), the number of days between becoming infectious and dying[6] (Extended Data Table 2 provides references for all parameters). This gives

$$N^{(d)}_{\text{deaths}} = r_d \cdot \sum_{i=1}^{m} \sum_{\tau=24(d-1)+1-\delta_d}^{24d-\delta_d} N^{(\tau)}_{E_{c_i} \to I_{c_i}}. \tag{3}$$

Because we assume that deaths occur $\delta_d = 18$ days after individuals become infectious, we compared with NYT death data starting on March 19, 2020 (18 days after our simulation begins). Figure S9 shows that the calibrated models can also fit the trends in the death counts well. Ranking models using RMSE on deaths was consistent with ranking models using RMSE on cases, with a median Spearman correlation over metro areas of 0.99, and as with the above sensitivity analyses (changing the likelihood model and the acceptance threshold), we found that our key results remained similar.

2.5 Parameter identifiability

We assess the identifiability of the fitted model parameters ψ, β_{base}, and p_0 as follows. First, we verify that the model-fitting procedure is able to recover the true parameters when fit on simulated data for which the true parameters are known. For each metro area, we simulate daily case counts using the best-fit parameters for that metro area (i.e., those with the minimum RMSE to daily case counts, as reported in Table S6). We then run our grid search fitting procedure on the simulated case counts. For all 10 metro areas, as Figure S13 illustrates, the parameters in our grid search that obtain the lowest RMSE on the simulated data are always the true parameters that were used to generate that data. In other words, for each metro area, we correctly estimate and recover the true parameters on the simulated data.

As a further assessment of model identifiability, in Figure S14 we plot RMSE on true (not simulated) daily case counts (that is, the metric used to perform model calibration) as a function of model parameters β_{base} and ψ. (We take the minimum RMSE over values of p_0 so the plots can be visualized in two dimensions.) As these plots illustrate, β_{base} and ψ are correlated, which is unsurprising because they scale the growth of infections at CBGs and POIs respectively. This correlation results in uncertainty in the parameter estimates in some metro areas. Throughout our analyses, we reflect this uncertainty by aggregating results from all parameter settings which achieve an RMSE within 20% of the best-fit model for each metro area.

3 Estimating the mobility network from SafeGraph data

As we discussed in Methods M2, the central technical challenge in constructing our mobility network is estimating the network weights $W^{(t)} = \{w_{ij}^{(t)}\}$ from SafeGraph data, since this visit matrix is not directly available from the data. In this section, we describe our estimation process, which utilizes the iterative proportional fitting procedure[7] (IPFP) to estimate a matrix for each metro area and each hour from March 1, 2020 to May 2, 2020.

Quantities from SafeGraph data. To begin, we describe the quantities from SafeGraph data that we use to make this estimation.

- The estimated visit matrix $\hat{W}^{(r)}$ aggregated for the month r, where we use r instead of t to denote time periods longer than an hour. This is taken from the Patterns dataset, and is aggregated at a monthly level. To account for non-uniform sampling from different CBGs, we weight the number of SafeGraph visitors from each CBG by the ratio of the CBG population and the number of SafeGraph devices with homes in that CBG.[8]

- $\hat{V}_{p_j}^{(t)}$: The number of visitors recorded in POI p_j at hour t. This is taken from the Weekly Patterns v1 dataset.

- $\hat{h}_{c_i}^{(t)}$: The estimated fraction of people in CBG c_i who left their home in day $\lfloor t/24 \rfloor$. This is derived by taking $1 - (\texttt{completely\_home\_device\_count}/\texttt{device\_count})$. These are daily (instead of hourly) metrics in the Social Distancing Metrics dataset.

- $\hat{\delta}_{p_j}$: The median length of a visit to a POI p_j. We estimate this by averaging over the weekly values in the `median_dwell` field in the Patterns datasets in March and April 2020. $\hat{\delta}_{p_j}$ is measured to minute-level resolution and expressed in units of hours, e.g., $\hat{\delta}_{p_j} = 1.5$ means a median visit time of 1.5 hours = 90 minutes.

3.1 Data preprocessing and dwell time computation

Hourly visits. The raw SafeGraph data records the number of visitors that *newly arrive* at each POI p_j at each hour. However, $\hat{V}_{p_j}^{(t)}$ above represents the number of visitors that *are present* at a POI in an hour t; these visitors may have arrived prior to t. The aggregate visit matrix $\hat{W}^{(r)}$, as well as the visit matrix $W^{(t)}$ used in our model, are defined similarly. To compute these quantities from the raw data, we make two assumptions: first, that every visitor to p_j stays for exactly $\hat{\delta}_{p_j}$ hours, where $\hat{\delta}_{p_j}$ is the median length of a visit to p_j, and second, that a visitor who newly arrives in an hour t is equally likely to arrive at any time from $[t, t+1)$. With these assumptions, we can convert the number of visitor arrivals in each hour into the expected number of visitors present at each hour: for example, if $\hat{\delta}_{p_j} = 1.5$ hours, then we assume that a visitor who arrives sometime during an hour t will also be present in hour $t+1$ and be present half the time, on expectation, in hour $t+2$. Note that under our definition, visits are still counted even if a visitor does not stay for the entire hour. For example, a visitor that arrives at 9:30am and leaves at 10:10am will be counted as two visits: one during the 9-10am hour and one during the 10-11am hour.

The dwell time correction factor d_{p_j}. To estimate the mean occupancy at each POI p_j in an hour t, we multiply the expected number of visitors present at p_j in hour t by the dwell time correction factor d_{p_j}, which measures the expected fraction of an hour that a visitor present at p_j at any hour will spend there. In other words, conditioned on a visitor being at p_j at some time within an hour t, d_{p_j} is the expected fraction of the hour t that the visitor physically spends at p_j. The same two assumptions above allow us to calculate d_{p_j}: since each visitor stays for exactly $\hat{\delta}_{p_j}$ hours, and on average is counted as being present in $\hat{\delta}_{p_j} + 1$ different hours, we have $d_{p_j} = \hat{\delta}_{p_j}/(\hat{\delta}_{p_j} + 1)$.

Truncating outliers. As described in Methods M3, our model necessarily makes parametric assumptions about the relationship between POI characteristics (area, hourly visitors, and dwell time) and transmission rate at the POI; these assumptions may fail to hold for POIs which are outliers, particularly if SafeGraph data has errors. We mitigate this concern by truncating extreme values for POI characteristics to prevent data errors from unduly influencing our conclusions. Specifically, we truncate each POI's area (i.e., square footage) to the 5th and 95th percentile of areas in the POI's category; for every hour, we truncate the number of visitor arrivals for each POI to its category's 95th percentile of visitor arrivals in that hour; and we truncate each POI's median dwell time to its category's 90th percentile of median dwell times in that period.

3.2 Estimating the visit matrix $W^{(t)}$

Overview. We estimate the visit matrix $W^{(t)} = \{w_{ij}^{(t)}\}$, which captures the number of visitors from CBG c_i to POI p_j at each hour t from March 1, 2020 to May 2, 2020, through the iterative proportional fitting procedure (IPFP).[7] The idea is as follows:

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

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

3. We then use IPFP to estimate an hourly visit matrix $W^{(t)}$ that is consistent with the hourly marginals $U^{(t)}$ and $V^{(t)}$ but otherwise "as similar as possible" to the distribution of visits in the aggregate visit matrix $\bar{W}$. Here, similarity is defined in terms of Kullback-Leibler divergence; we provide a precise definition below.

Estimating the aggregate visit matrix $\bar{W}$. The estimated monthly visit matrices $\hat{W}^{(r)}$ are typically noisy and sparse: SafeGraph only matches a subset of visitors to POIs to their home CBGs, either for privacy reasons (if there are too few visitors from the given CBG) or because they are unable to link the visitor to a home CBG.[9] To mitigate this issue, we aggregate these visit matrices, which are available at the monthly level, over the $R = 14$ months from January 2019 to February 2020:

$$\bar{W} := \frac{1}{R} \sum_r \hat{W}^{(r)}. \tag{4}$$

Each entry $\bar{w}_{ij}$ of $\bar{W}$ represents the estimated number of visitors from CBG c_i that are present at POI p_j in an hour, averaged over each hour. After March 2020, SafeGraph reports the visit matrices $\hat{W}^{(r)}$ on a weekly level in the Weekly Patterns v1 dataset. However, due to inconsistencies in the way SafeGraph processes the weekly versus monthly matrices, we only use the monthly matrices up until February 2020.

Estimating the POI marginals $V^{(t)}$. We estimate the POI marginals $V^{(t)} \in \mathbb{R}^n$, whose j-th element $V_{p_j}^{(t)}$ represents our estimate of the number of visitors at POI p_j (from any CBG) at time t. The number of visitors recorded at POI p_j at hour t in the SafeGraph data, $\hat{V}_{p_j}^{(t)}$, is an underestimate because the SafeGraph data only covers on a fraction of the overall population. To correct for this, we follow Benzell et al.[10] and compute our final estimate of the visitors at POI p_j in time t as

$$V_{p_j}^{(t)} = \frac{\text{US population}}{\text{total number of SafeGraph devices}} \cdot \hat{V}_{p_j}^{(t)}. \tag{5}$$

This correction factor is approximately 7, using population data from the most recent 1-year ACS (2018).

Estimating the CBG marginals $U^{(t)}$. Next, we estimate the CBG marginals $U^{(t)} \in \mathbb{R}^m$. Here, the i-th element $U_{c_i}^{(t)}$ represents our estimate of the number of visitors in CBG c_i who are out visiting a POI at time t. We first use the POI marginals $V^{(t)}$ to calculate the total number of people who are out visiting any POI from any CBG at time t,

$$N_{\text{POIs}}^{(t)} := \sum_{j=1}^{n} V_{p_j}^{(t)}, \tag{6}$$

where n is the total number of POIs. Next, we estimate the number of people from each CBG c_i who are not at home at time t as $\hat{h}_{c_i}^{(t)} N_{c_i}$; recall that N_{c_i} is the total population of c_i, as derived from US Census data. In general, the total number of people who are not at home in their CBGs, $\sum_{i=1}^{m} \hat{h}_{c_i}^{(t)} N_{c_i}$, will not be equal to $N_{\text{POIs}}^{(t)}$, the number of people who are out visiting any POI. This discrepancy occurs for several reasons: for example, some people might have left their homes to travel to places that SafeGraph does not track, SafeGraph might not have been able to determine the home CBG of a POI visitor, etc.

To correct for this discrepancy, we assume that the relative proportions of POI visitors coming from each CBG follows the relative proportions of people who are not at home in each CBG. We thus estimate $U_{c_i}^{(t)}$ by apportioning the $N_{\text{POIs}}^{(t)}$ total POI visitors at time t according to the proportion of people who are not at home in each CBG c_i at time t:

$$U_{c_i}^{(t)} := N_{\text{POIs}}^{(t)} \cdot \frac{\hat{h}_{c_i}^{(t)} N_{c_i}}{\sum_{k=1}^{m} \hat{h}_{c_k}^{(t)} N_{c_k}}, \tag{7}$$

This construction ensures that the POI and CBG marginals match, i.e., $N_{\text{POIs}}^{(t)} = \sum_{j=1}^{n} V_{p_j}^{(t)} = \sum_{i=1}^{m} U_{c_i}^{(t)}$.

Iterative proportional fitting procedure (IPFP). IPFP is a classic statistical method[7] for adjusting joint distributions to match pre-specified marginal distributions, and it is also known in the literature as biproportional fitting, the RAS algorithm, or raking.[11] In the social sciences, it has been widely used to infer the characteristics of local subpopulations (e.g., within each CBG) from aggregate data.[12–14]

We estimate the visit matrix $W^{(t)}$ by running IPFP on the aggregate visit matrix $\bar{W}$, the CBG marginals $U^{(t)}$, and the POI marginals $V^{(t)}$ constructed above. Our goal is to construct a non-negative matrix $W^{(t)} \in \mathbb{R}^{m \times n}$ whose rows sum up to the CBG marginals $U^{(t)}$,

$$U_{c_i}^{(t)} = \sum_{j=1}^{n} w_{ij}^{(t)}, \tag{8}$$

and whose columns sum up to the POI marginals $V_{p_j}^{(t)}$,

$$V_{p_j}^{(t)} = \sum_{i=1}^{m} w_{ij}^{(t)}, \tag{9}$$

but whose distribution is otherwise "as similar as possible", in the sense of Kullback-Leibler divergence, to the distribution over visits induced by the aggregate visit matrix $\bar{W}$.

IPFP is an iterative algorithm that alternates between scaling each row to match the row (CBG) marginals $U^{(t)}$ and scaling each column to match the column (POI) marginals $V^{(t)}$. We provide pseudocode in Algorithm 1. For each value of t used in our simulation, we run IPFP sep-

Algorithm 1: Iterative proportional fitting procedure to estimate visit matrix $W^{(t)}$

Input: Aggregate visits $\bar{W} \in \mathbb{R}^{m \times n}$
CBG marginals $U^{(t)} \in \mathbb{R}^m$; POI marginals $V^{(t)} \in \mathbb{R}^n$
Number of iterations τ_{max}

Initialize $W^{(t,0)} = \bar{W}$
for $\tau = 1, \ldots, \tau_{max}$ **do**
  if τ *is odd* **then**
    for $i = 1, \ldots, m$ **do**
      $\alpha_i \leftarrow U_{c_i}^{(t)} / \sum_{j=1}^{n} w_{ij}^{(t,\tau-1)}$ `// Compute scaling factor for row` i
      $W_{i,:}^{(t,\tau)} \leftarrow \alpha_i * W_{i,:}^{(t,\tau-1)}$ `// Rescale row` i
    end
  else if τ *is even* **then**
    for $j = 1, \ldots, n$ **do**
      $\beta_j \leftarrow V_{p_j}^{(t)} / \sum_{i=1}^{m} w_{ij}^{(t,\tau-1)}$ `// Compute scaling factor for col` j
      $W_{:,j}^{(t,\tau)} \leftarrow \beta_i * W_{:,j}^{(t,\tau-1)}$ `// Rescale col` j
    end
  end
end
$W^{(t)} \leftarrow W^{(t,\tau_{\text{max}})}$

arately for $\tau_{\text{max}} = 100$ iterations. Note that IPFP is invariant to scaling the absolute magnitude of the entries in $\bar{W}$, since the total number of visits it returns is fixed by the sum of the marginals; instead, its output depends only on the distribution over visits in $\bar{W}$. The notion of similarity invoked above has a maximum likelihood interpretation: if IPFP converges, then it returns a visit matrix $W^{(t)}$ whose induced distribution minimizes the Kullback-Leibler divergence to the distribution induced by $\bar{W}$.[15]

Convergence of IPFP. For completeness, we briefly review the convergence properties of IPFP. Consider the L_1-error function

$$E^{(t,\tau)} := \underbrace{\sum_i \left| U_{c_i}^{(t)} - \sum_j w_{ij}^{(t)} \right|}_{\text{Error in row marginals}} + \underbrace{\sum_j \left| V_{p_j}^{(t)} - \sum_i w_{ij}^{(t)} \right|}_{\text{Error in column marginals}}, \tag{10}$$

which sums up the errors in the row (CBG) and column (POI) marginals of the visit matrix $W^{(t,\tau)}$ from the τ-th iteration of IPFP. Each iteration of IPFP monotonically reduces this L_1-error $E^{(t,\tau)}$, i.e., $E^{(t,\tau)} \geq E^{(t,\tau+1)}$ for all $\tau \geq 0$.[16] In other words, the row and column sums of $W^{(t,\tau)}$ (which is initialized as $W^{(t,0)} = \bar{W}$) progressively get closer to (or technically, no further from) the target marginals as the iteration number τ increases. Moreover, IPFP maintains the cross-product ratios of the aggregate matrix $\bar{W}$, i.e.,

$$\frac{w_{ij}^{(t,\tau)} w_{k\ell}^{(t,\tau)}}{w_{i\ell}^{(t,\tau)} w_{kj}^{(t,\tau)}} = \frac{\bar{w}_{ij}\bar{w}_{k\ell}}{\bar{w}_{i\ell}\bar{w}_{kj}} \tag{11}$$

for all matrix entries indexed by i, j, k, ℓ, for all t, and for all iterations τ.

IPFP converges to a unique solution, in the sense that $W^{(t)} = \lim_{\tau\to\infty} W^{(t,\tau)}$, if there exists a matrix $W^{(t)}$ that fits the row and column marginals while maintaining the sparsity pattern (i.e., location of zeroes) of $\bar{W}$.[16] If IPFP converges, then the L_1-error also converges to 0 as $\tau \to \infty$,[16] and $W^{(t)}$ is the maximum likelihood solution in the following sense. For a visit matrix $W = \{w_{ij}\}$, let P_W represent a multinomial distribution over the mn entries of W with probability proportional to w_{ij}, and define $\mathcal{U}^{(t)} \subseteq \mathbb{R}_{+}^{m\times n}$ and $\mathcal{V}^{(t)} \subseteq \mathbb{R}^{m\times n}$ as the set of non-negative matrices whose row and column marginals match $U^{(t)}$ and $V^{(t)}$ respectively. Then, if IPFP converges,

$$W^{(t)} = \underset{W\in\mathcal{U}^{(t)}\cap\mathcal{V}^{(t)}}{\arg\min} \ \mathrm{KL}\left(P_W \| P_{\bar{W}}\right), \tag{12}$$

where $\mathrm{KL}\left(p\|q\right)$ is the Kullback-Leibler divergence $\mathrm{KL}\left(p\|q\right) = \mathbb{E}_p\left[\log \frac{p(x)}{q(x)}\right]$. In other words, IPFP returns a visit matrix $W^{(t)}$ whose induced distribution $P_{W^{(t)}}$ is the I-projection of the aggregate visit distribution $P_{\bar{W}}$ on the set of distributions with compatible row and column marginals.[15] In fact, IPFP can be viewed as an alternating sequence of I-projections onto the row marginals and I-projections onto the column marginals.[15,17]

However, in our setting, IPFP typically does not return a unique solution and instead oscillates between two accumulation points, one that fits the row marginals and another that fits the column marginals.[17] This is because $\bar{W}$ is highly sparse (there is no recorded interaction between most CBGs and POIs), so the marginals are sometimes impossible to reconcile. For example, suppose there is some CBG c_i and POI p_j such that $\bar{w}_{ij}$ is the only non-zero entry in the i-th row and j-th column of $\bar{W}$, i.e., visitors from c_i only travel to p_j and conversely visitors from p_j are all from c_i. Then, if $U_{c_i}^{(t)} \neq V_{p_j}^{(t)}$, there does not exist any solution $W^{(t)}$ such that $U_{c_i}^{(t)} = V_{p_j}^{(t)} = w_{ij}^{(t)}$. Note that in this scenario, IPFP still monotonically decreases the L_1-error.[16]

In our implementation (Algorithm 1), we take $\tau_{\text{max}} = 100$, so IPFP ends by fitting the column (POI) marginals. This ensures that our visit matrix $W^{(t)}$ is fully compatible with the POI marginals

$V^{(t)}$, i.e.,

$$V_{p_j}^{(t)} = \sum_{i=1}^{m} w_{ij}^{(t)}, \tag{13}$$

while still minimizing the L_1-error $E^{(t,\tau)}$ with respect to the CBG marginals $U^{(t)}$. Empirically, we find that $\tau_{\text{max}} = 100$ iterations of IPFP are sufficient to converge to this oscillatory regime.

Supplementary Discussion

4 Plausibility of predicted racial/socioeconomic disparities

To assess the plausibility of the model's predicted disparities in infection rates, we compared the model's predicted racial disparities to observed racial disparities in mortality rates. (Data on socioeconomic disparities in mortality was not systematically available on a national level.) The model's predicted racial disparities are generally of the same magnitude as reported racial disparities in mortality rates—for example, the overall reported black mortality rate is $2.4\times$ higher than the white mortality rate,[18] which is similar to the median racial disparity across metro areas of $3.0\times$ that our model predicts (Main Figure 3b). However, we note that this is an imperfect comparison because many factors besides mobility contribute to racial disparities in death rates.

In addition, we observed that our model predicted unusually large socioeconomic and racial disparities in infection rates in the Philadelphia metro area. To understand why the model predicted such large disparities, we inspected the mobility factors discussed in the main text; namely, how much each group was able to reduce their mobility, and whether disadvantaged groups encountered higher transmission rates at POIs. First, we found that higher-income CBGs and more white CBGs in Philadelphia were able to reduce their mobility substantially more than lower-income CBGs and less white CBGs, respectively (Extended Data Figure 6). While these trends were true for every metro area, the gap between income groups and racial groups was especially noticeable for Philadelphia. The other key to Philadelphia's outlier status lay in the comparison of predicted transmission rates. Generally, we found that individuals from lower-income and less white CBGs tended to visit POIs with higher predicted transmission rates (Extended Data Tables 3 and 4). This was particularly true for Philadelphia; in 19 out of 20 POI categories, individuals from lower-income CBGs in Philadelphia encountered higher predicted transmission rates than individuals from high-income CBGs, and CBGs with the lowest percentage of white residents encountered higher predicted transmission rates than the CBGs with the highest percentage of white residents in 18 out of 20 categories. The predicted transmission rates encountered by individuals from lower-income CBGs in Philadelphia are often dramatically higher than those encountered by higher-income CBGs; for example, up to $10.4\times$ higher for grocery stores. Digging deeper, this is because the average grocery store visited by lower-income CBGs has $5.3\times$ the number of hourly visitors per square foot, and visitors tend to stay 86% longer. Furthermore, Philadelphia's large discrepancy in density between lower-income and higher-income POIs in SafeGraph data is consistent with US Census data, which shows that the discrepancy in *population* density between lower- and higher-income CBGs is larger in Philadelphia than in any of the other metro areas that we examine. In

Philadelphia, CBGs in the bottom income decile have a population density $8.2\times$ those in the top income decile, a considerably larger disparity than the overall median across metro areas ($3.3\times$) or the next-highest CBG ($4.5\times$).

Since there are many other factors contributing to disparity that we do not model, we do not place too much weight on our model's prediction that Philadelphia's disparities will be larger than those of other cities. However, we consider this a valuable finding in terms of Philadelphia's mobility patterns, suggesting that mobility may play an especially strong role in driving socioeconomic and racial infection disparities in this metro area, and we encourage policy-makers to be aware of how differences in mobility patterns may exacerbate the disproportionate impact of SARS-CoV-2 on disadvantaged groups.

5 Model limitations

In this section, we discuss limitations in the dataset and model which are relevant to interpreting our results. The cell phone mobility dataset we use has limitations: it does not cover all populations (e.g., prisoners, children under 13, or adults without smartphones), does not contain all POIs (e.g., nursing homes are undercovered, and we exclude schools and hospitals from our analysis of POI category risks), and cannot capture sub-CBG heterogeneity in demographics. Individuals may also be double-counted in the dataset if they carry multiple cell phones. While the dataset allows us to illuminate mobility-related mechanisms which contribute to racial and socioeconomic disparities, these disparities are also driven by differences our dataset cannot capture (e.g., public transit use, or working at a restaurant as opposed to dining there) as well as non-mobility-related factors including differences in household size, access to healthcare, and comorbidities. These limitations notwithstanding, cell phone mobility data in general and SafeGraph data in particular have been instrumental and widely used in modeling SARS-CoV-2 spread.[10,19–26]

Our model itself is parsimonious, and does not include such relevant features as asymptomatic transmission; variation in household size; travel and seeding between metro areas; differentials in susceptibility due to pre-existing conditions or access to care; age-related variation in mortality rates or susceptibility (e.g., for modeling transmission at elementary and secondary schools); various time-varying transmission-reducing behaviors (e.g., hand-washing, mask-wearing, and holding events in outdoor spaces); and some POI-specific risk factors (e.g., ventilation). Although our model recovers case trajectories and known infection disparities even without incorporating these features, we caution that this predictive accuracy does not mean that our predictions should be interpreted in a narrow causal sense. Because certain types of POIs or subpopulations may disproportionately select for certain types of omitted processes, our findings on the relative risks of different POIs should be interpreted with due caution, and the potential public health benefits of restricting access to POIs should always be assessed in conjunction with the short-run and long-run economic impacts of doing so. 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 because they tend to visit smaller, denser POIs and because they have not been able to reduce mobility by as much (likely in part because they cannot as easily work from home[27])—to hold robustly.

Supplementary Tables

Category	% visits	% POIs
Full-Service Restaurants	14.82%	10.86%
Limited-Service Restaurants	8.08%	3.69%
Elementary and Secondary Schools	6.36%	3.06%
Other General Stores	5.97%	1.37%
Gas Stations	4.56%	2.94%
Fitness Centers	4.55%	2.98%
Grocery Stores	4.16%	2.17%
Cafes & Snack Bars	4.01%	2.70%
Hotels & Motels	2.93%	1.57%
Religious Organizations	2.31%	5.04%
Parks & Similar Institutions	1.93%	2.31%
Hardware Stores	1.79%	1.87%
Department Stores	1.78%	0.32%
Child Day Care Services	1.71%	2.76%
Offices of Physicians	1.63%	4.02%
Pharmacies & Drug Stores	1.54%	0.95%
Sporting Goods Stores	1.16%	1.05%
Automotive Parts Stores	1.16%	1.80%
Used Merchandise Stores	1.15%	1.01%
Colleges & Universities	1.12%	0.44%
Convenience Stores	1.09%	0.66%
Pet Stores	0.93%	0.85%
New Car Dealers	0.73%	0.43%
Hobby & Toy Stores	0.73%	0.36%
Offices of Dentists	0.70%	2.67%
Commercial Banking	0.70%	2.05%
Gift Stores	0.69%	0.57%
Liquor Stores	0.61%	0.82%
Women's Clothing Stores	0.59%	1.00%
Home Health Care Services	0.55%	1.02%
Furniture Stores	0.53%	0.89%
Electronics Stores	0.51%	0.72%
Used Car Dealers	0.50%	1.08%
Book Stores	0.49%	0.32%
Musical Instrument Stores	0.49%	0.50%
Optical Goods Stores	0.47%	0.76%
Family Clothing Stores	0.46%	0.49%
Car Repair Shops	0.41%	1.83%
Offices of Mental Health Practitioners	0.41%	1.05%
Tobacco Stores	0.41%	0.31%
Office Supplies	0.40%	0.33%
Beauty Salons	0.39%	1.58%
Paint and Wallpaper Stores	0.38%	0.56%
Other Gas Stations	0.37%	0.20%
Sports Teams and Clubs	0.37%	0.03%
Cosmetics & Beauty Stores	0.36%	0.71%
Jewelry Stores	0.34%	0.60%
Junior Colleges	0.34%	0.07%
Sewing & Piece Goods Stores	0.34%	0.39%
Senior Homes	0.34%	0.41%
Libraries & Archives	0.3%	0.3%

Table S1: The 50 POI categories accounting for the largest fraction of visits in the full SafeGraph dataset. Collectively they account for 88% of POI visits and 76% of POIs.

Google category	Google description	NAICS categories
Retail & recreation	Restaurants	Full-Service Restaurants
	Cafes	Limited-Service Restaurants
	Shopping centers	Snack and Nonalcoholic Beverage Bars
	Theme parks	Drinking Places (Alcoholic Beverages)
	Museums	Malls, Amusement and Theme Parks
	Libraries	Museums, Libraries and Archives
	Movie theaters	Motion Picture Theaters (except Drive-Ins)
Grocery & pharmacy	Grocery markets	Supermarkets and Other Grocery (except Convenience) Stores
	Food warehouses	Food (Health) Supplement Stores
	Farmers markets	Fish and Seafood Markets
	Specialty food shops	All Other Specialty Food Stores
	Drug stores	Pharmacies and Drug Stores
	Pharmacies	

Table S2: Mapping of Google mobility data categories to NAICS categories. Google descriptions taken from `https://www.google.com/covid19/mobility/data_documentation.html`.

State	Retail & recreation	Grocery & pharmacy	Residential
California	0.947	0.834	0.876
Delaware	0.957	0.847	0.856
Florida	0.963	0.814	0.885
Georgia	0.948	0.682	0.868
Illinois	0.964	0.710	0.899
Indiana	0.956	0.741	0.877
Maryland	0.956	0.825	0.886
New Jersey	0.951	0.720	0.935
New York	0.958	0.763	0.909
Pennsylvania	0.971	0.850	0.875
Texas	0.965	0.789	0.886
Virginia	0.967	0.840	0.877
Washington, DC	0.959	0.889	0.780
West Virginia	0.960	0.740	0.814
Wisconsin	0.967	0.783	0.886
Median	0.959	0.789	0.877

Table S3: Pearson correlations between the Google and SafeGraph mobility timeseries. We report correlations over the period of March 1–May 2, 2020 for the 15 states that we model. See SI Methods 1 for details.

Metro area	7 days earlier	3 days earlier	3 days later	7 days later
Atlanta	0.586 (0.397, 0.834)	0.803 (0.639, 0.956)	1.359 (1.075, 1.741)	1.981 (1.189, 2.761)
Chicago	0.641 (0.563, 0.711)	0.848 (0.769, 0.933)	1.226 (1.143, 1.365)	1.542 (1.446, 1.639)
Dallas	0.642 (0.495, 0.782)	0.855 (0.693, 1.013)	1.298 (1.09, 1.577)	1.722 (1.487, 1.966)
Houston	0.656 (0.500, 0.812)	0.848 (0.663, 1.021)	1.288 (1.079, 1.541)	1.731 (1.493, 2.064)
Los Angeles	0.608 (0.407, 0.848)	0.816 (0.639, 0.984)	1.265 (1.041, 1.554)	1.692 (1.216, 2.137)
Miami	0.576 (0.424, 0.795)	0.792 (0.669, 0.919)	1.317 (1.117, 1.559)	1.856 (1.281, 2.27)
New York City	0.818 (0.795, 0.856)	0.909 (0.890, 0.927)	1.113 (1.094, 1.133)	1.27 (1.246, 1.307)
Philadelphia	0.799 (0.731, 0.868)	0.916 (0.823, 1.005)	1.12 (1.031, 1.206)	1.287 (1.246, 1.351)
San Francisco	0.609 (0.408, 0.798)	0.815 (0.666, 1.012)	1.271 (1.048, 1.527)	1.689 (1.452, 2.029)
Washington DC	0.671 (0.447, 0.879)	0.848 (0.627, 1.045)	1.207 (0.959, 1.586)	1.488 (1.158, 1.789)

Table S4: Effects of shifting past mobility reduction earlier or later. We report the expected ratio of the number of infections predicted under the counterfactual to the number of infections predicted using observed mobility data; a ratio lower than 1 means that fewer predicted infections occurred under the counterfactual. The numbers in parentheses indicate the 2.5th and 97.5th percentiles across sampled parameters and stochastic realizations. See Methods M5 for details.

Metro area	0%	25%	50%
Atlanta	16.593 (3.088, 30.532)	7.714 (1.73, 15.833)	2.265 (1.17, 3.673)
Chicago	6.202 (5.2, 7.088)	3.329 (2.761, 3.759)	1.587 (1.421, 1.704)
Dallas	18.026 (10.361, 27.273)	5.908 (3.75, 8.857)	1.87 (1.532, 2.349)
Houston	18.964 (11.949, 32.755)	5.725 (3.761, 9.233)	1.659 (1.362, 2.109)
Los Angeles	12.926 (3.15, 24.207)	5.097 (1.779, 9.721)	1.665 (1.176, 2.309)
Miami	10.781 (3.382, 15.935)	4.85 (1.886, 7.405)	1.777 (1.208, 2.3)
New York City	2.037 (1.902, 2.174)	1.73 (1.603, 1.811)	1.333 (1.258, 1.389)
Philadelphia	2.976 (2.734, 3.39)	1.894 (1.747, 2.137)	1.211 (1.141, 1.305)
San Francisco	9.743 (7.089, 15.596)	4.282 (3.124, 6.781)	1.714 (1.427, 2.255)
Washington DC	5.85 (2.329, 9.713)	3.032 (1.541, 4.646)	1.509 (1.132, 1.959)

Table S5: Scaling the magnitude of past mobility reduction. Each column represents a counterfactual scenario where the magnitude of mobility reduction is only a some percentage of the observed mobility reduction, i.e., 0% corresponds to no mobility reduction, and 100% corresponds to the real, observed level of mobility reduction. We report the expected ratio of the number of infections predicted under the counterfactual to the number of infections predicted using observed mobility data; a ratio lower than 1 means that fewer infections occurred under the counterfactual. The numbers in parentheses indicate the 2.5th and 97.5th percentiles across sampled parameters and stochastic realizations. See Methods M5 for details.

Metro area	**# sets**	β_{base}	ψ	$\mathbf{p_0}$
Atlanta	16	0.004 (0.001, 0.014)	2388 (515, 3325)	$5 \times 10^{-4}(1 \times 10^{-4}, 2 \times 10^{-3})$
Chicago	4	0.009 (0.006, 0.011)	1764 (1139, 2076)	$2 \times 10^{-4}(2 \times 10^{-4}, 5 \times 10^{-4})$
Dallas	5	0.009 (0.004, 0.011)	1452 (1139, 2388)	$2 \times 10^{-4}(1 \times 10^{-4}, 2 \times 10^{-4})$
Houston	8	0.001 (0.001, 0.009)	2076 (1139, 2076)	$2 \times 10^{-4}(1 \times 10^{-4}, 5 \times 10^{-4})$
Los Angeles	25	0.006 (0.001, 0.016)	2076 (515, 3637)	$2 \times 10^{-4}(2 \times 10^{-5}, 1 \times 10^{-3})$
Miami	7	0.001 (0.001, 0.011)	2388 (515, 2388)	$2 \times 10^{-4}(2 \times 10^{-4}, 2 \times 10^{-3})$
New York City	7	0.001 (0.001, 0.009)	2700 (1452, 3013)	$1 \times 10^{-4}(5 \times 10^{-5}, 1 \times 10^{-3})$
Philadelphia	3	0.009 (0.001, 0.009)	827 (827, 1452)	$5 \times 10^{-4}(1 \times 10^{-4}, 5 \times 10^{-4})$
San Francisco	5	0.006 (0.001, 0.009)	1139 (827, 1764)	$5 \times 10^{-4}(2 \times 10^{-4}, 1 \times 10^{-3})$
Washington DC	17	0.016 (0.001, 0.019)	515 (515, 3949)	$5 \times 10^{-4}(2 \times 10^{-5}, 5 \times 10^{-4})$

Table S6: Estimated model parameters in each metro area. # sets counts the number of parameter sets that are within 20% of the RMSE of the best-fit parameter set, as described in Methods M4. For each of β_{base} (which scales the transmission rates at CBGs), ψ (which scales the transmission rates at POIs), and p_0 (the initial proportion of infected individuals), we show the best-fit parameter set and, in parentheses, the corresponding minimum and maximum within the 20% threshold.

Supplementary Figures

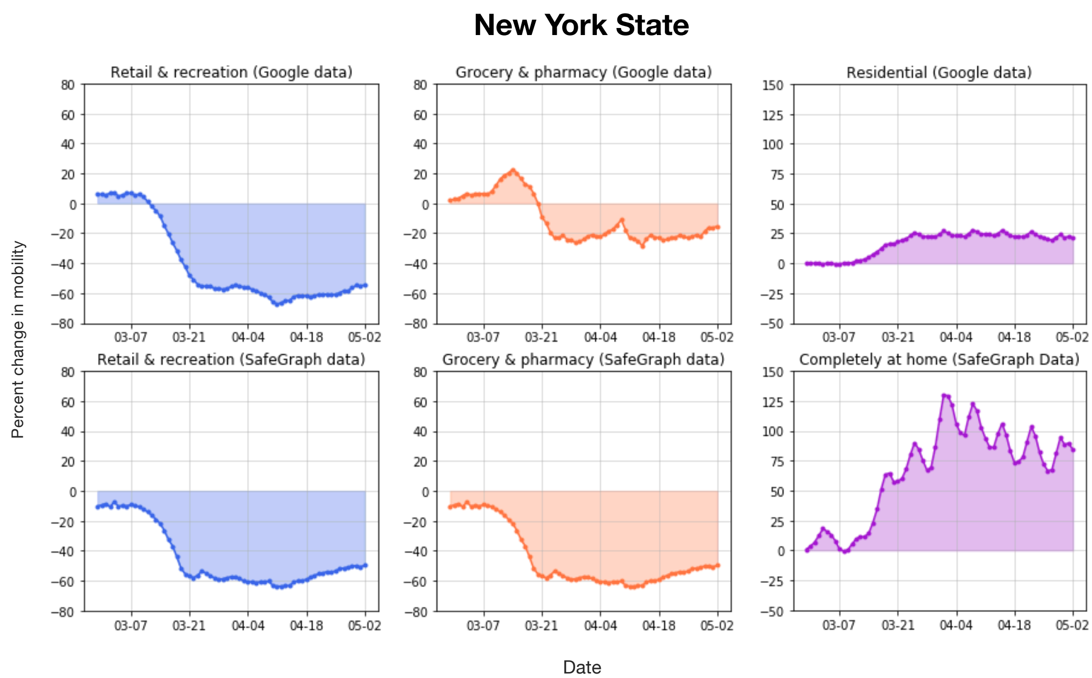

Figure S1: Google versus SafeGraph mobility trends for New York state. The x-axis is the same across plots, showing the date from March 1–May 2, 2020. The y-axis represents percent change in mobility levels compared to baseline activity in January and February 2020. For the categories from left to right, the Pearson correlations between the datasets in New York state are 0.96, 0.76, and 0.91. See SI Methods 1 for details.

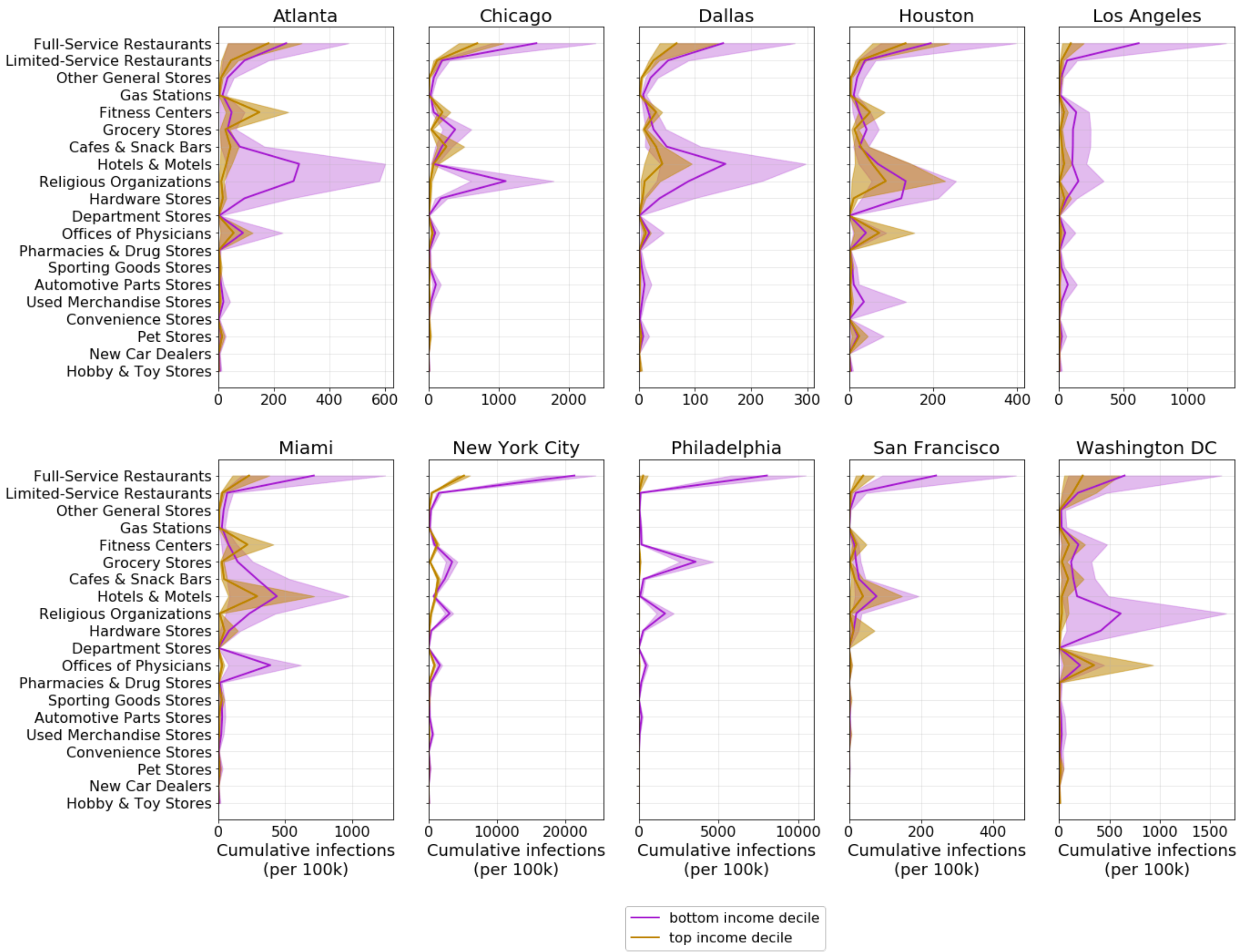

Figure S2: For each POI category, we plot the predicted cumulative number of infections (per 100k population) that occurred at that category for CBGs in the bottom- (purple) and top- (gold) income deciles. Shaded regions denote 2.5th and 97.5th percentiles across sampled parameters and stochastic realizations. See Methods M5 for details.

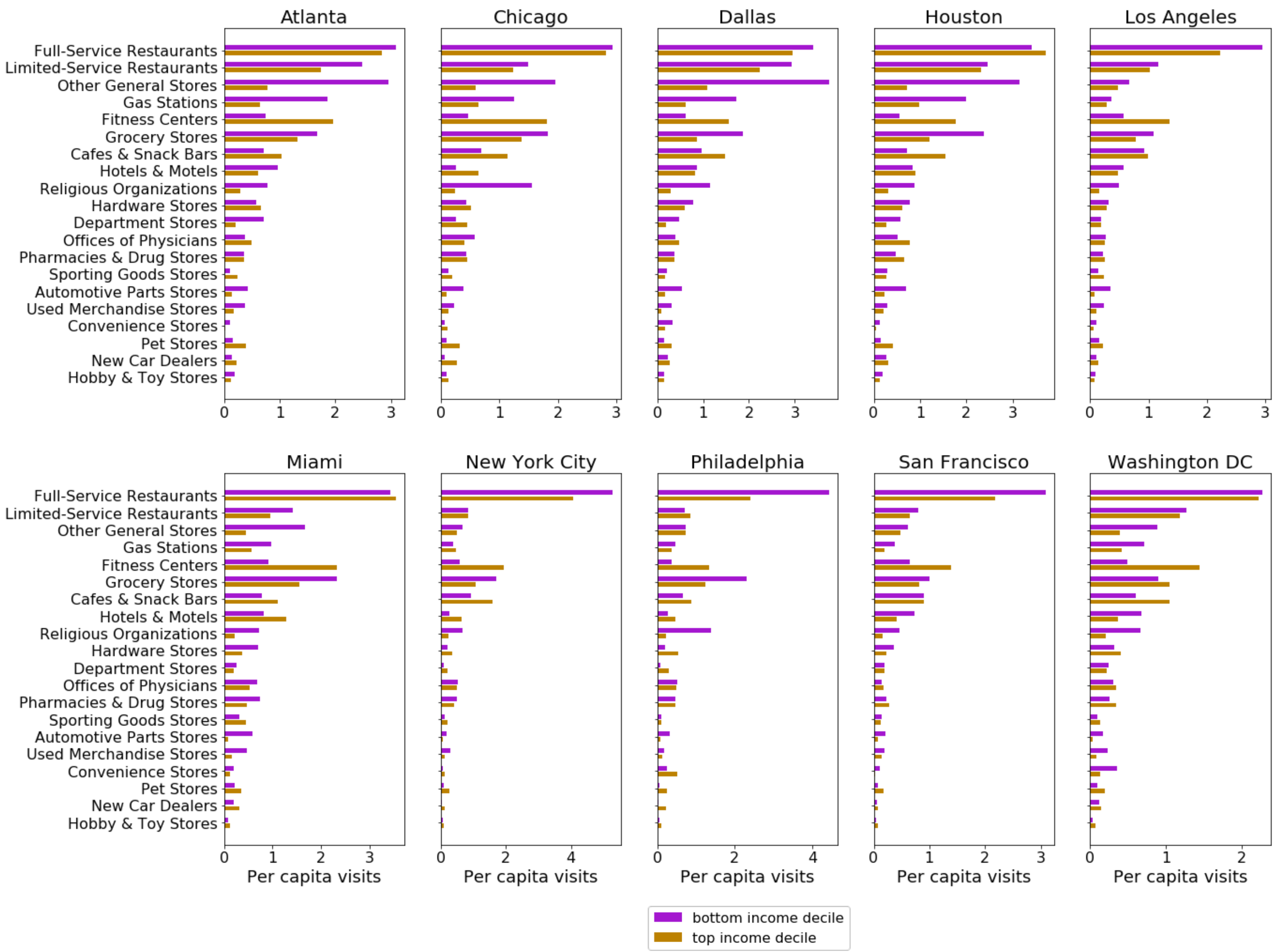

Figure S3: Visits per capita from CBGs in the bottom- (purple) and top- (gold) income deciles to each POI category, accumulated from March 1–May 2, 2020. See Methods M5 for details.

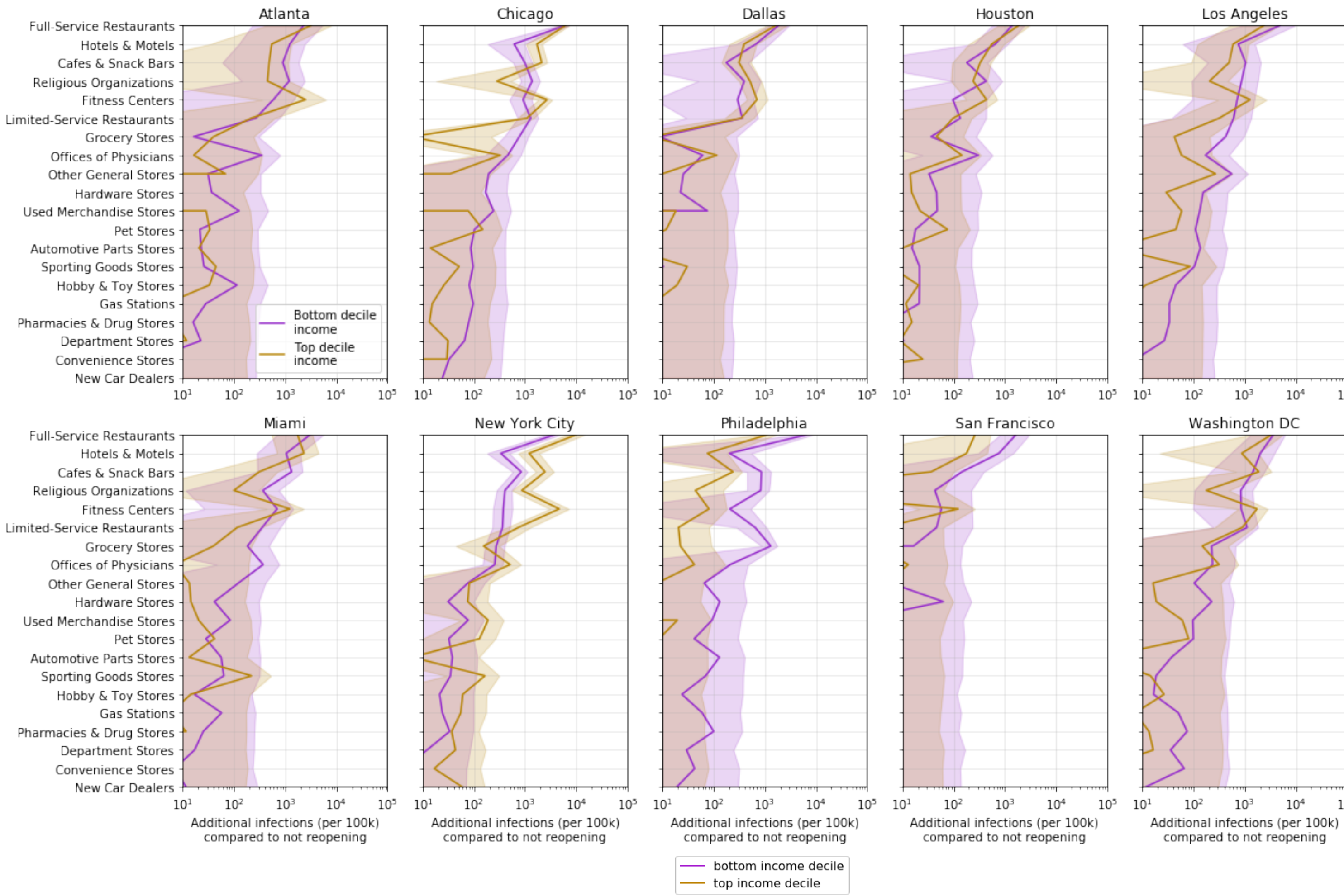

Figure S4: Model predicted additional infections (per 100k population) from reopening each POI category, for CBGs in the top- (gold) and bottom- (purple) income deciles. Predicted reopening impacts are generally worse for lower-income CBGs. Shaded regions denote 2.5th and 97.5th percentiles across sampled parameters and stochastic realizations. See Methods M5 for details.

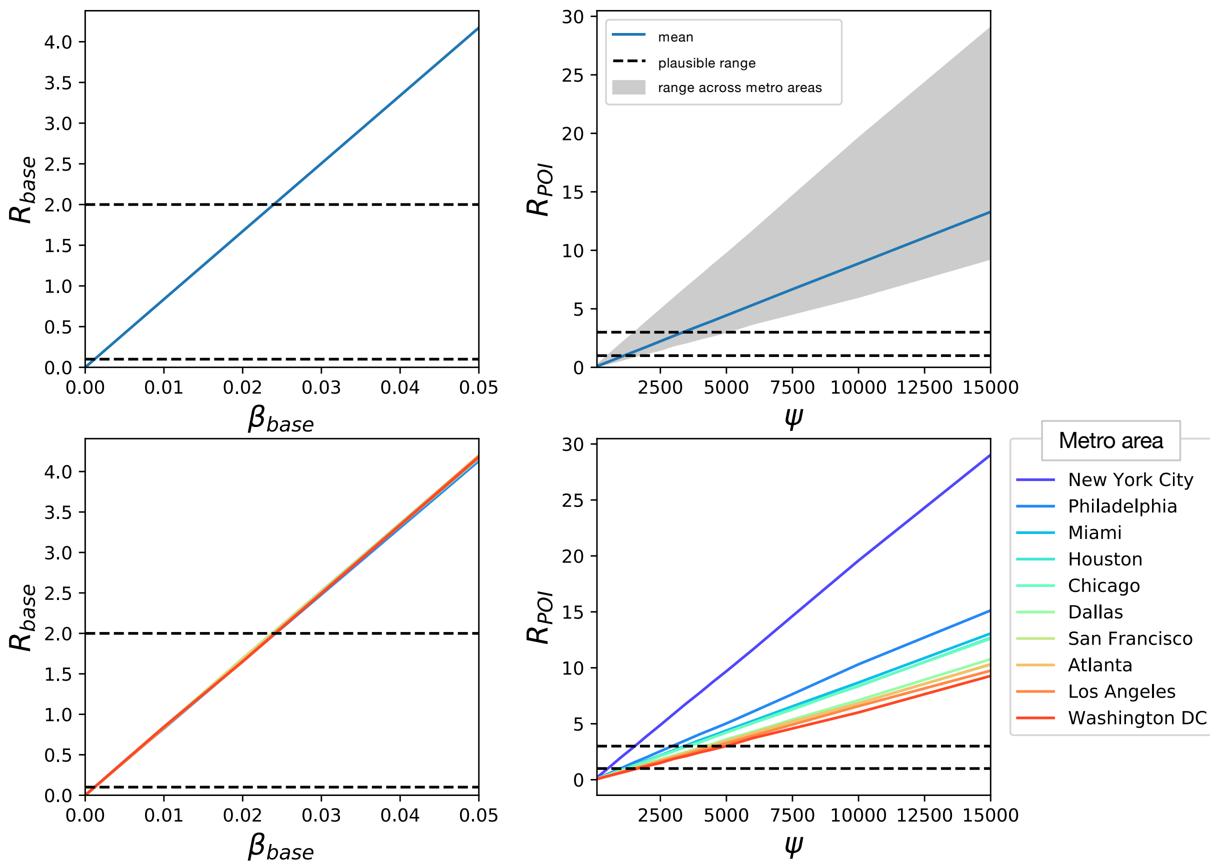

Figure S5: R_{base} and R_{POI} implied by model parameter settings, where ψ is the scaling factor for POI transmission and β_{base} is the base CBG transmission rate. In the top two plots, dotted black lines denote plausible ranges from prior work, the blue line shows the mean across metro areas, and the grey shaded area indicates the range across metro areas. R_{base} does not vary across metro areas because it does not depend on metro area-specific social activity. The bottom two plots show the same results broken down by metro area. See Methods M4.1 for details.

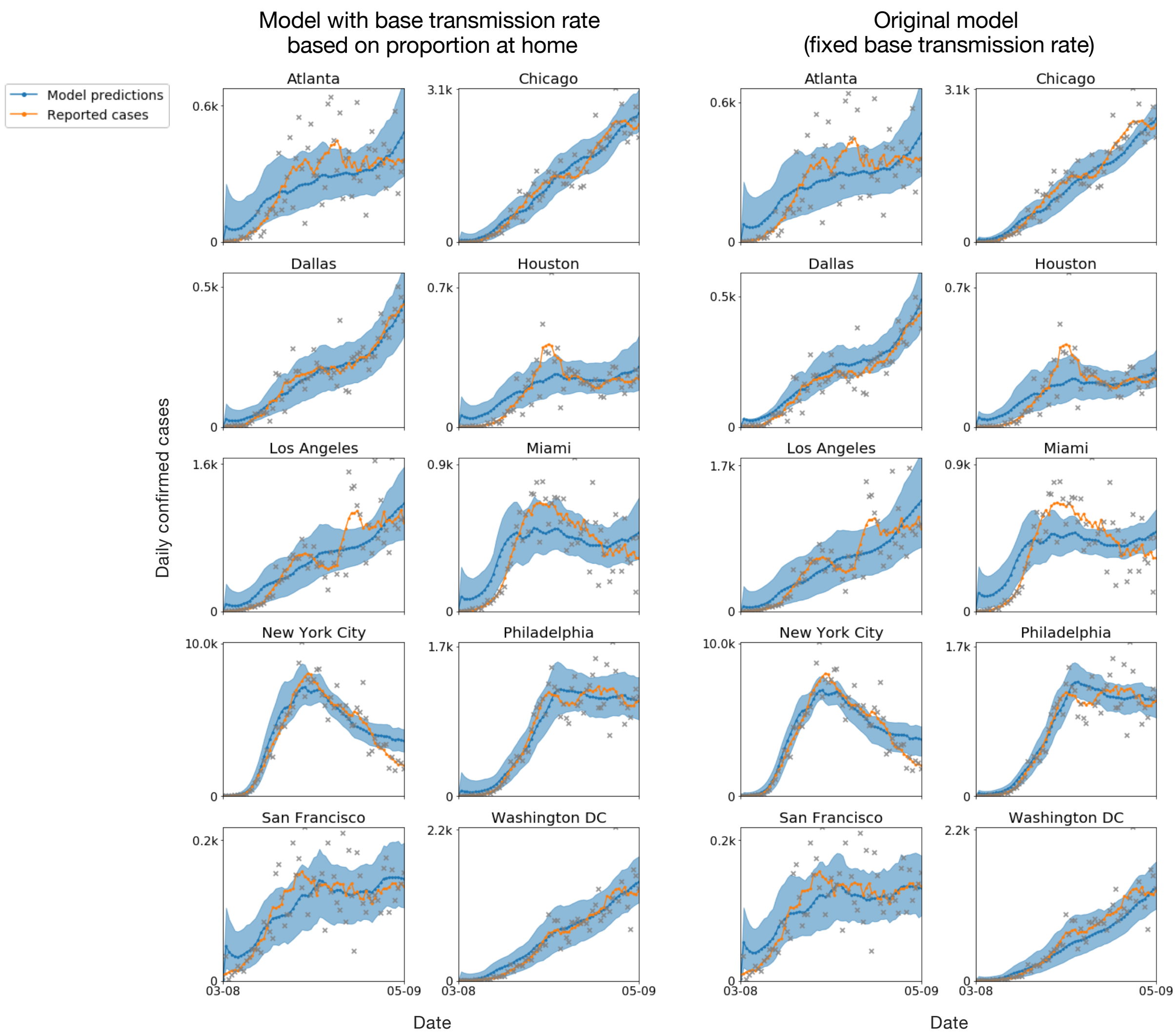

Figure S6: Sensitivity analysis of time-varying base transmission rate. Instead of assuming a fixed base transmission rate, we designed an alternate model where each CBG's base transmission rate varied with the proportion of the CBG that was at home at time t; see SI Methods 2.1 for details. We found that the predictions of this modified model (left) were highly similar to the predictions of the original model (right). The x-axis is the same across plots, showing the date from March 8–May 9, 2020. The grey x's represent the daily reported cases; since they tend to have great variability, we also show the smoothed weekly average (orange). Shaded regions denote 2.5th and 97.5th percentiles across sampled parameters and stochastic realizations.

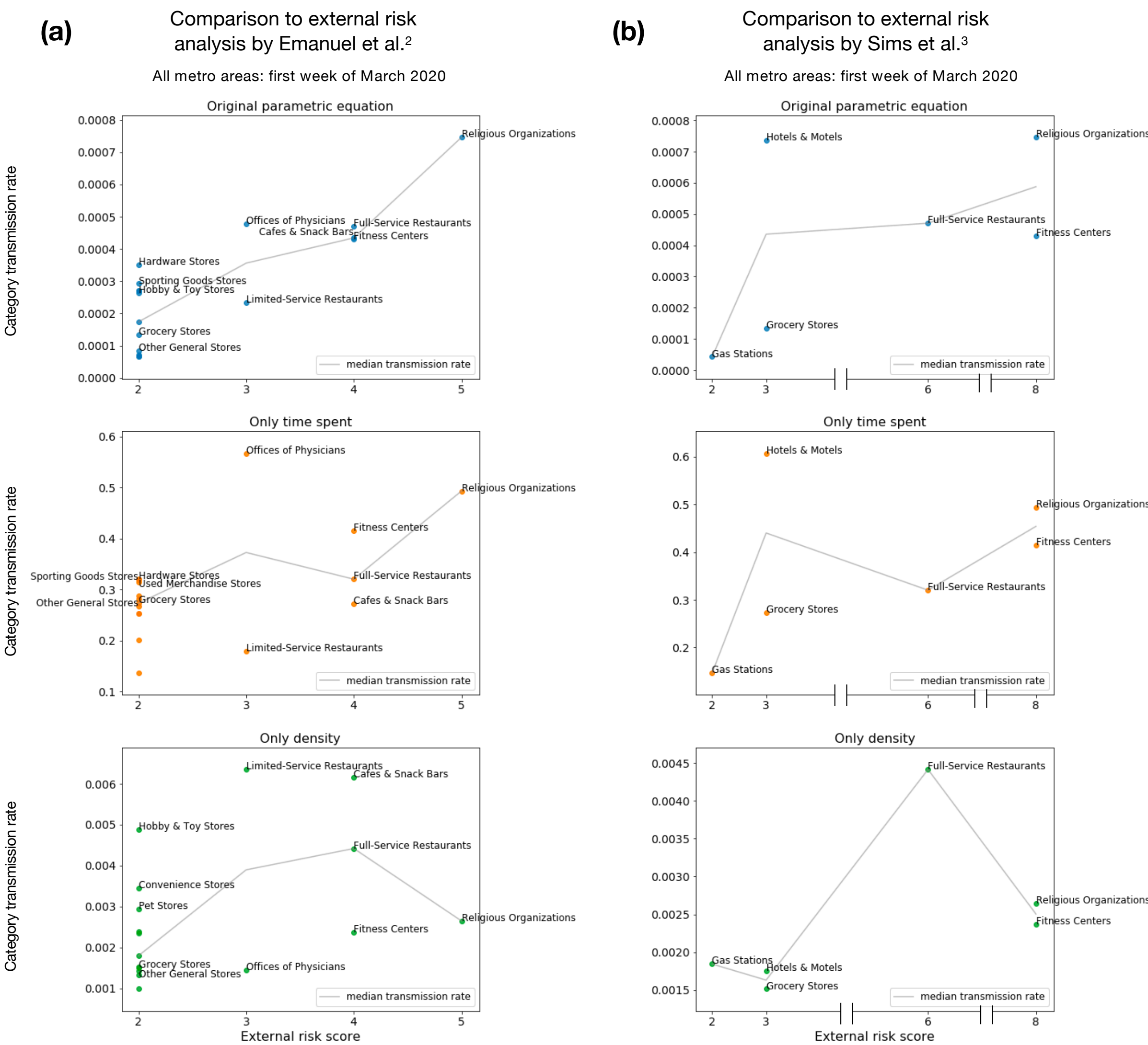

Figure S7: Sensitivity analysis on the parametric form for POI transmission rate. Our model assumes that POI transmission rates depend on two factors: time spent at the POI and the density of individuals per square foot. We tested this assumption by computing an alternate transmission rate that only included time spent (removing density) and another version that only included density (removing time spent); see SI Methods 2.2 for details. We found that the relative risks predicted by our original transmission rate formula concorded best with the assessments of risk proposed by independent epidemiological experts.[2,3] The x-axis represents their proposed risk scores; some scores are missing (e.g., 3 and 4 on the right) because there was no overlap between the categories they assigned that score and categories that we analyzed. The y-axis represents each category's predicted average transmission rate in the first week of March, taking the median over metro areas. Due to space constraints, only a subset of the categories scored at 2 by Emanuel et al. (left) are labeled – the labels are reserved for either the 2 most visited categories in this group (Grocery Stores and Other General Stores) and/or the 3 categories with highest predicted transmission rates within the group.

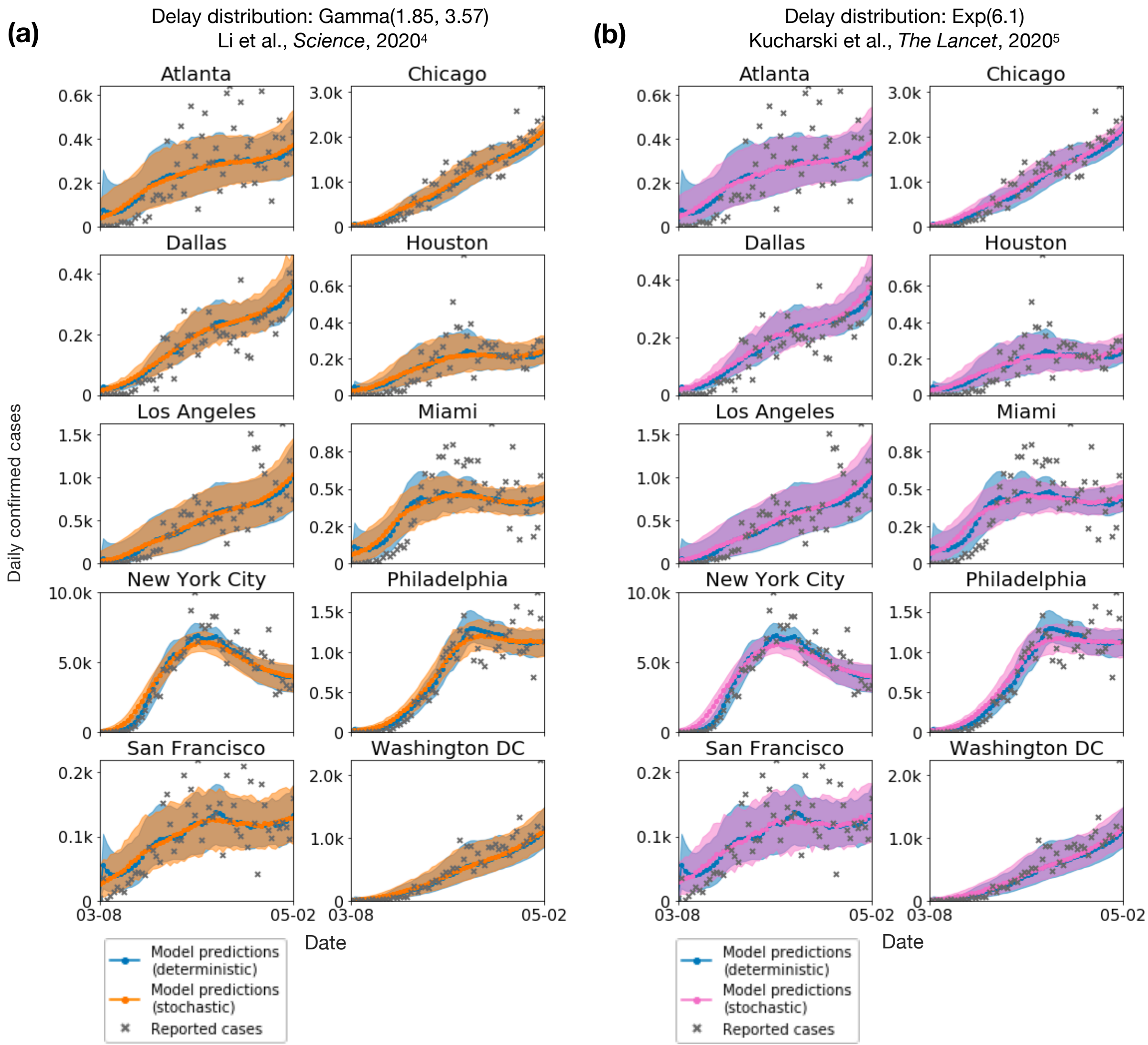

Figure S8: Sensitivity analysis on confirmation rate and delay. Instead of assuming a constant confirmation rate and constant infectious-to-confirmation delay on cases, we tested sampling the number of confirmed cases and delay distribution stochastically. The number of confirmed cases was sampled from a Binomial distribution, and we tried two different delay distributions that were fitted on empirical line list data, (a) Li et al.[4] and (b) Kucharski et al.[5] (see SI Methods 2.3 for details). For both delay distributions, we find that model predictions under the stochastic setting are highly similar to the predictions made under the constant rate and delay setting (labeled as "deterministic" in the plot). Note that the "deterministic" and "stochastic" labels only apply to the computation of confirmed cases from infectious individuals to confirmed cases; the underlying SEIR models are all stochastic, as described in Methods M3. The x-axis is the same across plots, showing the date from March 8–May 2, 2020. Shaded regions denote 2.5th and 97.5th percentiles across sampled parameters and stochastic realizations.

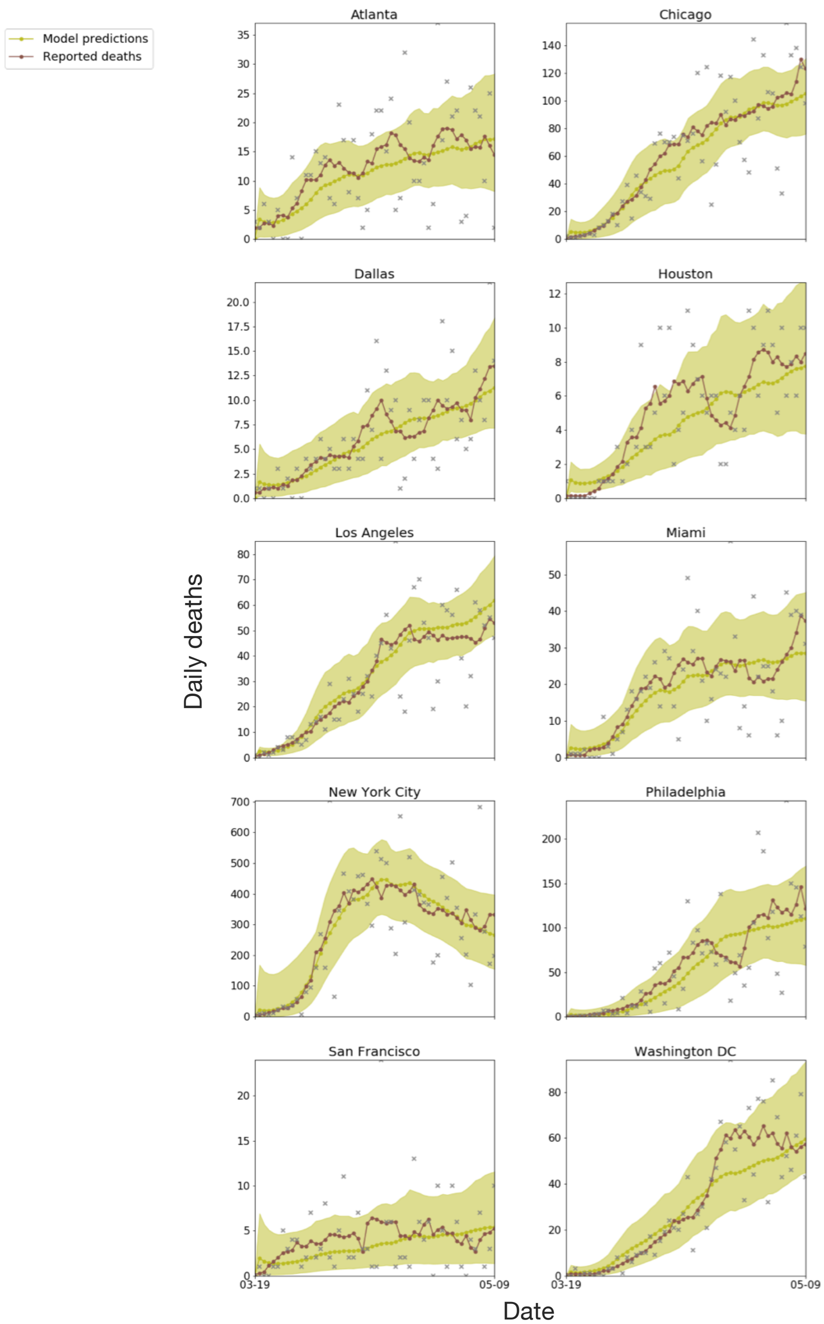

Figure S9: Predicted (green) and true (brown) daily death counts, when our model is calibrated on observed *death* counts. The x-axis is the same across plots, showing the date from March 19–May 9, 2020. The grey x's represent the daily reported deaths; since they tend to have great variability, we also show the smoothed weekly average (brown). Shaded regions denote 2.5th and 97.5th percentiles across sampled parameters and stochastic realizations. See SI Methods 2.4 for details.

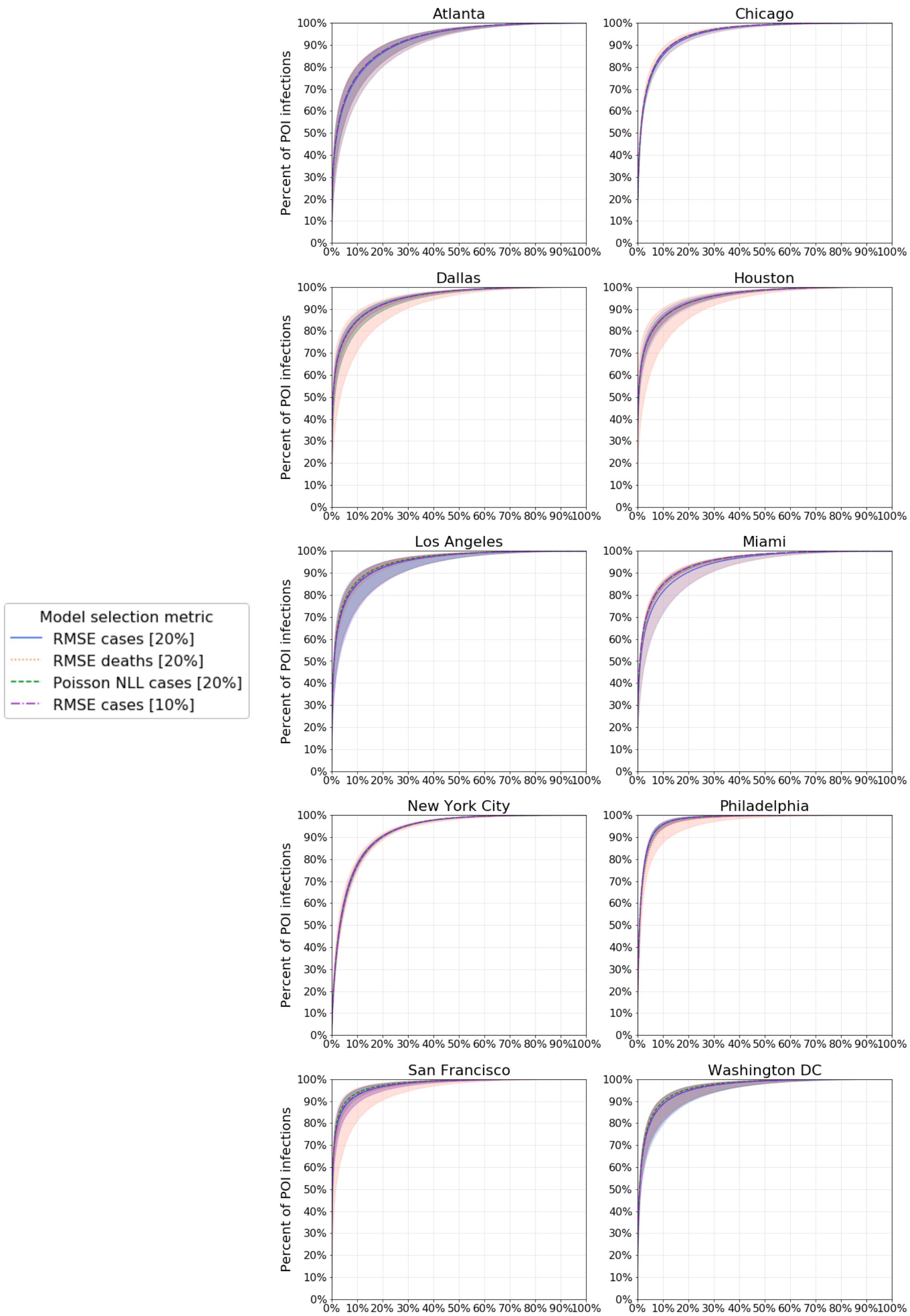

Figure S10: A small fraction of POIs account for a large fraction of the predicted infections at POIs. We additionally conducted a sensitivity analysis on which metric was used for model calibration (SI Methods 2.4) and show that this key finding holds across all metrics. For each metric, we ran the fitted models on the observed mobility data from March 1–May 2, 2020 and recorded the predicted number of infections that occurred at each POI (Methods M5). Shaded regions denote 2.5th and 97.5th percentiles across sampled parameters and stochastic realizations.

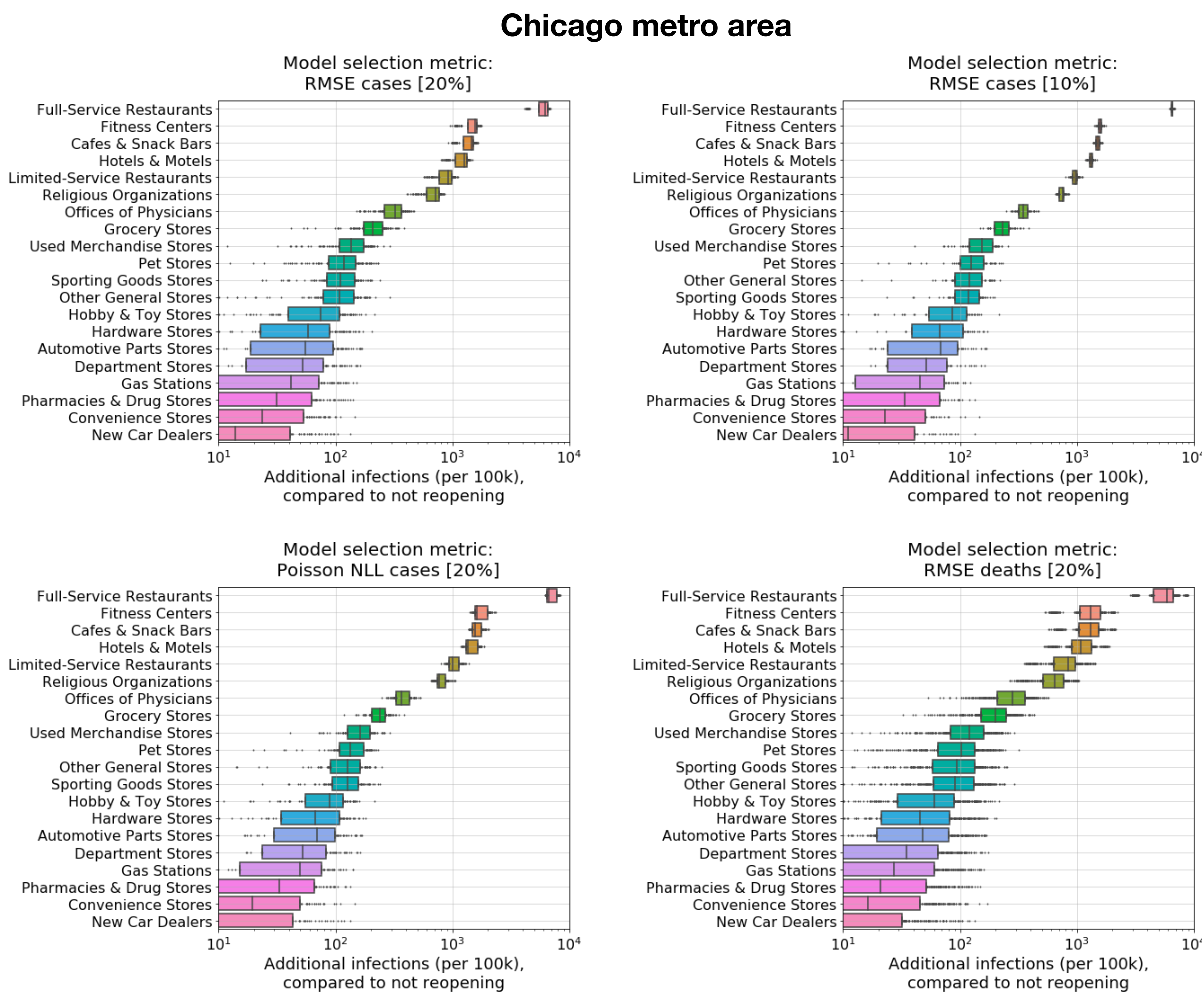

Figure S11: Sensitivity analysis on model calibration metrics and reopening risks. We conducted a sensitivity analysis on which metric was used for model calibration, comparing our default metric (top left) to three other metrics (SI Methods 2.4). We ran our reopening experiments forward with the model parameters selected by each metric (Methods M5). The predicted ranking of risk from reopening different POI categories remains consistent across all metrics. All boxes denote the interquartile range across parameter sets and stochastic realizations, with data points outside the range individually shown. For the Chicago metro area, with 30 stochastic realizations per parameter set, our original metric (top left) selected 4 parameter sets (N = 120); RMSE cases with 10% threshold (top right) selected 2 parameter sets (N = 60); Poisson negative log-likelihood (bottom left) selected 3 parameter sets (N = 90); and RMSE deaths (bottom right) selected 12 parameter sets (N = 360).

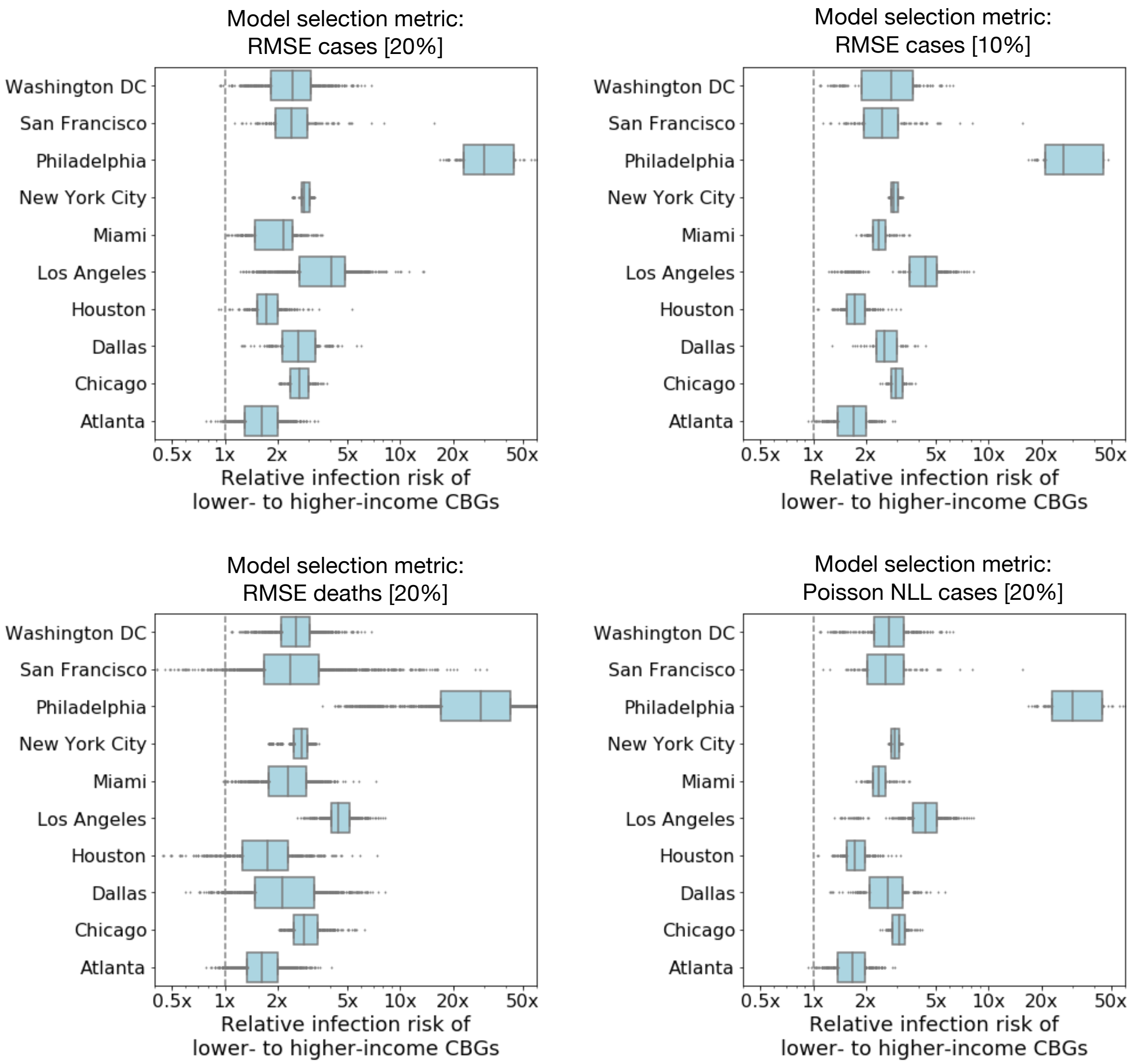

Figure S12: Sensitivity analysis on model calibration metrics and predicted socioeconomic disparities. We conducted a sensitivity analysis on which metric was used for model calibration, comparing our default metric (top left) to three other metrics (SI Methods 2.4). We then analyzed the socioeconomic disparities in each metro area predicted by the model parameters selected by each metric. The predicted disparities remain remarkably consistent across all metrics, and, for every metric, the best fit models predict that lower-income CBGs are at higher infection risk. All boxes denote the interquartile range across parameter sets and stochastic realizations, with data points outside the range individually shown. Across metro areas, with 30 stochastic realizations per parameter set, our original metric (top left) selected 97 parameter sets (N = 2,910); RMSE cases with 10% treshold (top right) selected 45 parameter sets (N = 1,350); Poisson negative log-likelihood (bottom left) selected 52 parameter sets (N = 1,560); and RMSE deaths (bottom right) selected 251 parameter sets (N = 7,530).

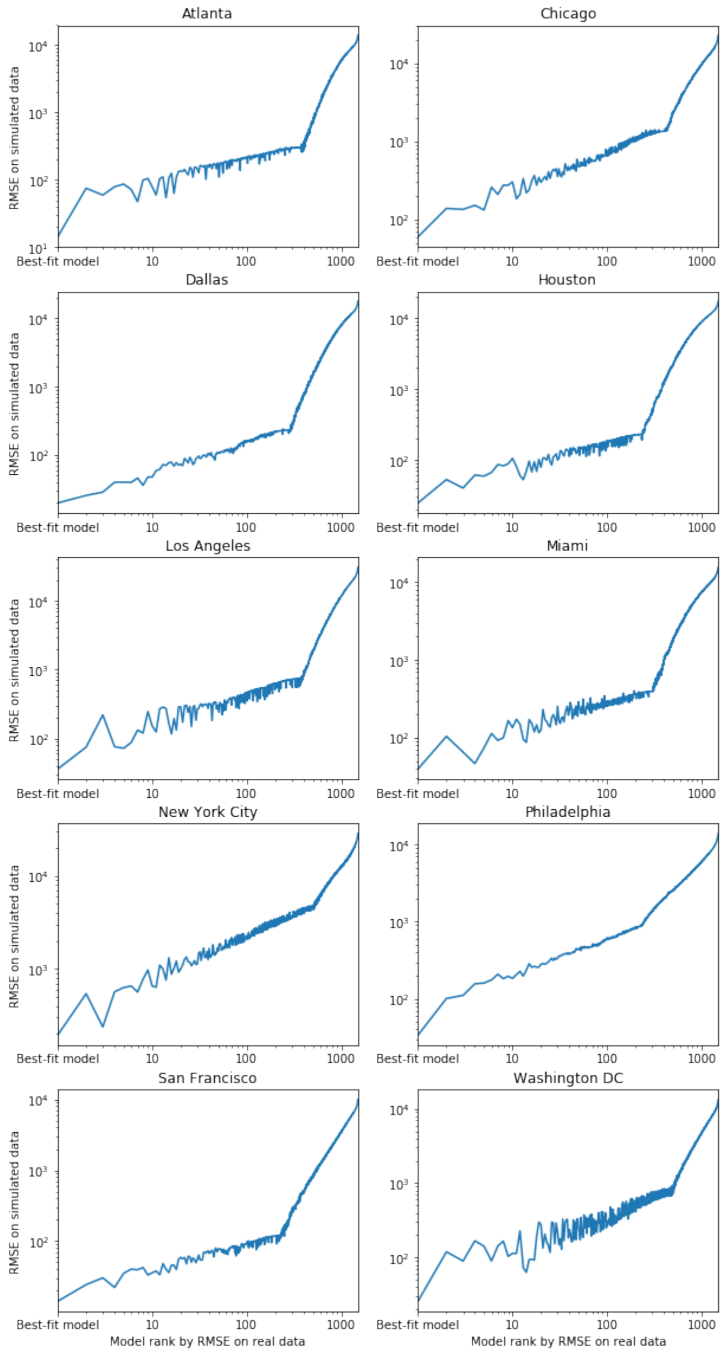

Figure S13: Assessing model identifiability on simulated data. The x-axis ranks parameter settings by how well they fit *real* data (measured by RMSE to daily cases). The y-axis plots plots RMSE on *simulated* case count data generated using the best-fit parameter settings. For all 10 metro areas, the leftmost point—which corresponds to best-fit parameter setting, i.e., the parameters we use as ground truth for the simulated data—also obtains the lowest loss on the simulated data. This demonstrates that the model and fitting procedure can correctly recover the true parameters in simulated data. SI Methods 2.5 provides more details.

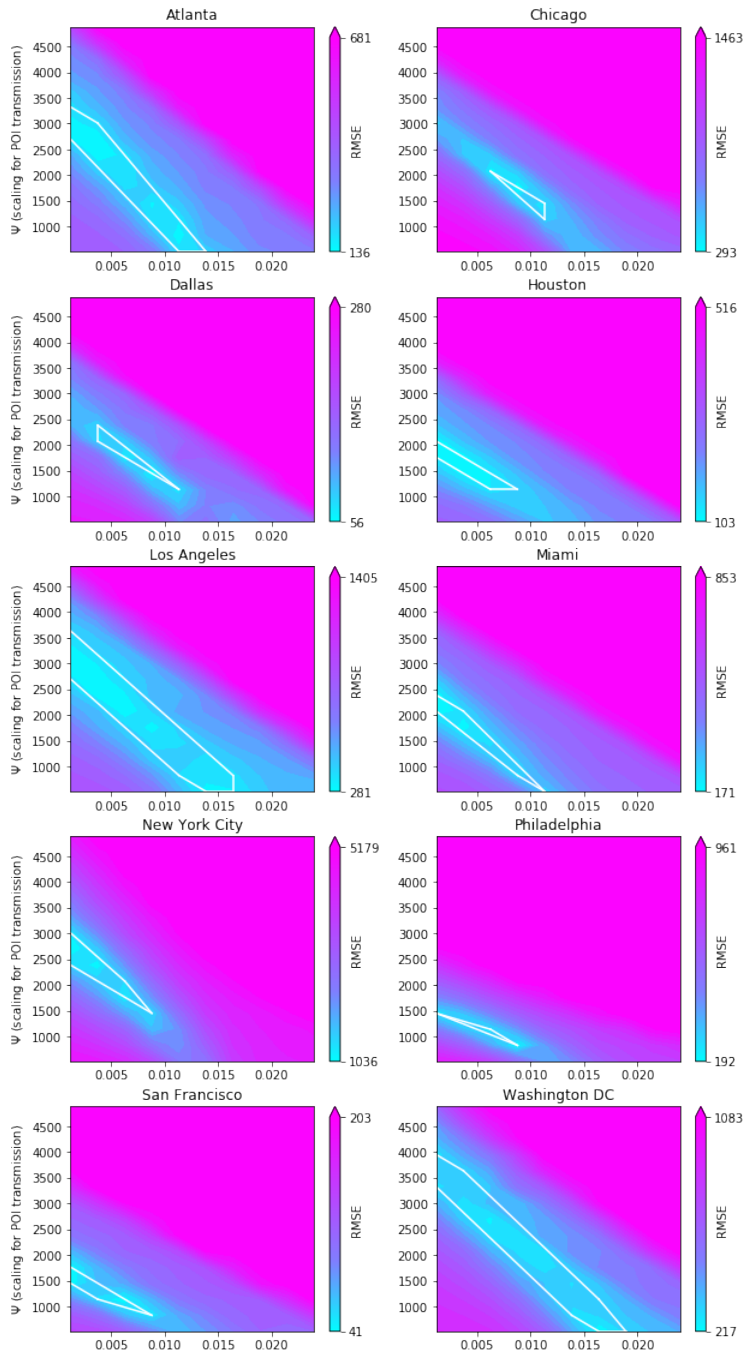

Figure S14: RMSE on daily case count data as a function of parameters ψ (y-axis), which scales POI transmission rates, and β_{base} (x-axis), which is the base CBG transmission rate. Color indicates the RMSE, normalized such that blue represents the RMSE of the best-fit model. The white polygon shows the convex hull of the parameter settings used to generate results: i.e., all models with an RMSE less than $1.2\times$ that of the best-fit model. For all parameter combinations, we take the minimum RMSE over p_0. SI Methods 2.5 provides more details.

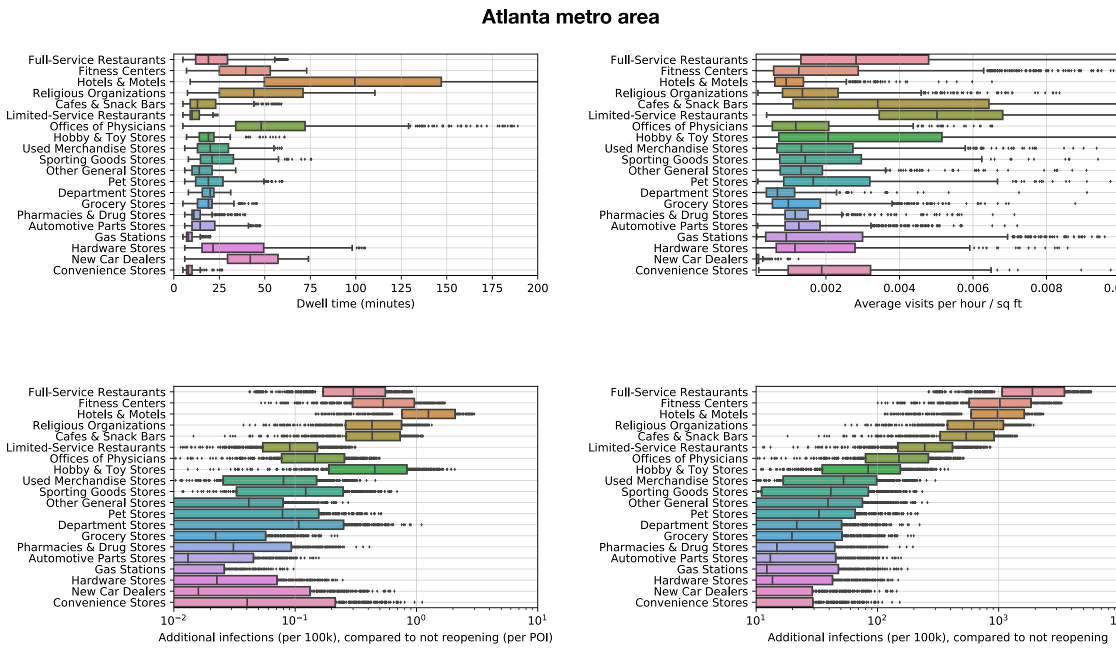

Figure S15: POI attributes in Atlanta metro area. The top two plots display quantities from the mobility data: the dwell time and the average number of hourly visitors divided by POI area. Each point represents one POI; boxes depict the interquartile range across POIs, with data points outside the range individually shown. The bottom two plots show model predictions for the increase in infections from reopening a POI category: per POI (left bottom) and for the category as a whole (right bottom). Each point represents one model realization; boxes depict the interquartile range across realizations, with data points outside the range individually shown. In Atlanta, we model 39,411 POIs in total, and we sample 16 parameter sets and 30 stochastic realizations (N=480).

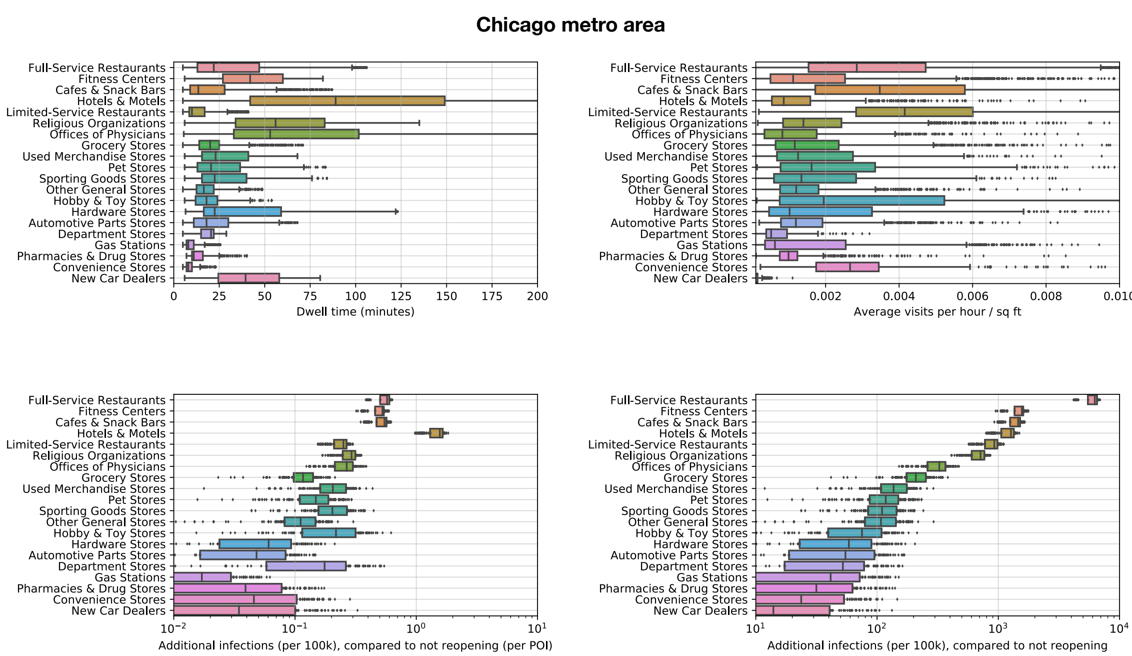

Figure S16: POI attributes in the Chicago metro area. See Figure S15 for details. In Chicago, we model 62,420 POIs in total, and we sample 4 parameter sets and 30 stochastic realizations (N=120).

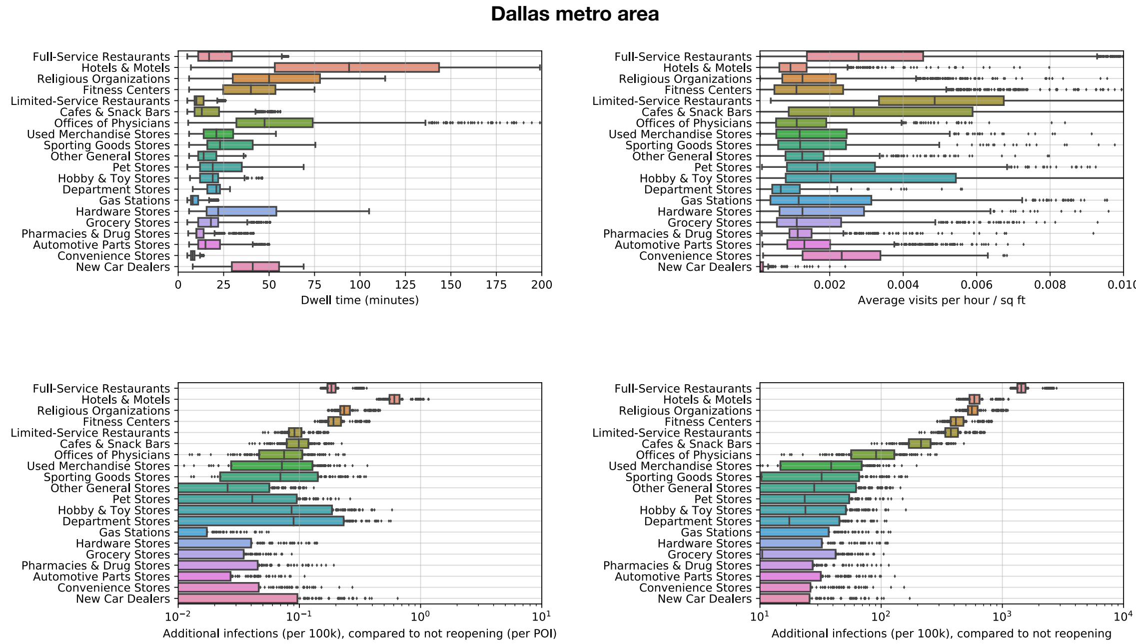

Figure S17: POI attributes in the Dallas metro area. See Figure S15 for details. In Dallas, we model 52,999 POIs in total, and we sample 5 parameter sets and 30 stochastic realizations (N=150).

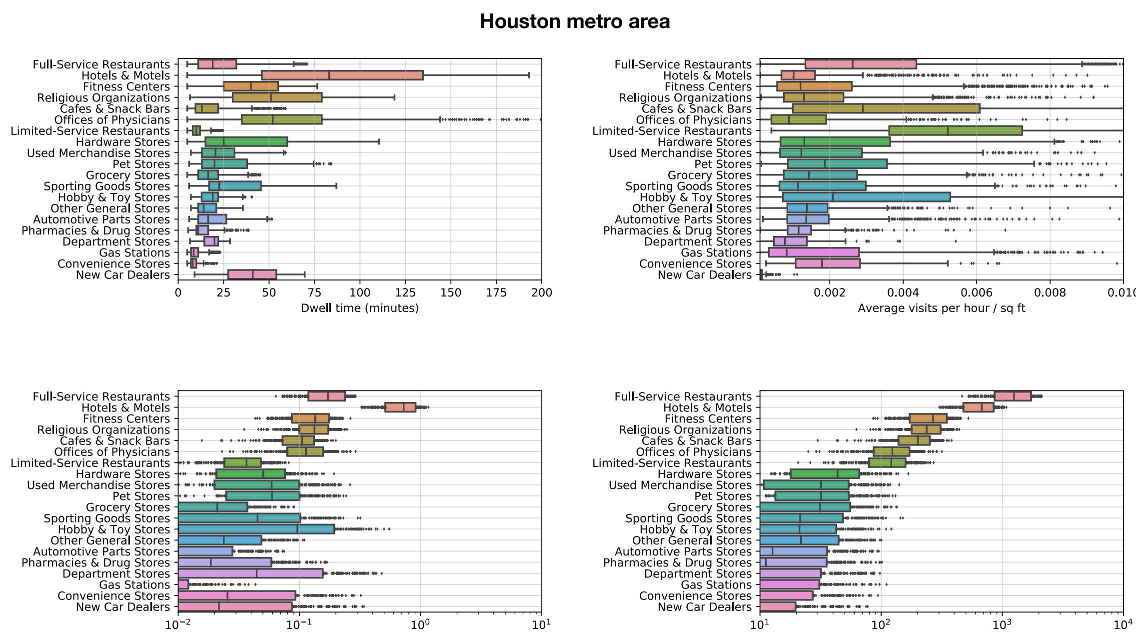

Figure S18: POI attributes in the Houston metro area. See Figure S15 for details. In Houston, we model 49,622 POIs in total, and we sample 8 parameter sets and 30 stochastic realizations (N=240).

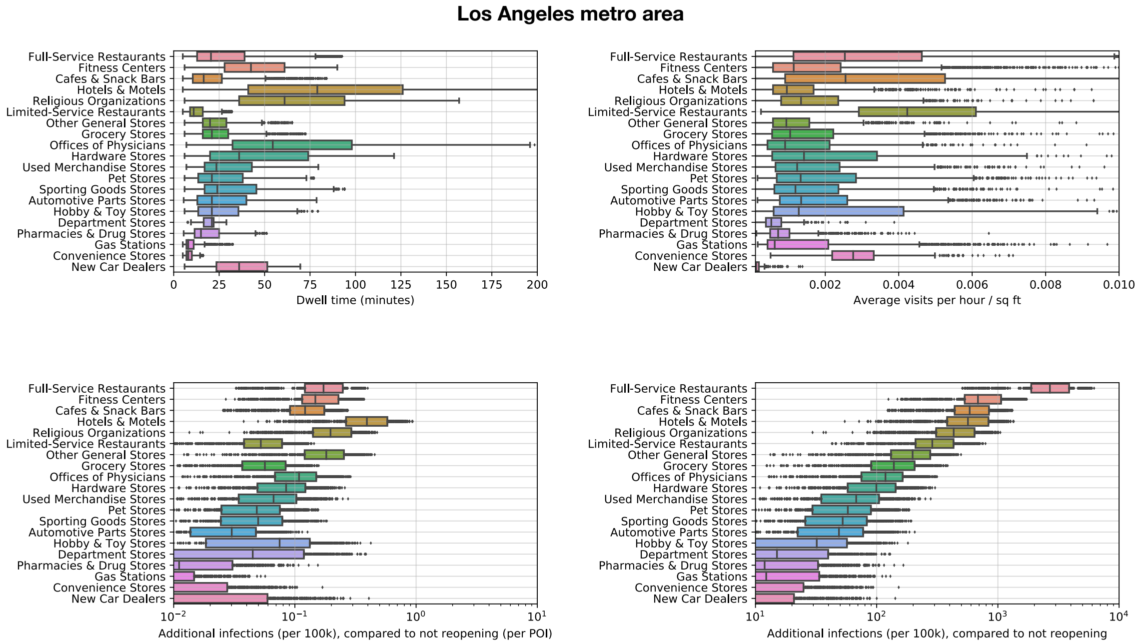

Figure S19: POI attributes in Los Angeles metro area. See Figure S15 for details. In Los Angeles, we model 83,954 POIs in total, and we sample 25 parameter sets and 30 stochastic realizations (N=750).

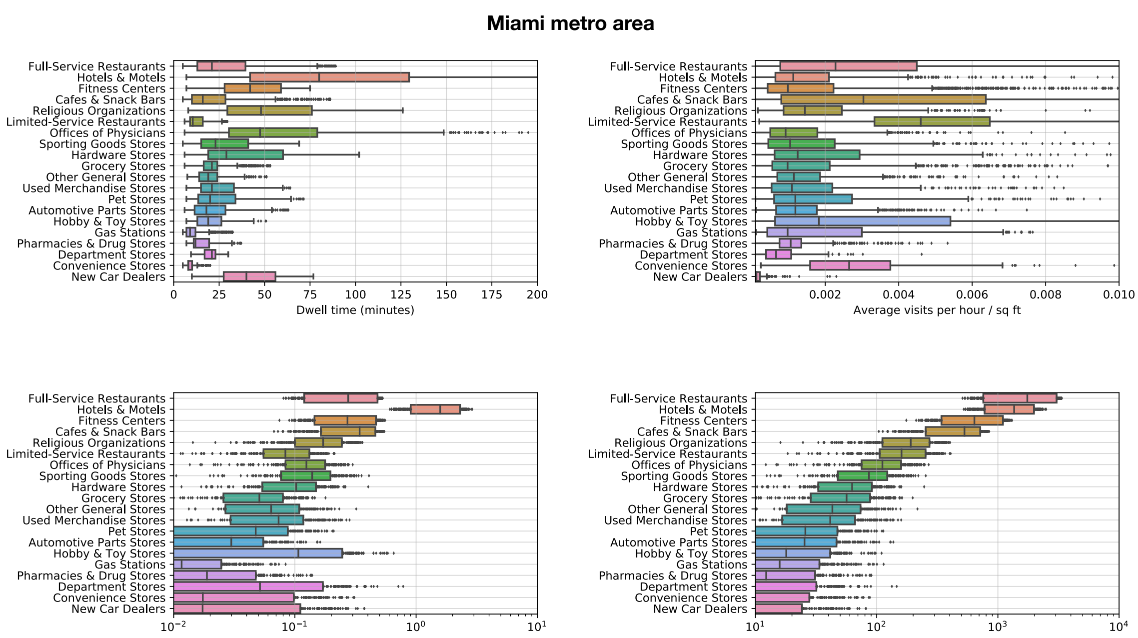

Figure S20: POI attributes in Miami metro area. See Figure S15 for details. In Miami, we model 40,964 POIs in total, and we sample 7 parameter sets and 30 stochastic realizations (N=210).

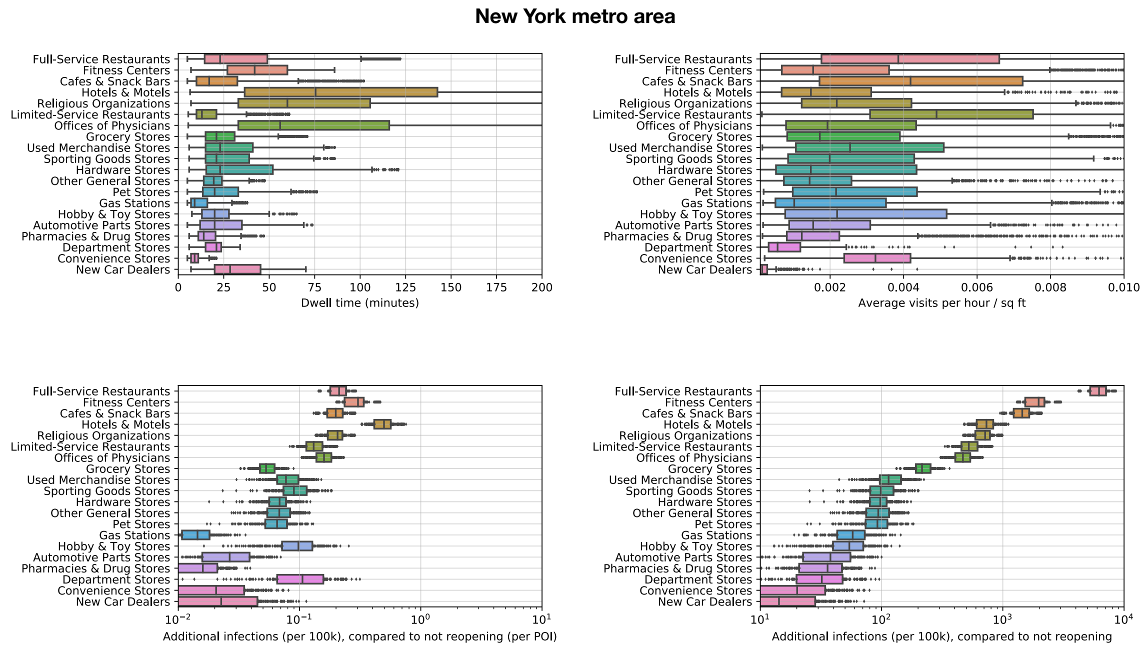

Figure S21: POI attributes in New York metro area. See Figure S15 for details. In New York, we model 122,428 POIs in total, and we sample 7 parameter sets and 30 stochastic realizations (N=210).

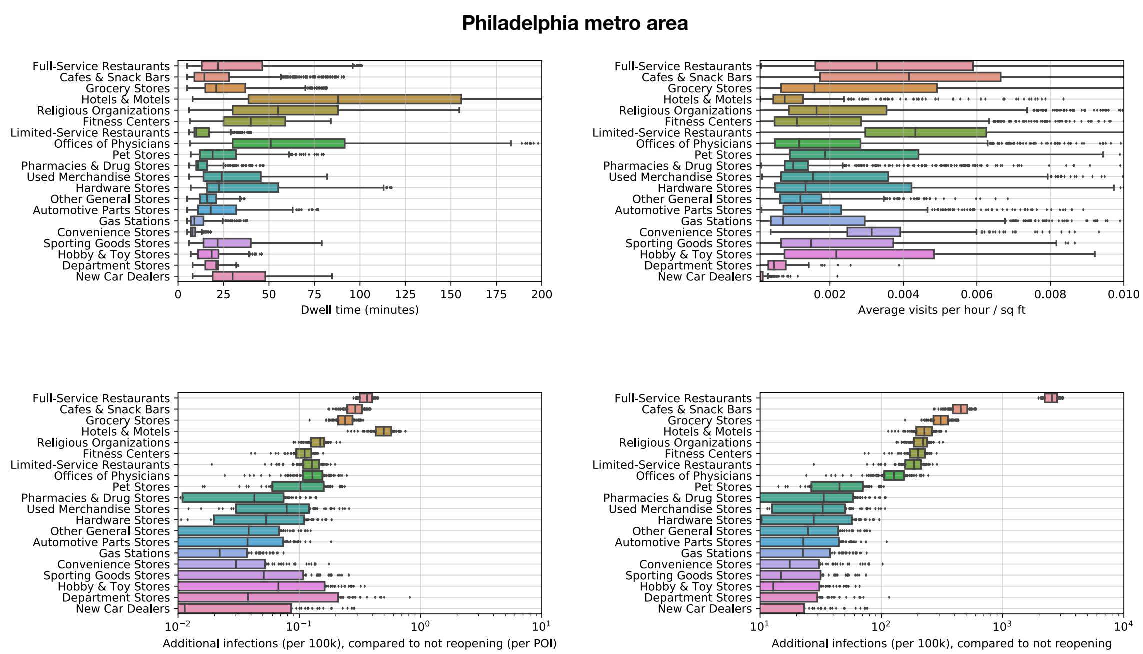

Figure S22: POI attributes in Philadelphia metro area. See Figure S15 for details. In Philadelphia, we model 37,951 POIs in total, and we sample 3 parameter sets and 30 stochastic realizations (N=90).

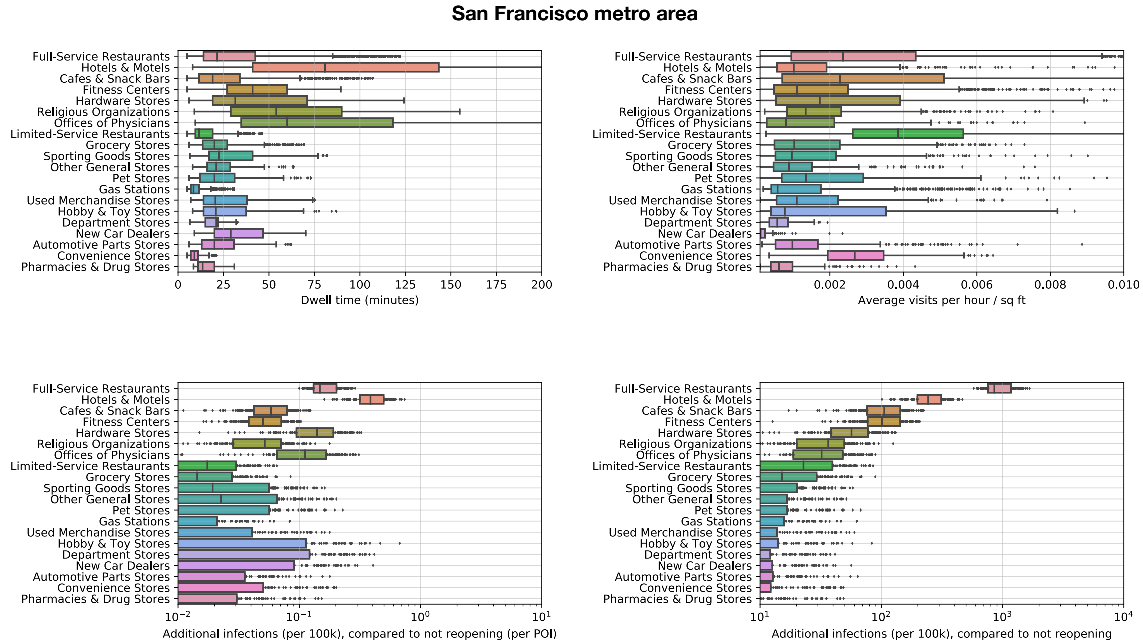

Figure S23: POI attributes in San Francisco metro area. See Figure S15 for details. In San Francisco, we model 28,713 POIs in total, and we sample 5 parameter sets and 30 stochastic realizations (N=150).

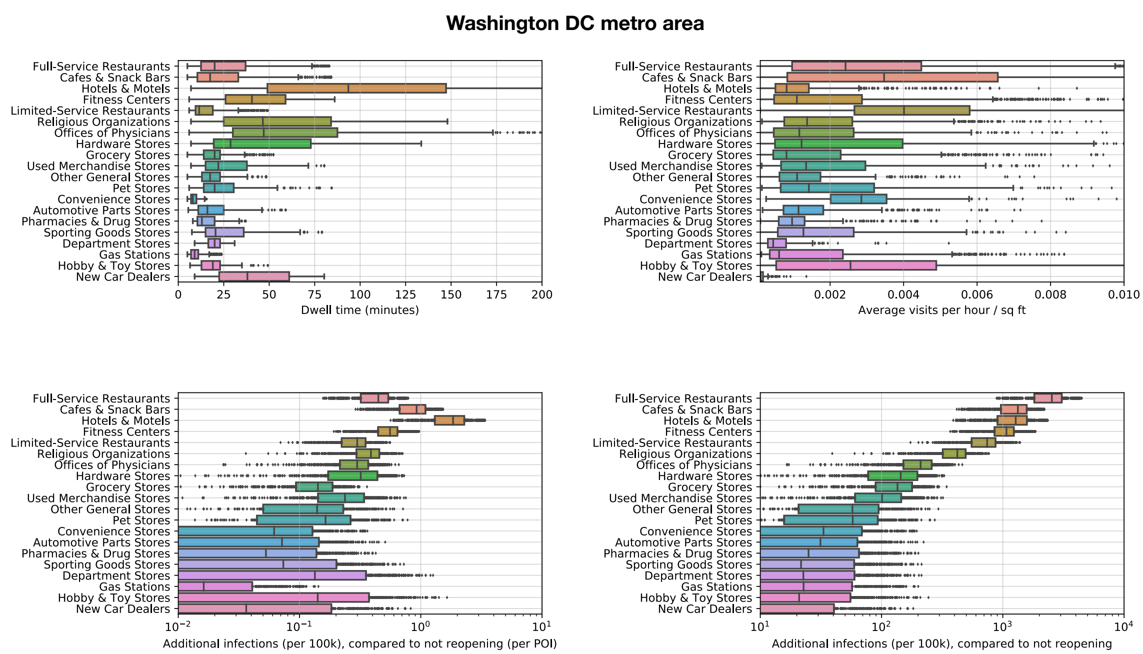

Figure S24: POI attributes in Washington DC metro area. See Figure S15 for details. In DC, we model 34,296 POIs in total, and we sample 17 parameter sets and 30 stochastic realizations (N=510).

Supplementary References

1. Google. COVID-19 community mobility reports (2020). Available at https://google.com/covid19/mobility/.

2. Emanuel, E. J., Phillips, J. P. & Popescu, S. COVID-19 Activity Risk Levels (2020). Available at http://www.ezekielemanuel.com/writing/all-articles/2020/06/30/covid-19-activity-risk-levels.

3. DesOrmeau, T. From hair salons to gyms, experts rank 36 activities by coronavirus risk level. *MLive* (2020). Available at https://www.mlive.com/public-interest/2020/06/from-hair-salons-to-gyms-experts-rank-36-activities-by-coronavirus-risk-level.html.

4. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* **368**, 489–493 (2020).

5. 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).

6. Verity, R. *et al.* Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet* **20**, 669–677 (2020).

7. 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).

8. SafeGraph. Measuring and Correcting Sampling Bias in Safegraph Patterns for More Accurate Demographic Analysis (2020). Available at https://safegraph.com/blog/measuring-and-correcting-sampling-bias-for-accurate-demographic-analysis.

9. SafeGraph. Places Manual (2020). Available at https://docs.safegraph.com/docs/places-manual#section-visitor-home-cbgs.

10. 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).

11. Bishop, Y. M., Fienberg, S. E. & Holland, P. W. Discrete multivariate analysis (1975).

12. 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).

13. Wong, D. W. The reliability of using the iterative proportional fitting procedure. *The Professional Geographer* **44**, 340–348 (1992).

14. 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).

15. Csiszár, I. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability* 146–158 (1975).

16. Pukelsheim, F. Biproportional scaling of matrices and the iterative proportional fitting procedure. *Annals of Operations Research* **215**, 269–283 (2014).

17. Gietl, C. & Reffel, F. P. Accumulation points of the iterative proportional fitting procedure. *Metrika* **76**, 783–798 (2013).

18. APM Research Lab. The color of coronavirus: COVID-19 deaths by race and ethnicity in the U.S. (2020). Available at https://apmresearchlab.org/covid/deaths-by-race.

19. Buckee, C. O. *et al.* Aggregated mobility data could help fight COVID-19. *Science* **368**, 145 (2020).

20. 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.

21. 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).

22. 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.

23. Galeazzi, A. *et al.* Human Mobility in Response to COVID-19 in France, Italy and UK (2020). Available at arxiv.org/abs/2005.06341.

24. 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).

25. Woody, S. *et al.* Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile phones. *medRxiv* (2020). Available at doi.org/10.1101/2020.04.16.20068163.

26. Fenichel, E. P., Berry, K., Bayham, J. & Gonsalves, G. A cell phone data driven time use analysis of the COVID-19 epidemic. *medRxiv* (2020). Available at doi.org/10.1101/2020.04.20.20073098.

27. 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/.