Simulation of SARS-CoV2 Spread and Intervention Effects in the Continental US with Variable Contact Rates, March 24, 2020

Authors: Sen Pei, Jeffrey Shaman

Date: March 26, 2020

Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University

Here we use a metapopulation model applied at county resolution to simulate the spread and growth of COVID-19 incidence in the continental United States. We calibrate the model against county-level incidence data collected between February 21, 2020 and March 24, 2020, and project the outbreak in the continental US for 4 weeks after March 24, 2020. Projections for daily reported cases, hospital bed demand, ICU bed demand and mortality are generated. We also evaluate the effects of social distancing on the outbreak.

Model

We use a metapopulation SEIR model[1] to simulate the transmission of COVID-19 among 3,108 US counties. In this model, we consider two types of movement: daily work commuting and random movement. Information on county-to-county work commuting is publicly available from the US Census Bureau[2]. We further assume the number of random visitors between two counties is proportional to the average number of commuters between them. As population present in each county is different during daytime and nighttime, we model the transmission dynamics of COVID-19 separately for these two time periods.

We formulate the transmission as a discrete Markov process during both day and night times. The daytime transmission lasts for dt_1 day and the nighttime transmission dt_2 day ($dt_1 + dt_2 = 1$). Here, we assume daytime transmission lasts for 8 hours and nighttime transmission lasts for 16 hours, i.e., $dt_1 = 1/3$ and $dt_2 = 2/3$. The transmission dynamics are depicted by the following equations.

Daytime transmission:

$$S_{ij}(t+dt_1) = S_{ij}(t) - \frac{\beta S_{ij}(t)\sum_k I^r_{ki}(t)}{N^D_i(t)}dt_1 - \frac{\mu\beta S_{ij}(t)\sum_k I^u_{ik}(t)}{N^D_i(t)}dt_1 + \theta dt_1 \frac{N_{ij} - I^r_{ij}(t)}{N^D_i(t)}\sum_{k\neq i}\frac{\overline{N}_{ik}\sum_l S_{kl}(t)}{N^D_k(t) - \sum_l I^r_{lk}(t)} - \theta dt_1 \frac{S_{ij}(t)}{N^D_i(t) - \sum_l I^r_{li}(t)}\sum_{k\neq i}\overline{N}_{ki} \quad (1)$$

$$E_{ij}(t+dt_1) = E_{ij}(t) + \frac{\beta S_{ij}(t)\sum_k I^r_{ki}(t)}{N^D_i(t)}dt_1 + \frac{\mu\beta S_{ij}(t)\sum_k I^u_{ik}(t)}{N^D_i(t)}dt_1 - \frac{E_{ij}(t)}{Z}dt_1$$
$$+ \theta dt_1 \frac{N_{ij} - I^r_{ij}(t)}{N^D_i(t)} \sum_{k\neq i} \frac{\bar{N}_{ik}\sum_l E_{kl}(t)}{N^D_k(t) - \sum_l I^r_{lk}(t)} - \theta dt_1 \frac{E_{ij}(t)}{N^D_i(t) - \sum_l I^r_{li}(t)} \sum_{k\neq i} \bar{N}_{ki} \quad (2)$$

$$I^r_{ij}(t+dt_1) = I^r_{ij}(t) + \alpha\frac{E_{ij}(t)}{Z}dt_1 - \frac{I^r_{ij}(t)}{D}dt_1 \quad (3)$$

$$I^u_{ij}(t+dt_1) = I^u_{ij}(t) + (1-\alpha)\frac{E_{ij}(t)}{Z}dt_1 - \frac{I^u_{ij}(t)}{D}dt_1$$
$$+ \theta dt_1 \frac{N_{ij} - I^r_{ij}(t)}{N^D_i(t)} \sum_{k\neq i} \frac{\bar{N}_{ik}\sum_l I^u_{kl}(t)}{N^D_k(t) - \sum_l I^r_{lk}(t)} - \theta dt_1 \frac{I^u_{ij}(t)}{N^D_i(t) - \sum_l I^r_{li}(t)} \sum_{k\neq i} \bar{N}_{ki} \quad (4)$$

$$N^D_i(t) = N_{ii} + \sum_{k\neq i} I^r_{ki}(t) + \sum_{k\neq i}(N_{ik} - I^r_{ik}(t)) \quad (5)$$

Nighttime transmission:

$$S_{ij}(t+1) = S_{ij}(t+dt_1) - \frac{\beta S_{ij}(t+dt_1)\sum_k I^r_{kj}(t+dt_1)}{N^N_j}dt_2$$
$$- \frac{\mu\beta S_{ij}(t+dt_1)\sum_k I^u_{kj}(t+dt_1)}{N^N_j}dt_2 + \theta dt_2 \frac{N_{ij}}{N^N_j}\sum_{k\neq j} \frac{\bar{N}_{jk}\sum_l S_{lk}(t+dt_1)}{N^N_k - \sum_l I^r_{lk}(t+dt_1)}$$
$$- \theta dt_2 \frac{S_{ij}(t+dt_1)}{N^N_j - \sum_k I^r_{kj}(t+dt_1)} \sum_{k\neq j} \bar{N}_{kj} \quad (6)$$

$$E_{ij}(t+1) = E_{ij}(t+dt_1) + \frac{\beta S_{ij}(t+dt_1)\sum_k I^r_{kj}(t+dt_1)}{N^N_j}dt_2$$
$$+ \frac{\mu\beta S_{ij}(t+dt_1)\sum_k I^u_{kj}(t+dt_1)}{N^N_j}dt_2 - \frac{E_{ij}(t+dt_1)}{Z}dt_2$$
$$+ \theta dt_2 \frac{N_{ij}}{N^N_j}\sum_{k\neq j} \frac{\bar{N}_{jk}\sum_l E_{lk}(t+dt_1)}{N^N_k - \sum_l I^r_{lk}(t+dt_1)} - \theta dt_2 \frac{E_{ij}(t+dt_1)}{N^N_j - \sum_k I^r_{kj}(t+dt_1)} \sum_{k\neq j} \bar{N}_{kj} \quad (7)$$

$$I^r_{ij}(t+1) = I^r_{ij}(t+dt_1) + \alpha\frac{E_{ij}(t+dt_1)}{Z}dt_2 - \frac{I^r_{ij}(t+dt_1)}{D}dt_2 \quad (8)$$

$$I^u_{ij}(t+1) = I^u_{ij}(t+dt_1) + (1-\alpha)\frac{E_{ij}(t+dt_1)}{Z}dt_2 - \frac{I^u_{ij}(t+dt_1)}{D}dt_2$$
$$+ \theta dt_2 \frac{N_{ij}}{N^N_j}\sum_{k\neq j} \frac{\bar{N}_{jk}\sum_l I^u_{lk}(t+dt_1)}{N^N_k - \sum_l I^r_{lk}(t+dt_1)} - \theta dt_2 \frac{I^u_{ij}(t+dt_1)}{N^N_j - \sum_k I^r_{kj}(t+dt_1)} \sum_{k\neq j} \bar{N}_{kj} \quad (9)$$

$$N^N_i = \sum_k N_{ki} \quad (10)$$

Here, S_{ij}, E_{ij}, I^r_{ij}, I^u_{ij} and N_{ij} are the susceptible, exposed, reported infected, unreported infected and total population in the subpopulation commuting from county j to county i

($i \leftarrow j$); β is the transmission rate of reported infections; μ is the relative transmissibility of unreported infections; Z is the average latency period (from infection to contagiousness); D is the average duration of contagiousness; α is the fraction of documented infections; θ is a multiplicative factor adjusting random movement; $\overline{N}_{ij} = (N_{ij} + N_{ji})/2$ is the average number of commuters between counties i and j; and N_i^D and N_i^N are the daytime and nighttime populations of county i. We integrate Eqs. 1-10 using a Poisson process to represent the stochasticity of the transmission process. A similar model has been used to generate forecasts for influenza in the United States[3].

To account for reporting delay, we mapped simulated documented infections to confirmed cases using a separate observational delay model. In this delay model, we account for the time interval between a person transitioning from latent to contagious (i.e. $E \rightarrow I_i^r$) and observational confirmation of that individual infection. To estimate this delay period, T_d, we examined line-list data from early-confirmed cases in China[4]. Prior to January 23, 2020, the time-to-event distribution of the interval (in days) from symptom onset to confirmation is well fit by a Gamma distribution[1] ($a = 1.85, b = 3.57, LL = -252.24$). Consequently, we adopted a Gamma distribution to model T_d, but tested longer mean periods (ab) as symptom onset often lags the onset of contagiousness.

From February 21 – March 13, 2020 there was no heterogeneity in the parameters β and α across counties. However, after March 13, 2020, we introduced separate estimates for β and α in certain counties in order to represent the variability in contact rates and reporting rates.

For the contact rate, we defined a separate β_i for each of the 16 counties with the most cumulative confirmed cases as of March 24 2020. The contact rates in these counties were scaled based on their population density using the following relation:

$$\beta_i = \frac{0.8 \times log10(PD_i)}{median\big(log10(PD)\big)} \times \beta.$$

Here PD_i is the population density in county i, $median\big(log10(PD)\big)$ is the median value of log-transformed population density among all counties, and β is the contact rate shared by other counties.

For reporting rate, we observed a surge of testing in certain counties in New York State. To account for this surge, we introduced four levels of reporting rate: α_1 defined for New York County NY; α_2 defined for Nassau County NY, Suffolk County NY, Westchester County NY and Orange County NY; α_3 defined for other US counties with cumulative cases >=10 as of March 24 2020; and α_4 defined for counties with cumulative cases lower than 10 as of March 24 2020. On March 13 2020, we set $\alpha_1 = 6\alpha$, $\alpha_2 = 4\alpha$, $\alpha_3 = 2\alpha$ and $\alpha_4 = \alpha$ as priors.

Estimate prior parameters on March 13 2020

To derive an estimate of prior parameters on March 13 2020, we calibrated the transmission model against county-level incidence data reported from February 21,

2020 through March 13, 2020[5]. Specifically, we estimated model parameters using an iterated filtering (IF) framework[6,7]. The metapopulation model is high dimensional with 59,998 subpopulations. We therefore applied an efficient data assimilation algorithm – the Ensemble Adjustment Kalman Filter (EAKF)[8], which is applicable to high dimensional model structures, in multiple iterations to infer parameters β, μ, Z, D, α and θ. This iterated filtering (IF)-EAKF framework has previously been used to infer parameters in a large-scale agent-based model for antimicrobial-resistant pathogens[9], as well as a metapopulation model depicting the spread of SARS-CoV-2 in China[1]. Details of its implementation can be found in Ref. 1.

The prior ranges of model parameters were set as: $\beta \in [0.3, 1.5]$, $\mu \in [0.2, 1.0]$, $Z \in [2, 5]$, $D \in [2, 5]$, $\alpha \in [0.02, 1.0]$, and $\theta \in [0.01,0.3]$. In the inference, we fixed the shape parameter of the Gamma distribution for T_d as $a = 1.85$, and vary the mean value of the distribution. We tested a range of mean T_d values from 6 days to 10 days.

To initialize the model, we seeded exposed individuals (E) and unreported infections (I^u) in counties with at least one confirmed case. Unlike the situation in China, where the outbreak originated from a single city, there was importation to multiple locations in the US that could have initiated community transmission. To reflect this potential ongoing community transmission before the reporting of the first local infection, for each county with confirmed cases, we randomly drew E and I^r from uniform distributions $[0, 12C]$ and $[0, 10C]$ 8 days prior to the reporting date (T_0) of the first case. Here C is the total number of reported cases between day T_0 and $T_0 + 4$.

The rationale for this seeding strategy is as follows. If an average reporting delay of 8 days is assumed, we can estimate I^r on day $T_0 - 8$ by $\frac{C}{5} \times D$, where $\frac{C}{5}$ is the average number of daily cases during the first five days with reporting (T_0 to $T_0 + 4$). If we use the upper bound of the prior for D (i.e., 5 days), I^r is estimated as C, which is also an upper bound. Using parameters obtained from China[1], we assume the mean I^u on day $T_0 - 8$ is $5C$, implying a reporting rate of 1/6=16.7%. Drawing I^u from $[0, 10C]$ leads to a broader prior range of the reporting rate. As both I^r and I^u were evolved from the exposed population E, we draw E from the range $[0, 12C]$. This crude calculation gives us an estimate of seeding in US counties. During inference, this seeding can be adjusted up or down by the filter, and best-fitting models produce simulations that capture observed outcomes.

Model calibration after March 13 2020

Due to changes of control measures and laboratory testing performed, both of which vary on a county-by-county basis as well as through time, certain model parameters (e.g., β and α) are unlikely to remain constant. As a result, instead of using IF-EAKF to estimate a single set of constant parameters, we applied the EAKF to adjust model states and parameters sequentially after March 13. This fluid adjustment of state variables and parameters allows variations over time and grants the system more flexibility to fit observation.

Estimation of hospitalization, ICU admission and mortality

We estimated daily numbers of hospitalizations, ICU admissions and mortality using the probabilities reported in Table 1, which were compiled using data from the Diamond Princess cruise[10]. For each county, we generated average values of these probabilities based on demographic information. In addition, we modeled the duration of events, including hospital length of stay, ICU length of stay, time from onset to hospitalization, time from onset to death and time from onset to ICU using gamma distributions. Parameters for those distributions are reported in Table 2[11-13]. Using the event duration generated from these gamma distributions, we were able to compute the numbers of daily admission and discharge from hospital and ICU beds, and thus represent the daily demand for hospital and ICU beds.

Results

We generated projections of daily demand for hospital and ICU beds, as well as mortality for the 4 weeks after March 24 2020 under three scenarios: 1) no further intervention after March 24 2020, 2) a 25% reduction of contact rates based on the values estimated as of March 24 2020, and 3) a 50% reduction of contact rates based on the values estimated as of March 24 2020.

Projections for national hospital bed demand, ICU demand and mortality are reported in Figs. 1-3. We also display the projections for the same targets in the New York City metropolitan area in Figs. 4-6. The New York metropolitan area in the figures includes 11 counties: Kings County NY, Queens County NY, New York County NY, Bronx County NY, Richmond County NY, Westchester County NY, Bergen County NJ, Hudson County NJ, Passaic County, NJ, Putnam County NY, Rockland County NY.

Model output and movies are posted at: https://github.com/shaman-lab/COVID-19Projection_0324

Interpretation Considerations

Several qualifications with respect to these projections must be noted and considered during interpretation. Firstly, the model is optimized using observations through March 24, 2020; however, those observations, i.e. confirmed cases by county, represent infections that were acquired by individuals 1-2 weeks earlier during a time prior to the implementation of many of the social distancing and isolation measures in place on March 24, 2020. Because of this long delay between infection acquisition and case confirmation, any flattening of the curve due to these effects is not yet apparent in observations nor communicated to the model during optimization. Consequently, the no intervention scenario roughly represents the transmission potential of the virus around March 12th projected forward. As many new control policies were effected after March 12, 2020, the 25% and 50% contact reduction projections may depict paths that some

counties are already following and thus may be more representative of future activity. In this fashion, the 3 scenarios provide references against which the effectiveness of control measures already enacted can be assessed. We also ask readers to note the cones of uncertainty associated with each scenario projection, which broaden into the future.

Secondly, the landscape to which this model has been optimized is highly variable in space and time, due to differences in contact behavior, population density, control measures and testing practices. These differences in space and time make the fitting of any model of this scale challenging. In this shifting landscape, it is very important to reiterate that it will take 10-14 days before the effects of real-world interventions—any flattening of the curve weeks—become apparent.

Acknowledgement

We thank Dr. Wan Yang for compiling the data in Tables 1 and 2.

References

1. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J (2020) Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science,* eabb3221 DOI: 10.1126/science.abb3221.
2. https://www.census.gov/topics/employment/commuting.html
3. Pei S, Kandula S, Yang W, Shaman J. Forecasting the spatial transmission of influenza in the United States. PNAS 115(11): 2752-7 (2018).
4. M. Kramer, D. Pigott, B. Xu, S. Hill, B. Gutierrez, O. Pybus, Epidemiological data from the nCoV-2019 Outbreak: Early Descriptions from Publicly Available Data. Available: http://virological.org/t/epidemiological-data-from-the-ncov-2019-outbreak-early-descriptions-from-publicly-available-data/337 Accessed Feb 24, 2020.
5. USAFACTS https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/ Accessed March 24 2020.
6. Ionides EL, Bretó C, King AA. Inference for nonlinear dynamical systems. PNAS 103(49):18438-43 (2006).
7. King AA, Ionides EL, Pascual M, Bouma MJ. Inapparent infections and cholera dynamics. Nature 454(7206):877-80 (2008).
8. Anderson JL. An ensemble adjustment Kalman filter for data assimilation. Monthly Weather Review 129(12):2884-903 (2001).
9. Pei S, Morone F, Liljeros F, Makse H, Shaman JL. Inference and control of the nosocomial transmission of methicillin-resistant Staphylococcus aureus. eLife 7:e40977 (2018).
10. Russell TW, Hellewell J, Jarvis CI, van-Zandvoort K, Abbott S, Ratnayake R, Flasche S, Eggo RM, Kucharski AJ, CMMID nCov working group. Estimating the infection and case fatality ratio for COVID-19 using age-adjusted data from the outbreak on the Diamond Princess cruise ship. Euro Surveill. 2020;25(12):pii=2000256. https://doi.org/10.2807/1560-7917.ES.2020.25.12.2000256
11. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet. 2020. https://doi.org/10.1016/S0140-6736(20)30566-3
12. Yang X, Yu Y, Xu J, Shu H, Liu H, Wu Y, Zhang L, Yu Z, Fang M, Yu T, Wang Y. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. The Lancet Respiratory Medicine. 2020. https://doi.org/10.1016/S2213-2600(20)30079-5

13. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. JAMA. 323(11):1061-1069 (2020).

Table 1. Probabilities of infected persons being hospitalized, admitted to ICU and dying.

Age group	P(Hosp|Infection)	P(ICU|Infection)	P(Death|Infection)
0-9	0.017	0.006	0.002
10-19	0.0271	0.0078	0.0022
20-29	0.0565	0.0161	0.0022
30-39	0.0754	0.0215	0.002
40-49	0.078	0.0222	0.0032
50-59	0.0601	0.0171	0.0382
60-69	0.0816	0.0232	0.0155
70-79	0.3221	0.0889	0.0262
80-89	0.1267	0.0352	0.0191

Table 2. Parameters of gamma distributions for event duration.

	Mean	SD	Max
ICU length of stay, days	21	5.9	30
Hospital length of stay, days	24	5.2	30
Time from onset to hospitalization, days	7	3	12
Time from onset to death, days	18.5	5.2	30
Time from onset to ICU, days	11	5	20

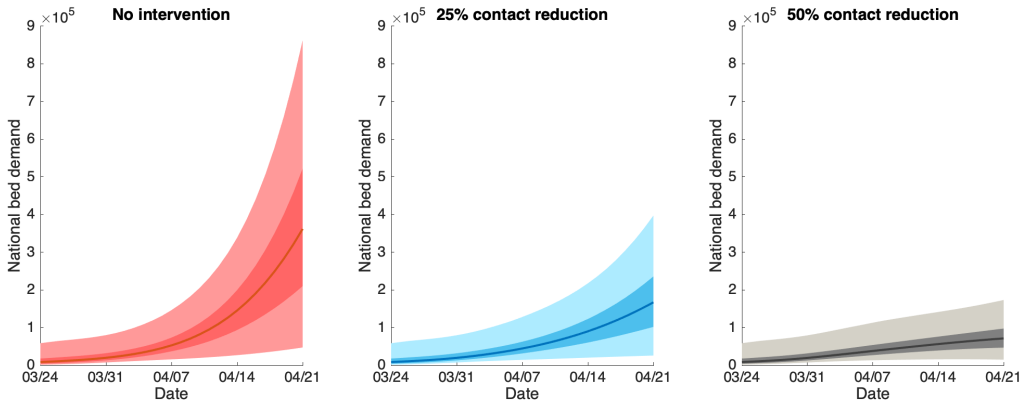

Fig. 1. Projections of hospital bed demand in the continental US.

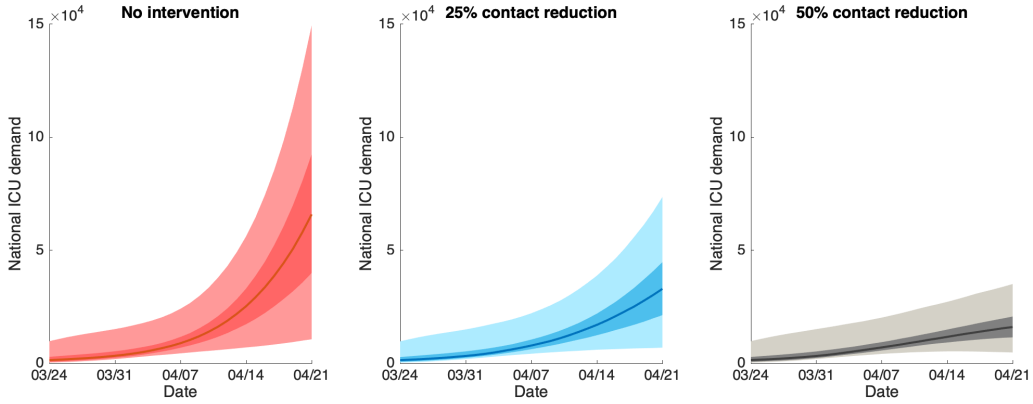

Fig. 2. Projections of ICU demand in the continental US.

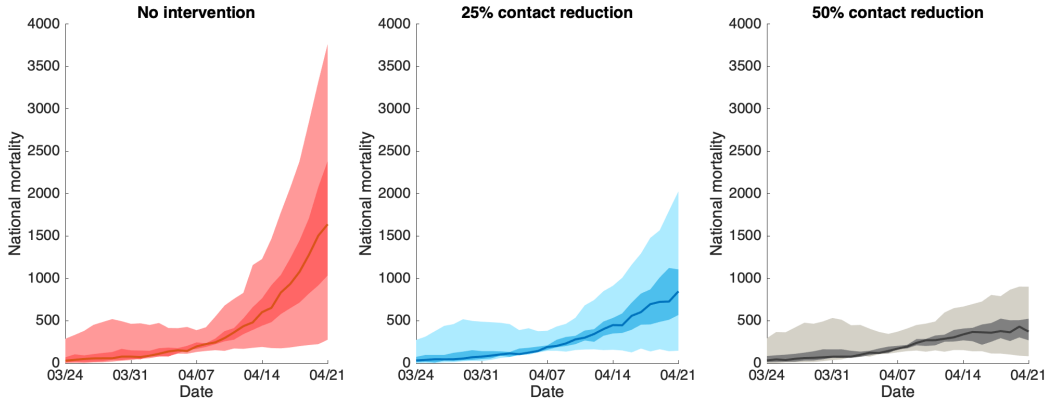

Fig. 3. Projections of mortality in the continental US.

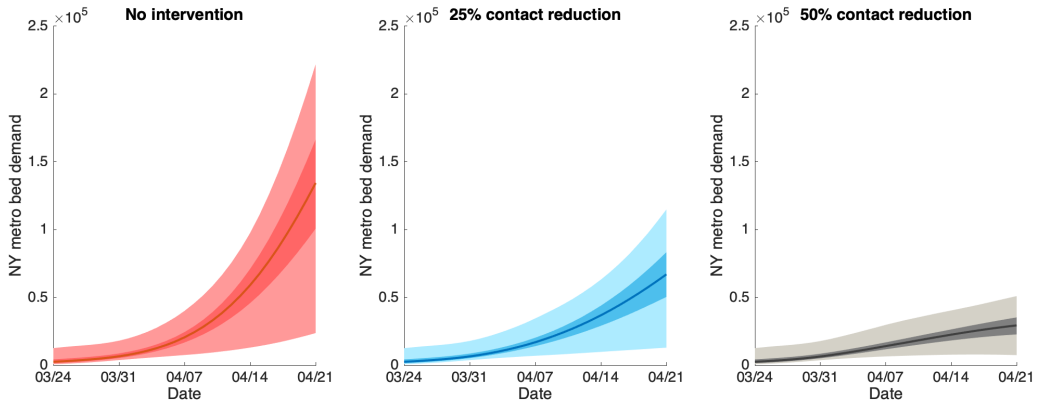

Fig. 4. Projections of hospital bed demand in New York metropolitan area.

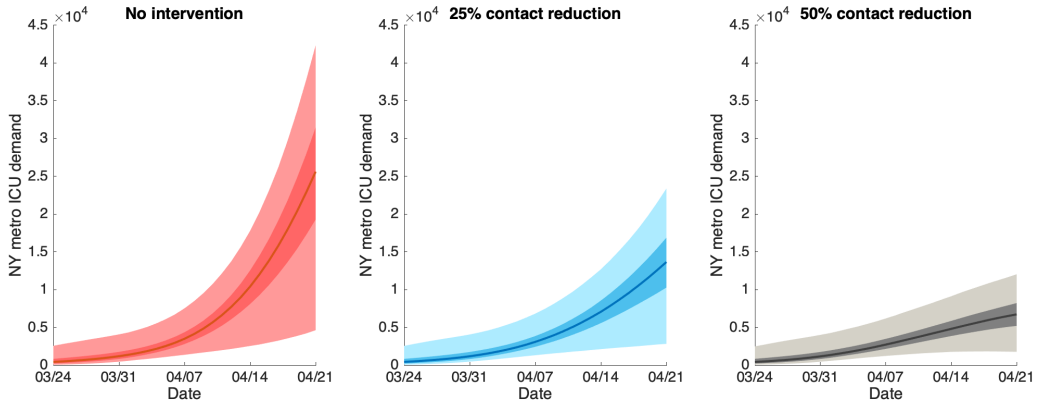

Fig. 5. Projections of ICU demand in New York metropolitan area.

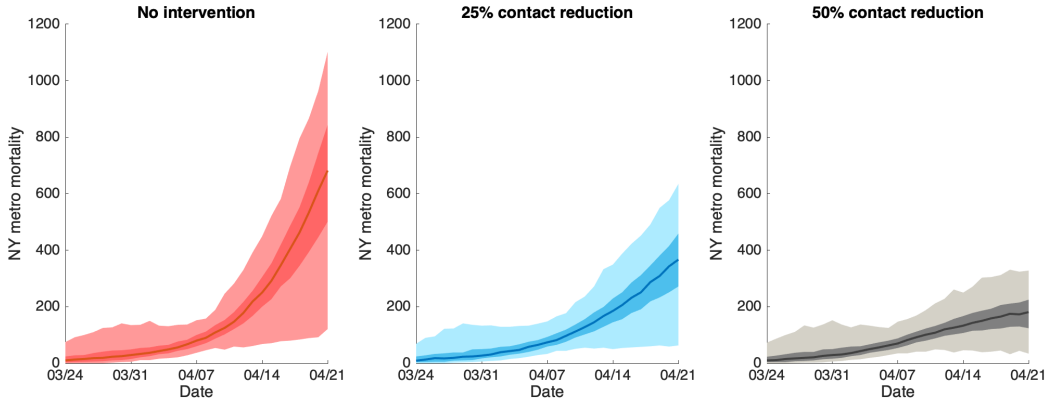

Fig. 6. Projections of mortality in New York metropolitan area.