medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Data-driven Optimized Control of the COVID-19 Epidemics

Afroza Shirin[1,2], Yen Ting Lin[3], and Francesco Sorrentino[1]

[1]Department of Mechanical Engineering, University of New Mexico, New Mexico, 87131
[2]Electrical and Computer Engineering, University of New Mexico, New Mexico, 87131
[3]Information Sciences Group, Computer, Computational and Statistical Sciences Division (CCS-3), Los Alamos National Laboratory, New Mexico, 87544

August 28, 2020

Abstract

Optimizing the impact on the economy of control strategies aiming at containing the spread of COVID-19 is a critical challenge. We use daily new case counts of COVID-19 patients reported by local health administrations from different Metropolitan Statistical Areas (MSAs) within the US to parametrize a simple model that well describes the propagation of the disease in each area. We then introduce a time-varying control input that represents the level of social distancing imposed on the population of a given area and solve an optimal control problem with the goal of minimizing the impact of social distancing on the economy in the presence of relevant constraints, such as a desired level of suppression for the epidemics at a terminal time. We find that with the exception of the initial time and of the final time, the optimal control input is well approximated by a constant, which is specific to each area. For all the areas considered, this optimal level corresponds to stricter social distancing than the 'current' level estimated from data. Proper selection of the time period for application of the control action optimally is important: depending on the particular MSA this period should be either short or long or intermediate. Finally, we see that small deviations from the optimal solution can lead to dramatic violations of the constraints.

The fast propagation of the COVID-19 pandemic has attracted unprecedented attention from both the public and the scientific community. Due to the high fatality rate of SARS-CoV-2 [32], governments throughout the world have adopted measures such as lock-down, stay-at-home, and shelter-in-place, which in turn have led to substantial economic losses, see e.g., [22]. In many countries control interventions have been articulated in phases, usually phase 1 to phase 3, with higher phases corresponding to progressively lower restrictions [12]. A fundamental challenge is to balance the need to suppress the spread of COVID-19 and the need to contain the economic impact of measures aiming at limiting the spread of the disease. In this manuscript, we apply optimal

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

control theory on a mathematical model for the propagation of the epidemics, parametrized by real-world data describing different regions, and compute control strategies which are optimal for each region from an economic standpoint.

A number of papers and reports have focused on both modeling and controlling the pandemic. Flaxman et al. [10] looked at the effect of non-pharmaceutical interventions including school closures, banning of mass gathering, social distancing, etc. on the reproductive number R_0 of COVID-19. Sanche et al. [36] used a mathematical model with data on individual cases, real-time human travel, and infections, as well as estimated epidemiology parameters to compute R_0 and found that it is higher than initially estimated. Chang et al. [8] adopted an agent-based model to determine the efficacy of several intervention strategies on the spread of COVID-19 in Australia. Anderson et al. [6] performed the Bayesian analysis to estimate the impact of social distancing on number of reported cases and hospitalizations in British Columbia and found that when 78% participation in social distancing has been accomplished the cases would decrease; it also noted that if the participation were below 45%, an exponential growth would restart. Morris et al. [25] explored COVID-19 intervention methods which are robust to implementation errors and found that these methods in conjunction with optimal time-limited methods derived from the standard SIR model can be used to mitigate the spread of the virus. Another study analyzed an open-loop optimal control policy updated weekly using real-world feedback, and found that this method is effective in reducing fatalities even if some measurements are inaccurate [16]. A study published in March 2020 estimated the ICU occupancy and ventilator use from a statistical model under the conditions of social distancing and found that the demand for hospital beds and ventilators will exceed the current supply [9]. Another study explored optimal policies for decreasing economic cost and mortality rates from a multi-risk SIR model and found that strict lock-down policies which specifically target the elderly population were most effective in minimizing deaths and economic losses [5]. Previous work has not computed optimal controls for data driven models. On the one hand, Refs. [8], [25], [16], [5] explicitly compute optimal control strategies, but their models were stylized and not parametrized/calibrated by data; On the other hand, Refs. [10] [36], [6], [9] used data to parametrize the models, but the models do not have controllers which can be used to infer optimal control strategies.

In what follows, we first construct a mathematical model, which is parametrized by historical and regional daily new case report. After parametrization, the data-driven model is capable to reproduce the regional progression of the COVID-19 epidemics up to the present. Then, we apply optimal control theory to the parametrized model to compute an optimal control strategy over the course of a pre-determined period into the future to suppress the epidemics to a desired level, while minimizing economic costs. This type of approach is suitable for long-term planning (i.e., over the course of several months) as opposed to short-term planning, which can be difficult to implement by the government and by businesses.

While it appears likely that an effective vaccine for SARS-CoV-2 is under way, most estimates indicate that the vaccine will be distributed to the population in summer or fall 2021 [38], which points out the need for planning interventions to contain the epidemics for several months to come in the absence of a vaccine. Thus our proposed workflow aims to bridge the gap until the time T_{vac} when an effective vaccine is discovered, massively manufactured, and administered to the majority of the population. Another temporal consideration regards the

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

time T_{herd} at which a population achieves herd immunity in the absence of a vaccine and without overflowing the medical facilities. While herd immunity from COVID-19 is the subject of much ongoing discussion [31], in the Methods we provide a rough estimate of T_{herd} from available data. In this paper we proceed under the assumption that the inference period T_{inf} and the control horizon T_{cont} are such that $T_{inf} + T_{cont} < T_{vac}$ and $T_{inf} + T_{cont} < T_{herd}$.

We set out the analysis by first introducing a compartmental model which describes key features of the COVID-19 epidemics. The whole population is divided into the following compartments. The susceptible population compartment (S) includes the people who can contract the pathogen SARS-CoV-2. The exposed population (E) are those who have been infected but have not progressed long enough into the disease to transmit it to susceptible people. Those who can transmit the disease ('carriers') are divided into the asymptomatic group (A) who do not show symptoms and the infected symptomatic groups (I) who show symptoms. Both the symptomatic and asymptomatic groups can transmit the disease, but with different infectiousness—the asymptomatic people are less infectious. The infected symptomatic population I is divided into three sub-compartments. The first sub-compartment I_{sq} includes those who just self quarantine and do not get tested. The second sub-compartment I_{tp} includes those who get tested, result positive, and get quarantined. In contrast to the previous two sub-compartments, the third sub-compartment I_s includes the rest of the symptomatic population, who do not change their behavior until they recover (or decease). Those who were tested positive and those who decided to self-quarantine are moved into a quarantined compartment (Q), and stop interacting with other populations. The removed compartment (R) includes those who are completely recovered from the disease and have acquired immunity, and those who have died because of the disease. Both groups are not susceptible to reinfection and are removed from the system.

Mathematically, the time evolution of the population density—defined as the compartmental population normalized by the total regional population—of each compartment is governed by the following set of coupled ordinary differential equations,

$$\dot{S}(t) = -\beta P^2(t)\, S(t)\left[I_s(t) + I_{sq}(t) + I_{tp}(t) + \mu A(t)\right] \tag{1a}$$

$$\dot{E}(t) = \beta P^2(t)\, S(t)\left[I_s(t) + I_{sq}(t) + I_{tp}(t) + \mu A(t)\right] - \lambda E(t) \tag{1b}$$

$$\dot{A}(t) = \lambda(1-\sigma)\, E(t) - \gamma_A A(t) \tag{1c}$$

$$\dot{I}_{sq}(t) = p_{sq}\lambda\sigma E(t) - \left[\gamma_I + \gamma_{sq}\right] I_{sq}(t) \tag{1d}$$

$$\dot{I}_{tp}(t) = p_{test}\lambda\sigma E(t) - \left[\gamma_I + \gamma_{tp}\right] I_{tp}(t) \tag{1e}$$

$$\dot{I}_s(t) = (1 - p_{sq} - p_{test})\,\lambda\sigma E(t) - \gamma_I I_s(t) \tag{1f}$$

$$\dot{Q}(t) = \gamma_{tp} I_{tp}(t) + \gamma_{sq} I_{sq}(t) + p_{sq}\lambda\sigma E(t) - \gamma_I Q(t) \tag{1g}$$

$$\dot{R}(t) = \gamma_A A(t) + \gamma_I\left[I_s(t) + I_{tp}(t) + Q(t)\right], \tag{1h}$$

where β is the transmissibility, λ is the transition rate from the exposed compartment to either the asymptomatic or symptomatic compartments, γ_I is the transition rate from the infected compartment to the recovered compartment, γ_A is the transition rate from the asymptomatic compartment to the recovered compartment, μ is the

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

relative infectiousness of asymptomatic individuals (compared to symptomatic individuals), σ is the fraction of exposed people who transition into the symptomatic compartments, p_{sq} is the fraction of symptomatic people who will self-quarantine, γ_{sq} is the transition rate from from the I_{sq} compartment into the quarantined compartment, γ_{tp} is the transition rate from the I_{tp} compartment into the quarantined compartment, and p_{test} is the fraction of infected who get tested (but only a fraction of them will be able to get a test and be confirmed as a positive, I_{tp}). The model assumes that testing resources are not scarce, i.e., availability of a testing kit to every person in I_{tp}; we also assume positive people are identified as such with 100% accuracy. The case with limited testing resources is discussed in the SI. Realistically, the time scale associated with γ_{test} can be several days, so we assume $\gamma_{test} = 0.5$ (2 days). We model social distancing by the control variable $0 \leq P(t) \leq 1$, which measures the reduction of contact probabilities between the susceptible and the infectious populations (which include both A and I). The model is structurally similar to the models analyzed in Refs. [6, 21], but simplified to allow for efficient calculations of optimal control solutions. Figure 1 illustrates a schematic diagram of the model.

In order to reduce the dimensionality of the dynamical system, we introduce the following simplification: we treat γ_{sq} as a very large number; for $\gamma_{sq} \to \infty$, we assume that those becoming symptomatic immediately transition into the Q compartment, yielding the following reduced-order model:

$$\dot{S}(t) = -\beta P^2(t) S(t) \left[I_s(t) + I_{tp}(t) + \mu A(t)\right] \tag{2a}$$

$$\dot{E}(t) = \beta P^2(t) S(t) \left[I_s(t) + I_{sq}(t) + I_{tp}(t) + \mu A(t)\right] - \lambda E(t) \tag{2b}$$

$$\dot{A}(t) = \lambda (1-\sigma) E(t) - \gamma_A A(t) \tag{2c}$$

$$\dot{I}_{tp}(t) = p_{test} \lambda \sigma E(t) - \left[\gamma_I + \gamma_{tp}\right] I_{tp}(t) \tag{2d}$$

$$\dot{I}_s(t) = (1 - p_{sq} - p_{test}) \lambda \sigma E(t) - \gamma_I I_s(t) \tag{2e}$$

$$\dot{Q}(t) = \gamma_{tp} I_{tp}(t) + p_{sq} \lambda \sigma E(t) - \gamma_I Q(t) \tag{2f}$$

$$\dot{R}(t) = \gamma_A A(t) + \gamma_I \left[I_s(t) + I_{tp}(t) + Q(t)\right]. \tag{2g}$$

The schematic diagram which fits the above model in Eq. (2) is shown in the Supplementary Note 1 1. To bridge the model and the data, we also integrate an auxiliary variable C_I which counts the cumulative confirmed cases and evolves according to

$$\dot{C}_{tp}(t) = \gamma_{tp} I_{tp}(t), \tag{3}$$

We will fit $\Delta C_{tp}(t) := C_{tp}(t+1) - C_{tp}(t)$ to the new case counts reported on each day, detailed in the section on parametrization below.

From an economic point of view, the measures of social distancing $P(t)$ and quarantining $Q(t)$ have very elevated costs for society. For example, the US real gross domestic product (GDP) dropped by roughly one third from year to year in the second quarter of 2020 [22], due mainly to the COVID-19 pandemic, see e.g., [26].

We thus formulate the following optimal control problem,

$$\min_P J = c_p \int_{t_i}^{t_f} \frac{1 - P(t)}{P(t)} dt + c_q \int_{t_i}^{t_f} \frac{Q(t)}{1 - Q(t)} dt, \tag{4}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

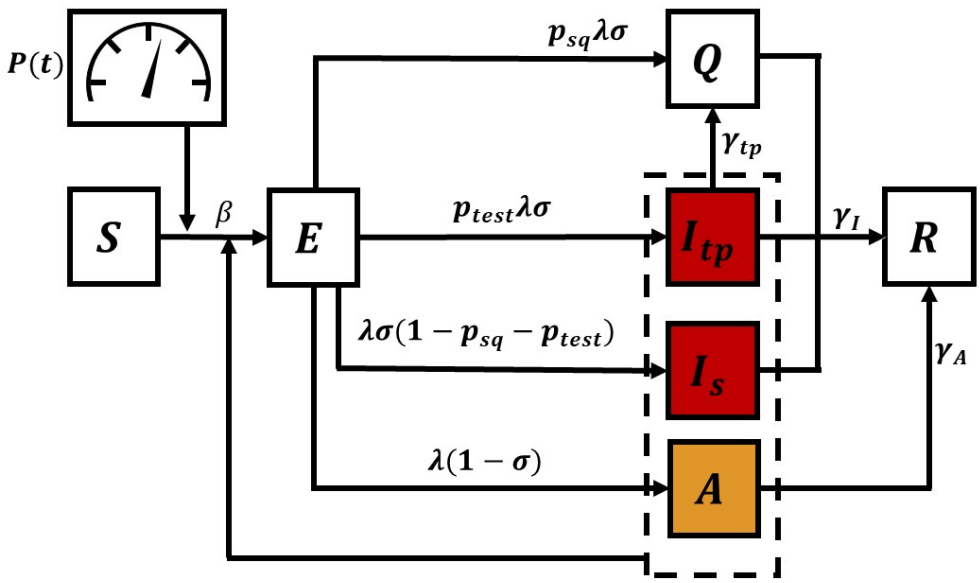

Figure 1: **Compartmental model corresponding to Eq. 2.** The transition from the S (susceptible) compartment to the E (exposed) compartment is affected by the population densities $S(t)$, $I_{tp}(t)$, $I_s(t)$, $A(t)$ and by the time-varying control input $P(t)$.

$c_p, c_q \geq 0$, subject to Eq. (2), the initial conditions, the terminal constraint

$$E(t_f) + I_s(t_f) + I_{tp}(t_f) + A(t_f) < \epsilon, \tag{5}$$

and the following path constraint

$$I_s(t) + I_{tp}(t) \leq I_{\max}, \tag{6}$$

where $I_{\max}$ is an upper limit on the number of infected people that can receive proper treatment in the hospitals. t_i is the time at which we perform the inference and start optimizing the control action, t_f is the final time of the optimization, the previously defined control horizon $T_{cont} = (t_f - t_i)$.

The objective function (4) takes into account an economic cost for social distancing with an appropriate coefficient $c_p > 0$ and an economic cost for quarantining with an appropriate coefficient $c_q > 0$. We set the cost associated with social distancing per unit time to be modeled by $(1 - P(t))/P(t)$. By choosing this functional form, the cost is linearly dependent on the scale of the reduction $(1 - P(t))$ when $1 - P(t) \ll 1$, and nonlinearly diverges near total shut-down ($P(t) \ll 1$). The economic cost increases with $1 - P(t)$ to indicate that stricter measures of social distancing affect more and more 'essential' workers, and so increasingly larger parts of the economy. For example, the cost of limiting large gatherings of people like concerts or sport events is lower than

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the cost of limiting customers' access to stores and restaurants. Similarly, imposing quarantine requires resources that are linearly proportional to the quarantined population when Q is small, and diverge nonlinearly when the quarantined population is close to the entire population $Q = 1$. We model such a cost by the functional form $Q(t)/\left(1 - Q(t)\right)$.

Both social distancing and quarantining have a cost associated with the lack of economic return generated by limiting person-to-person interactions. It is reasonable to assume $c_q \geq c_p$, as strict quarantining requires supervision costs as well as costs due to lowered productivity [13] while social distancing only incurs costs due to lowered productivity [17].

We will also consider the alternative objective function,

$$\min_P J^{\text{alt}} = c_p \int_{t_i}^{t_f} (1 - P(t))dt + c_q \int_{t_i}^{t_f} Q(t)dt, \tag{7}$$

for which the integrand functions are linear in $P(t)$ and $Q(t)$. While the formulation (7) is mathematically simpler, it does not take into account the fact that stricter measures of social distancing may result in progressively larger economic losses. The alternative objective function (7) is here mainly introduced in order to compare the results to those obtained with (4).

The tunable parameter ϵ represents the desired suppression level for the epidemics at the final time t_f. In general, selected values for ϵ may depend on a number of factors, such as the time horizon over which optimization is performed and the particular stage of the epidemics (initial exponential growth, intermediate growth, or plateau.) A possibility is to require complete eradication of the epidemics, which corresponds to setting $\epsilon = 1/N$, where N is the number of individuals in the population. However, the high basic reproductive number of COVID-19 makes eradication unlikely; instead we set ϵ to a small number indicating suppression of the disease. The other constraint given by $I_{\max}$ represents the need to contain the infected population below a given threshold at all times (or 'flatten the curve'.) In what follows, unless differently stated, we set the terminal constraint $\epsilon = 10^{-5}$, which corresponds to imposing that the number of infected people is below one person per 100000 population. This number is derived from the guidelines of European countries about re-opening, see for example [24], indicating that reopening occurred at about 2×10^{-5}. Also, European countries have official guidelines for reimposing stricter lock-down measures, see: [11], which sets a critical population equal to $50/100000 = 5 \times 10^{-4}$. The values of I_{max} are set regionally based on the capacity of the ICU beds in different metropolitan areas and are summarized in Table 3.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Parametrization of the model for different US metropolitan areas.

We partitioned the model parameters into two sets: a set of fixed parameters and a set of inferred parameters which are estimated by the daily case counts reported by regional health administration and registered in the repository curated by the New York Times [3].

The fixed parameter sets includes S_0, λ, γ_I, γ_A, μ, σ, p_{test}, and p_{sq}. S_0 is the regional total population, and we used the US Census Bureau-estimated regional population of each of the Metropolitan Statistical Areas (MSAs) or 'cities', which are delineated by the US Office of Management and Budget [4]. Lauer et al. [19] estimated the median of the incubation period to be about 5.2 days, however, there is evidence that patients become infectious roughly two days before the onset of symptoms [27], which corresponds to approximately a three-day progression into *contagiousness*. We thus set a rough estimation for λ to be 1/3 (d). In a more complex model with multiple stages of the disease progression [21], one could account for the fact that a patient can be both pre-symptomatic and infectious. The coefficients γ_I and γ_A, which are the transition rate to recovery of the symptomatic and asymptomatic populations, are estimated to be 0.12 (1/d) [43] and 0.26 (1/d) [35]. The relative infectiousness μ is set at 0.9 [42], and the fraction of the symptomatic population is set at 0.64 [2, 35]. The parametrization of γ_I, γ_A, μ, and σ are consistent with a more complex model which was used to perform daily forecasts of the disease spread [21]. We assume $p_{test} = 0.25$ and $p_{sq} = 0.4$, noting that these parameters were able to reproduce the infected population at the time of the inference (we estimated that about 15 to 20% of the total population infected in the New York City MSA on 07-Jul-2020 when we parametrize the model.)

We used the data from 21-Jan-2020 to 07-Jul-2020 to infer the inferred parameters, which corresponds to the previously defined inference period T_{inf}. We assume that the social distance function $P(t)$ before 07-Jul-2020 is piecewise-linear to avoid over-fitting (due to observation noise). At the time when the analysis was performed, multiple MSAs had shown clear second-phase resurrection of the epidemics [21, 1]. We found that a two-phase piecewise linear function is sufficient to reproduce the data of each of the MSAs:

$$P(t) = \begin{cases} 0, & t \leq t_0, \\ 1, & t_0 < t \leq t_1 \\ 1 + \frac{p_1 - 1}{t_2 - t_1}(t - t_1), & t_1 < t \leq t_2 \\ p_1, & t_2 < t \leq t_3 \\ p_1 + \frac{p_2 - p_1}{t_4 - t_3}(t - t_3), & t_3 < t \leq t_4 \\ p_2 & t_4 < t < t_i, \end{cases}$$

where t_0 is the time when the disease was introduced into a specific MSA, the monotonic t_1, t_2, t_3, and t_4 ($t_j \leq t_{j+1}$) are the times when the social-distancing behavior changes, and p_1 and p_2 are the two social-distancing strengths. We define $\Delta t_j := t_j - t_{j-1}$ for $j = 1, \ldots 4$. The time at which we perform the inference is t_4, which is also the time after which optimization of the control action begins. The two-phase piecewise linear model is the minimal model that we found capable to reproduce the data, and can be validated by more rigorous model-selection procedure detailed in [21].

To fit the model by the noisy daily report new counts, we adopt a negative-binomial noise model. We brief the procedure here, noting that the procedure is similar to the inference method detailed in [21]. Given a set

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

of parameters θ (a stylized notation of the set of the inferred model parameters), the model Eq. (2) predicts a deterministic trajectory of the *new* positive tested case on day j, $\Delta I_{tp}(j;\theta)$. This deterministic prediction is interpreted as the mean of a stochastic outcome, modeled by a negative binomial distribution $\mathrm{NB}(r, p_j)$ where r and p_j are the parameter of the distribution, and p_j is constrained by the deterministic model prediction

$$p_j = \frac{r}{r + \Delta C_{tp}(j;\theta)}. \tag{8}$$

Here, the r is the *dispersion* parameters which describes how disperse the noise distribution is; in the limit $r \to \infty$, the negative binomial statistics converges to Poissonian, and in the limit $r \to 0$ the distribution looks closer to an exponential. The negative binomial noise model is phenomenological: it has the capability to capture a wide variety of single-modal distributions. With the negative binomial noise model, the likelihood function given a set of N daily reported new case counts $\{\Delta C_{tp}(j)\}_{j=0}^{N-1}$ can be formulated [21]:

$$\mathcal{L}\left(\theta; \{\Delta C_{tp}(j)\}_{j=0}^{N-1}\right) \equiv \prod_{j=0}^{N-1} \begin{pmatrix} \Delta C_{tp}(j) + r - 1 \\ \Delta C_{tp}(j) - 1 \end{pmatrix} p_j^r \left(1 - p_j\right)^{\Delta C_{tp}(j)}. \tag{9}$$

In summary, the inferred parameters θ include the regional-specific disease transmissibility β, onset of the disease spread t_0, behavioral switching times t_1, t_2, t_3, t_4, the strengths of two social-distancing episodes p_1, p_2, and a dispersion parameter r of the negative binomial noise model. These parameters were inferred by the daily case reports from 21-Jan-2020 to 07-Jul-2020. With the formulated likelihood function (9), we used the standard Markov chain Monte Carlo procedure (MCMC) with an adaptive sampler ([7], detailed in [21]) to estimate the maximum likelihood estimator of the parameters θ for each of the interested MSA's.

Table 1 summarizes the set of model parameters that were estimated for each MSA.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

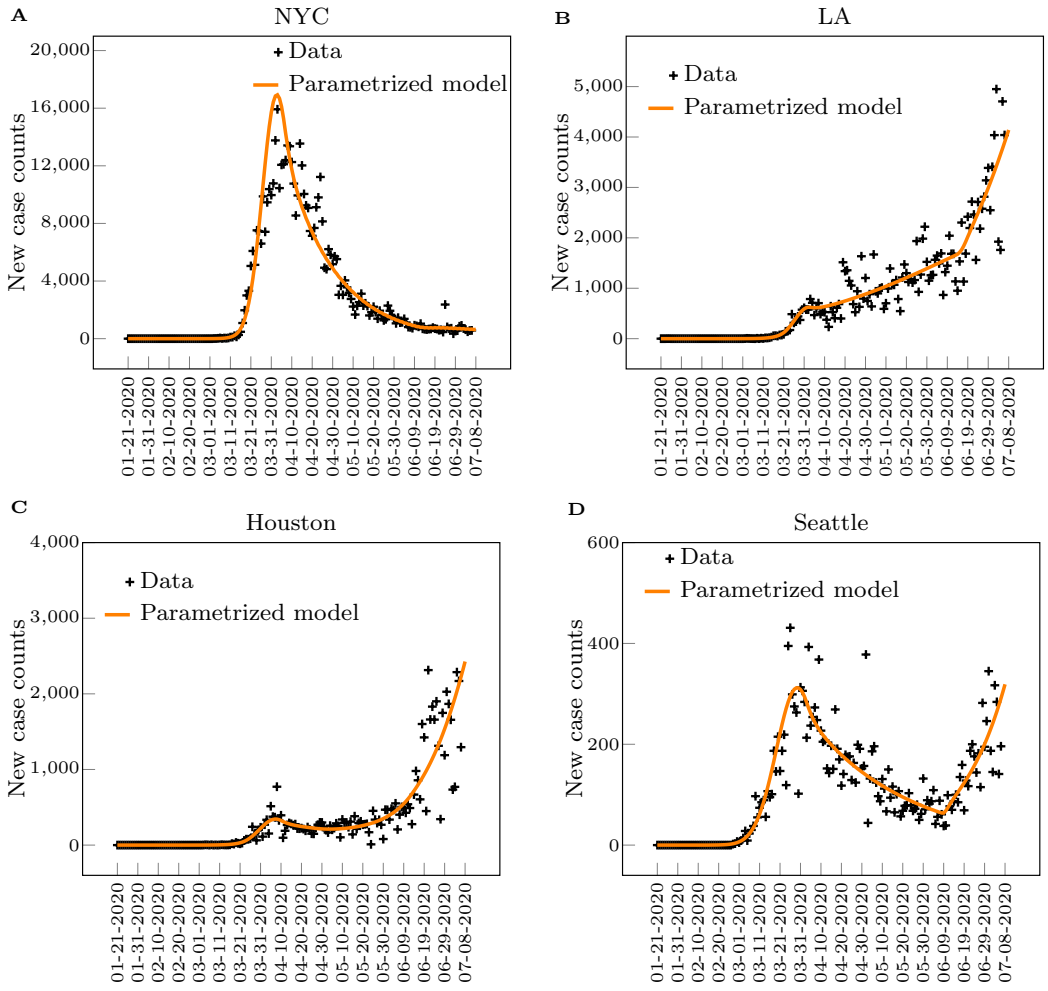

Figure 2: **New case counts from January 21, 2020 to July 8, 2020 in four Metropolitan Statistical Areas within the US.** Plus signs are daily new case counts reported by local health administrations. The solid line is the daily new cases obtained by integrating Eq. (2) after parametrization of the model.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 1: Parameters by Metropolitan Statistical Area (MSA)

Parameter	New York	Los Angeles	Houston	Seattle
S_0	1.92E+07	1.321E+07	7.066E+06	3.979E0+6
λ	3.333E-01	3.333E-01	3.333E-01	3.333E-01
γ_I	1.200E-01	1.200E-01	1.200E-01	1.200E-01
γ_A	2.60E-01	2.60E-01	2.60E-01	2.60E-01
μ	9.00E-01	9.00E-01	9.00E-01	9.00E-01
σ	5.60E-01	5.60E-01	5.60E-01	5.60E-01
p_{test}	2.50E-01	2.50E-01	2.50E-01	2.50E-01
p_{sq}	4.00E-01	4.00E-01	4.00E-01	4.00E-01
T	2.562E+04	1.762E+04	9.422+03	5.306E+03
D	7.117E+03	4.894E+03	2.617E+03	1.474E+03
r	1.183E+01	1.100E+01	5.287	1.067E+01
β	1.806	1.204	1.123	1.442
t_0	2.911E+01	2.865E+01	3.195E+01	2.339E+01
Δt_1	2.557E+01	3.237E+01	3.153E+01	1.388E+01
Δt_2	2.148E+01	1.153+01	1.633E+01	3.549E+01
Δt_3	6.423E+01	7.171E+01	1.685E+01	6.614E+01
Δt_4	8.800	4.289	4.658E+01	2.016
P_1	3.693E-01	5.371E-01	4.845E-01	4.160E-01
P_2	4.403E-01	5.968E-01	6.391E-01	5.555E-01

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Results: Optimal Control Solutions

Following previous work by the authors [40, 39], we use pseudo-spectral optimal control (see Supplementary Note 7) to compute solutions for the problem formulated in Eqs. (2),(4),(5),(6). We set $t_i = 169$ (days) and $t_f = 259$, corresponding to a ninety day control horizon, $(t_f - t_i) = 90$. We parametrize the solutions in $I_{\max}$, ϵ and c_q, after setting without loss of generality $c_p = 1$.

We focus on four different US metropolitan statistical areas: New York City (NYC), Los Angeles (LA), Houston, and Seattle. NYC is the largest US metropolitan area; it emerged as the main early hotspot of the epidemics in the US in March and April, but since early June has achieved stable control of the epidemics. LA is the second largest metropolitan area in the US, it has seen a steady rise in the number of cases by the time when we performed the inference. Houston has seen a rapid increase of the cases in June and July. Seattle was also a very early hotspot, which has seen a decrease in the number of cases in April and May, followed by an increase in June and July. We chose these four MSA's to cover a wide range of different dynamics before the time of the inference.

The solution of the optimal control problem is the function of time $P^*(t)$ that minimizes the objective function (4) subject to the constraints (2), (5), and (6). Different from the observation period $t \leq t_i$, for which we set $P(t)$ to be a piecewise linear function, in the optimization period $t \in [t_i, t_f]$ we let $P(t)$ be an arbitrary function of t, for the purpose of computing the optimal control solution.The optimal control solutions that we obtain for different MSAs are shown in Figs. 3. These solutions are robust to parameter variations (such as different values of the coefficient c_q, see the Supplementary Note 2) and also qualitatively consistent for different MSA's. Robustness is also found with respect to the choice of the constraint I_{max}, see the Supplementary Note 2. Typically we first observe a quick drop in $P^*(t)$ (initial tightening), followed by a long nearly constant trend at P_s^* (steady social distancing), and by another drop near the final time (final stranglehold). We remark that P_s^* corresponds to the level of reduction needed to suppress the time-varying reproduction number $R_t \approx 1$, and the cost per day associated with this level of reduction is $(1 - P_s^*)/P_s^*$. It follows that for each MSA, there is an almost constant value of P_s^* which well describes the optimal solution, except for the initial time and the final time. The value of P_s^* appears to be approximately the same for LA, Houston, and Seattle, while NYC has typically a slighter higher value of P_s^*. This also implies that the optimal cost function J^* is lower for NYC than for the other cities. In all the four metropolitan areas, P_s^* is lower than the 'current' value of $P(t)$ estimated from data, as can be seen from the initial dip in $P^*(t)$. NYC has the smallest initial drop, indicating that the control action at the time at which we performed the inference is the closest to optimal, followed by (in the order) Seattle, LA, and Houston. For each metropolitan area, $I_{\max}$ was estimated from available data about ICU beds for different US states, as shown in Table 3 in the Methods.

The most important parameter of the optimal control problem is the terminal suppression constraint ϵ, which describes the level at which one is trying to suppress the epidemic. The lower is ϵ, the closer the goal is to eradication of the disease. The optimal value attained by the objective function J^* versus ϵ is shown in Figs. 4(A) and (B). It should be noted that for each value of ϵ, J^* shown in Fig. 4 is the lowest possible cost that can be attained. For all the US cities considered, this lowest possible cost increases dramatically as ϵ is reduced, which

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

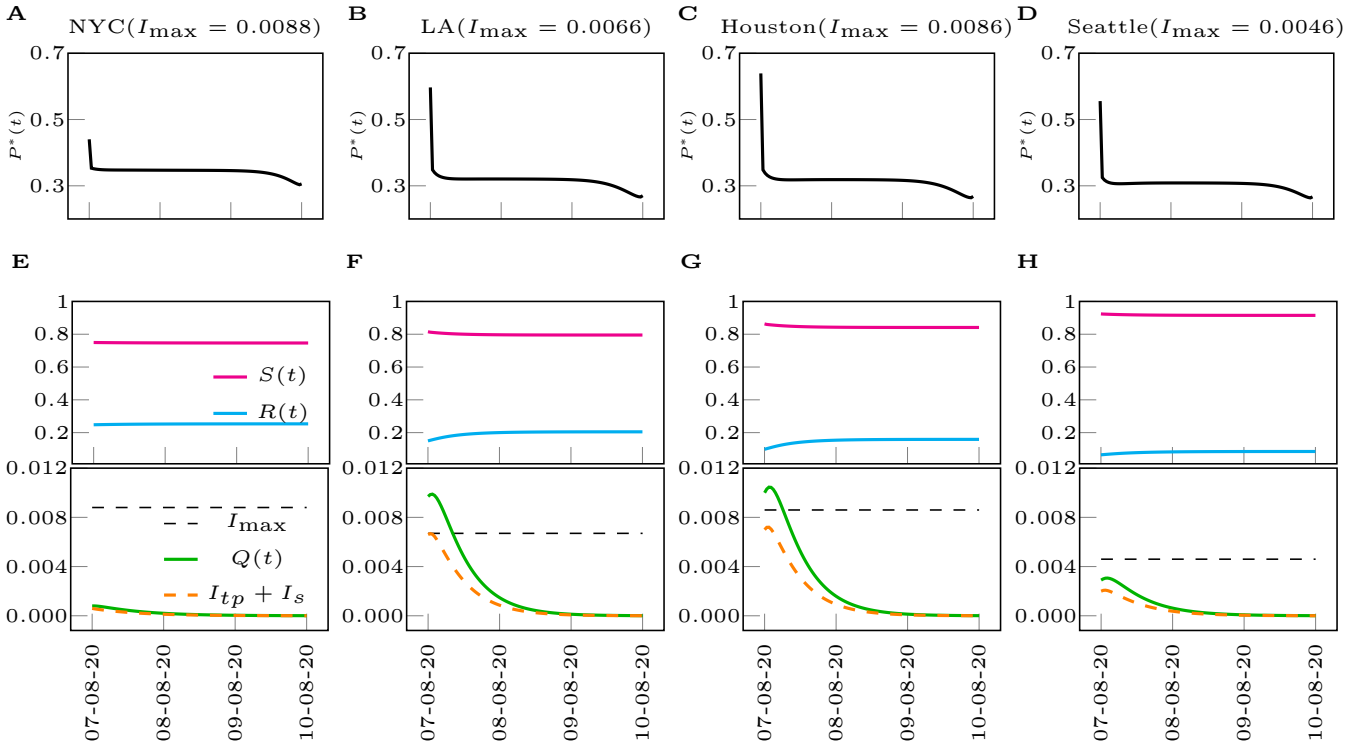

Figure 3: **Optimal control solutions for different Metropolitan Statistical Areas within the US.** (A-D) Optimal control inputs for the metropolitan cities of NYC, LA, Houston, Seattle (from left to right). (E-H) Time evolution of the states subject to the optimal control inputs. $I_{\max}$ are chosen as the minimum of the range in Table 3 ($\rho = 2/3$) and $c_p = c_q = 1$. The legends in (F-H) are same as the legend in (E).

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

exemplifies the dilemma between saving human lives and saving the economy. Again, here we are assuming that the measures of social distancing are optimally implemented, while the cost would be higher in case of non-optimal implementation (discussed in the next section). For large values of the suppression constraint ϵ, in all four cities considered, a different type of solution characterized by low cost emerges, which is discussed in more detail in what follows. Qualitatively similar results were obtained when the control input was chosen that minimizes the alternative objective function (7). A complete study of the effects of varying the suppression constraint parameter ϵ can be found in the Supplementary Note 3. There are deeper implications behind Figs. 4 (A) and (B), namely setting a larger value of ϵ corresponds to delaying the economic cost in time, rather than removing it.

From Fig. 4 (C) and (D) we also see the effects of changing the control horizon $(t_f - t_i)$ of the optimal control problem, from a minimum of 60 days to a maximum of 120 days. We see here that different cities behave differently. For Houston and LA we see that a longer t_f corresponds to a lower value of the optimal cost J^*, while for NYC the cost increases with t_f. A complete study of the effects of varying t_f can be found in the Supplementary Note 4. In particular, we see that increasing t_f has two contrasting effects on the objective function. On the one hand the cost is integrated over a longer time period, on the other hand the longer time period can be exploited to allow for less stringent measures of social distancing at any time (larger values of P^*), which may result in a lower value for the objective function overall. Thus it appears that finely tuning the time horizon of the objective function may be used to critically and selectively affect different cities. The implications are particularly significant for those cities, Houston and LA, that seem to need a longer time period to suppress the epidemics. For Seattle we see that J^* is minimized for intermediate values of t_f in the interval $[255, 285]$, indicating a specific advantage of choosing such a control horizon, see also Fig. 5. A remarkable observation is that by comparing the left plots and the right plots in Fig. 4, it is evident that the optimal control solution is quite independent of the particular form of the objective function (either J or J^{alt}.) The reason for this is that when the integrand in (4) is nonlinear the optimal solution maintains $P^*(t)$ as close as possible to the linear regime (high values of P^*), which is why we do not see much difference with the linear case (7).

In general, the main constraints of the problem are provided by $I_{\max}$ and ϵ. However, for each case considered, typically either one of these two constraints is dominant. In all the simulations shown in Figs. 3, the dominant constraint is provided by ϵ, with $I_s(t) + I_{tp}(t)$ remaining well below I_{max} over the entire period $[t_i, t_f]$. These solutions, which we will refer to as the type-1 solution, are characterized by strong measures of social distancing (low P^*) throughout the whole time interval considered, and the dominant constraint is to achieve suppression of the epidemics at the prescribed terminal time (t_f).

We have also seen the emergence of different solutions, which we refer to as the type-2 solution, when the dominant constraint is given by $I_{\max}$. In these solutions there is at least one time t at which the constraint (6) is satisfied with the equal sign. Overall, type 2 solutions are less expensive economic-wise than type 1 solutions, i.e., the value of J^* is lower. There are also cases when the optimal solution is actively affected by both constraints. In order to better understand the transition between type 1 and type 2 solutions, we have investigated the optimal control problem (4), for the cases of LA, NYC, and Seattle, as both the suppression parameter ϵ and the control horizon $(t_f - t_i)$ are varied. The transition is characterized by a gradual change in J^*, high for type

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

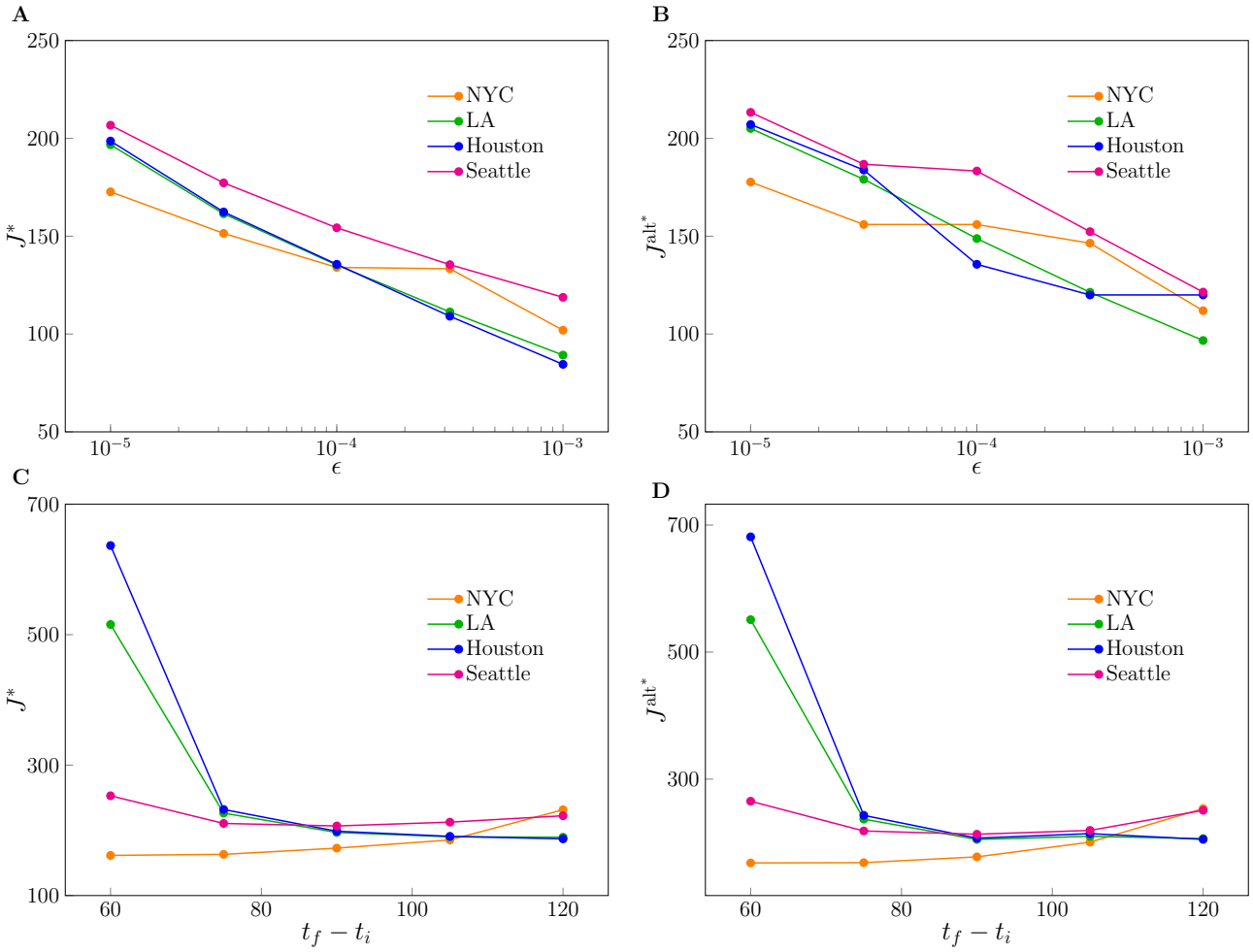

Figure 4: **Effects of varying the terminal suppression constraint ϵ and the control horizon $t_f - t_i$.** (A) The optimal cost J^* obtained as the solution of the optimization problem (4) vs. the parameter ϵ. (B) The optimal cost J^{alt^*} obtained as the solution of the optimization problem (7) vs. the parameter ϵ. (C) The optimal cost J^* obtained as the solution of the optimization problem vs. the control horizon $(t_f - t_i)$. (D) The optimal cost J^{alt^*} obtained as the solution of the optimization problem (7) vs. the control horizon $(t_f - t_i)$. The parameter c_q and c_p are both set to 1.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

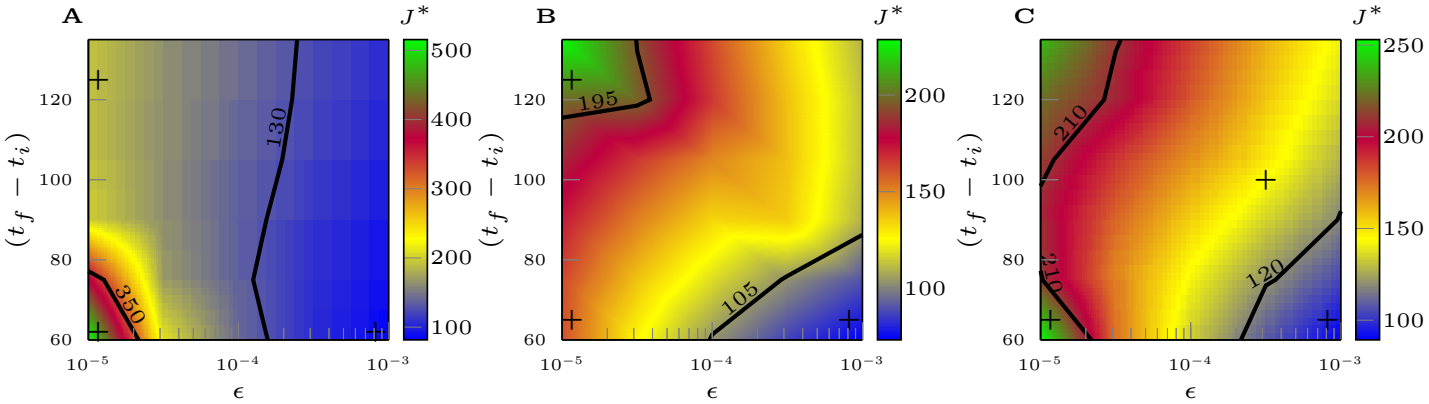

Figure 5: **The optimal cost J^* in the $(t_f - t_i)$, ϵ plane**. (A) shows the case of the Los Angeles Metropolitan Statistical Area. (B) shows the case of the NYC Metropolitan Statistical Area. (C) shows the case of the Seattle Metropolitan Statistical Area. For each city, $I_{\max}$ is chosen as the maximum of the range in Table 3. Type 1 solutions (in green) are more expensive than type 2 solutions (in blue.) The regions in yellow/red correspond to the transition between the two types of solutions. The parameter c_p and c_q are both set to 1.

1 solutions (in green) and low for type 2 solutions (in blue), with a transition area in between shown in yellow and red. This is illustrated in Fig. 5 and in more detail in Fig. 19 (LA), Fig. 20 (Seattle) and Fig. 21 (NYC) of the Supplementary Note 5. Our results show that considerations about the timescale of the control action apply differently to different cities. As can be seen, the particular emergence of type-1 or type-2 solutions is affected by both the choice of ϵ and $(t_f - t_i)$. From Fig. 5 (A) for LA we see that the most expensive control solutions are achieved when one is trying to suppress the epidemics to a low level in a short time (small ϵ and short $t_f - t_i$.) This is opposite to what seen for NYC in Fig. 5 (B), where longer time horizons are usually associated with more expensive solutions. The most convenient solution arises when the suppression constant ϵ is large but the control horizon is short (area shown in blue.) Finally, Fig. 5 (C) for Seattle shows that in the case of small ϵ, expensive solutions are obtained when the control horizon is either short or long, while an intermediate control horizon is to be preferred. We also see that setting the suppression constraint to a larger value and the control horizon to be short can reduce the cost considerably (which is similar to NYC). In general, these results point out the importance of carefully choosing the timescale over which one is seeking to suppress the epidemics, as well as the suppression level.

Implementation of non-optimal control solutions

We have provided an approach to robustly minimize the effects on the economy of social distancing measures in the presence of relevant constraints. However, it is possible that a number of considerations may limit the

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

implementation of such optimized interventions. We are interested in the effects of implementation of non-optimal controls solutions. We thus consider application of a variation of the optimal solution

$$\tilde{P}(t) = \min\{1, (1+\alpha)P^*(t)\}, \tag{10}$$

$\alpha > 0$, and analyze violations of the constraints $I_{\max}$ and ϵ as α is varied. Increasing values of α indicate stronger deviations of the control action from the optimal one. Table 2 summarizes application of such non-optimal interventions in all four cities, with α varying from 0.1 to 0.5. The letter 'S' stands for constraint satisfied, otherwise we report the percentage by which the constraint is violated. We see that the constraint on $I_{\max}$ remains always satisfied (this is expected as this constraint is not dominant) but strong violations of the suppression constraint ϵ are otherwise recorded. From Table 2 we see that for $\alpha = 0.5$ the constraint on ϵ in the four areas of interest is violated by an amount that varies from roughly 4000% for Houston to 80000% for NYC (which corresponds to a fraction of infected people at the final time in the order of 10^{-2}.) These results are opposite to those observed previously. The city that is less resilient to variation in the control input is NYC, which was previously reported to have received closer to optimal control interventions. This is due to the higher β for NYC (see Table 1) and to the fact that the optimal $P^*(t)$ for NYC is higher than for the other cities (corresponding to less strict social distancing required); as a result, increasing $P(t)$ further leads to the poorest outcome. This also highlights the risk for resurgences of the epidemics after partial suppression has been achieved [34], which is currently seen in different parts of the world.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 2: Violation in the constraints $I_{\max}$ and ϵ for the non-optimal solutions.

NYC: $I_{\max} = 0.0088, c_q = 1, 3, 6$

α	$I_{\max}$	ϵ	J
0	S	S	172.7213
0.1	S	272.4%	148.8441
0.2	S	1.37E+03%	128.9513
0.3	S	5.90E+03%	112.1306
0.4	S	2.39E+04%	97.7420
0.5	S	8.56E+04%	85.3439

LA: $I_{\max} = 0.0066, c_q = 1, 3, 6$

α	$I_{\max}$	ϵ	J
0	S	S	196.7950
0.1	S	13.6355%	170.7582
0.2	S	192.90%	149.0710
0.3	S	652.79%	130.7347
0.4	S	1.82E+03%	115.0389
0.5	S	4.54E+03%	101.4680

Houston: $I_{\max} = 0.0086, c_q = 1, 3, 6$

α	$I_{\max}$	ϵ	J
0	S	S	198.7045
0.1	S	9.17%	172.4965
0.2	S	177.67%	150.6672
0.3	S	602.96%	132.2108
0.4	S	1.67E+03%	116.4129
0.5	S	4.11E+03%	102.7529

Seattle: $I_{\max} = 0.0046, c_q = 1, 3, 6$

α	$I_{\max}$	ϵ	J
0	S	S	206.7624
0.1	S	65.96%	179.8001
0.2	S	432.45%	157.3389
0.3	S	1.65E+03%	138.3460
0.4	S	5.57E+03%	122.0895
0.5	S	1.73E+04%	108.0463

Discussion

In this paper, we have proposed an approach to optimize social distancing measures in time in order to contain the spread of the COVID-19 epidemics, while minimizing the impact on the economy. Our analysis has been applied to four different metropolitan statistical areas within the US, but can be directly applied to other geographical areas. Each MSA was shown to be well described by a stylized mathematical model whose parameters were inferred by daily new case counts reported by local health administrations. Other works have applied optimal control to the COVID-19 epidemics but none has considered data-driven models. Our approach based on modeling, inferring model parameters from real data, and computing optimal control solutions for the inferred model is general and may be applied to other complex systems of interest.

Our choice of the objective function is quite simplified, namely we assume that increasing levels of social distancing and quarantining result in progressively higher economical costs. In addition, we did not account for the cost associated with medical treatments, which is important yet significantly smaller than the cost of shutting down the economical activities of the entire society. Our approach can be easily generalized to other more complex, more realistic types of objective functions, see e.g., [39, 40]. Different from this study, these objective functions may also be specific to given regions or try to capture a particular socio-economical model of interest.

We have found that the optimal control solutions are quite robust to the specific choice of the objective

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

function and of its parameters, such as c_q. These control solutions tend to be qualitatively similar for different cities. However, they are affected by the choice of the constraints, in particular the constraint that we have associated to suppression of the epidemic, and by the time horizon of the control action. When these are varied, different cities behave differently, which points out the importance of our data-driven approach. We have also seen that small deviations from the optimal solution can lead to dramatic violations of the constraints.

It is possible to translate these optimal interventions in actual measures that can be imposed on the population, such as restricting the access to certain businesses or venues. While implementation of a time-varying control may be challenging in practice, we found that the optimal solution is typically characterized by an initial drop (due essentially to the non-optimality of current control interventions), followed by a nearly constant control (specific to each city) for a long time, and by a final drop, which is needed to achieve the desired suppression of the epidemics at the final time. Thus the optimal control solution is almost constant except for the initial time and the final time, which substantially increases the applicability of this study. The constant part of the solution could be practically achieved by dynamical regulation to control the time-varying reproduction number $R_t \approx 1$. A different type of optimal solution, which we call 'type 2', may arise either when the suppression constraint ϵ grows or the control horizon $(t_f - t_i)$ grows. A key observation is that the optimal solution P^* is generally lower than the value of P inferred by data. The initial drop in P^* may provide a measure of non-optimality of current interventions. This drop was seen to be smallest for NYC compared to other US cities, but was present in all the cities we have analyzed.

Our analysis reveals that the cost of eradicating the disease—i.e., suppressing the number of infected individuals down to a certain critical threshold, such as $\ll 1$ person in the formulation of our stylized mathematical model, even in the optimal case, is significantly higher than the cost of 'managing' the pandemic to avoid the saturation of regional medical resources. In light of the current progression of the pandemic in the US, our analysis brings rigorous scientific and quantitative foundation for the latter strategy, which is adopted by many local administrations (e.g., State offices.)

Our conclusion that the optimal control solution is well approximated by a constant level of social distancing is in partial contrast with the recommended system of reopening in phases [12]. The expected progression is from phase 1 to phase 2 and higher, but there have been several cases in which a premature reopening followed by a rise in COVID-19 numbers has led to folding back into phase 1 [37, 41, 15, 20]. Based on our results, such alternating control interventions are suboptimal because the economic benefits of momentarily relaxing the restrictions are lower than the costs associated with the successive tightening (such as e.g., a second lockdown.)

Occasionally, there have been claims by policy makers and/or scientists that seeking herd immunity by exposure to the virus (in the absence of a vaccine) may provide the best long-term path-forward [28, 18]. However, the timescales over which herd immunity can be achieved without violating the $I_{\max}$ constraint appear to be quite long (see Methods) and finally comparable with the timescale over which we expect availability of a vaccine. By setting a very large control horizon in our simulations, we have seen the emergence of these 'herd immunity solutions', shown in the Supplementary Note 6. However, the parameters of our realistic scenarios, estimated from real data, are far away from the point at which such solutions are optimal. In general, our study allows to quantify how 'far' one is in the parameter space from the point at which a herd immunity solution becomes

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

optimal.

The time-scale of control interventions, i.e., the control horizon, appears to play a fundamental role. Considerations about the control horizon may vary from area to area and may be affected by a number of factors, including expectations about the time at which a vaccine will become available. It appears that cities that have seen an increase of cases during the inference period need a longer control horizon to suppress the epidemics optimally. In certain instances, the impact on the economy can be minimized by tuning the control horizon; for example, for the city of Seattle we found that an optimal control horizon was equal to roughly 90 days when the suppression constraint $\epsilon = 10^{-5}$. We wish to emphasize that given the very large economical impact of social distancing measures, even a small improvement in the control strategy can lead to considerable economical benefits.

One relevant question is whether policy-makers can assess whether a currently employed control action is optimal or not. The so-called HAMVET procedure, initially proposed in [33] and presented in the Supplementary Note 8, can be used to validate a control strategy and evaluate its optimality.

References

[1] LANL COVID-19 Prediction GitHub.

[2] Ministry of Health, Official report on the cruise ship Diamond Princess, May 1, 2020.

[3] New York Times repository of Covid-19 data in the United States.

[4] U.S. Office of Management and Budget Delineation Files.

[5] Daron Acemoglu, Victor Chernozhukov, Iván Werning, and Michael D Whinston. A multi-risk sir model with optimally targeted lockdown. Technical report, National Bureau of Economic Research, 2020.

[6] Sean C Anderson, Andrew M Edwards, Madi Yerlanov, Nicola Mulberry, Jessica Stockdale, Sarafa A Iyaniwura, Rebeca C Falcao, Michael C Otterstatter, Michael A Irvine, Naveed Z Janjua, et al. Estimating the impact of covid-19 control measures using a bayesian model of physical distancing. *medRxiv*, 2020.

[7] Christophe Andrieu and Johannes Thoms. A tutorial on adaptive mcmc. *Statistics and computing*, 18(4):343–373, 2008.

[8] Sheryl L Chang, Nathan Harding, Cameron Zachreson, Oliver M Cliff, and Mikhail Prokopenko. Modelling transmission and control of the covid-19 pandemic in australia. *arXiv preprint arXiv:2003.10218*, 2020.

[9] IHME COVID, Christopher JL Murray, et al. Forecasting covid-19 impact on hospital bed-days, icu-days, ventilator-days and deaths by us state in the next 4 months. *MedRxiv*, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[10] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Unwin, Helen Coupland, T Mellan, Harisson Zhu, Tresnia Berah, J Eaton, P Perez Guzman, et al. Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on covid-19 in 11 european countries. 2020.

[11] France 24. 2020. "Germany reimposes local lockdowns after regional coronavirus outbreak". France 24. `https://www.france24.com/en/20200623-germany-reimposes-local-lockdown-after-coronavirus-outbreak`. August 2020.

[12] Scott Gottlieb, Caitlin Rivers, Mark B McClellan, Lauren Silvis, and Crystal Watson. National coronavirus response: a road map to reopening. *AEI Paper & Studies*, 2020.

[13] Anu G Gupta, Cheryl A Moyer, and David T Stern. The economic impact of quarantine: Sars in toronto as a case study. *Journal of Infection*, 50(5):386–393, 2005.

[14] Janet Christenbury. 2020. "Study in ICU Finds 30.9% Mortality Rate From COVID-19". Futurity. `https://www.futurity.org/covid-19-mortality-rate-2377362-2/`. August 2020.

[15] Ashifa Kassam. Spain warned of dire impact of second coronavirus lockdown. *The Guardian*, 2020.

[16] Johannes Köhler, Lukas Schwenkel, Anna Koch, Julian Berberich, Patricia Pauli, and Frank Allgöwer. Robust and optimal predictive control of the covid-19 outbreak. *arXiv preprint arXiv:2005.03580*, 2020.

[17] Miklós Koren and Rita Pető. Business disruptions from social distancing. *arXiv preprint arXiv:2003.13983*, 2020.

[18] Kin On Kwok, Florence Lai, Wan In Wei, Samuel Yeung Shan Wong, and Julian WT Tang. Herd immunity–estimating the level required to halt the covid-19 epidemics in affected countries. *Journal of Infection*, 80(6):e32–e33, 2020.

[19] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of internal medicine*, 2019, 2020.

[20] Jasmine C. Lee, Sarah Mervosh, Yuriria Avila, Barbara Harvey, and Alex Leeds Matthews. Spain warned of dire impact of second coronavirus lockdown. *The New York Times*, 2020.

[21] Yen Ting Lin, Jacob Neumann, Ely F Miller, Richard G Posner, Abhishek Mallela, Cosmin Stafa, Jaideep Ray, Gautam Thakur, Supriya Chinthavali, and William S Hlavacek. Daily Forecasting of New Cases for Regional Epidemics of Coronavirus Disease 2019 with Bayesian Uncertainty Quantification. *medRxiv*, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[22] Lisa Mataloni, Dave Wasshausen, Erich Strassner, Jeannine Aversa. 2020. "Gross Domestic Product, Second Quarter 2020 (Advance Estimate) and Annual Update". U.S. Bureau of Economic Analysis. `https://www.bea.gov/sites/default/files/2020-07/gdp2q20_adv_0.pdf`. August 2020.

[23] Xiaofan Liu, Hong Zhou, Yilu Zhou, Xiaojun Wu, Yang Zhao, Yang Lu, Weijun Tan, Mingli Yuan, Xuhong Ding, Jinjing Zou, et al. Risk factors associated with disease severity and length of hospital stay in covid-19 patients. *Journal of Infection*, 81(1):e95–e97, 2020.

[24] Lother Wieler, Ute Rexroth, Rene Gottschalk. 2020. "Emerging COVID-19 success story: Germany's strong enabling environment". Our World in data. `https://ourworldindata.org/covid-exemplar-germany`. August 2020.

[25] Dylan H Morris, Fernando W Rossine, Joshua B Plotkin, and Simon A Levin. Optimal, near-optimal, and robust epidemic control. *arXiv preprint arXiv:2004.02209*, 2020.

[26] Maria Nicola, Zaid Alsafi, Catrin Sohrabi, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, Maliha Agha, and Riaz Agha. The socio-economic implications of the coronavirus pandemic (covid-19): A review. *International journal of surgery (London, England)*, 78:185, 2020.

[27] Hiroshi Nishiura, Natalie M Linton, and Andrei R Akhmetzhanov. Serial interval of novel coronavirus (covid-19) infections. *International journal of infectious diseases*, 2020.

[28] Eric JW Orlowski and David JA Goldsmith. Four months into the covid-19 pandemic, sweden's prized herd immunity is nowhere in sight. *Journal of the Royal Society of Medicine*, 113(8):292–298, 2020.

[29] Javier Perez-Saez, Stephen A Lauer, Laurent Kaiser, Simon Regard, Elisabeth Delaporte, Idris Guessous, Silvia Stringhini, Andrew S Azman, Serocov-POP Study Group, et al. Serology-informed estimates of sars-cov-2 infection fatality risk in geneva, switzerland. *medRxiv*, 2020.

[30] Pipetius Quah, Andrew Li, and Jason Phua. Mortality rates of patients with covid-19 in the intensive care unit: a systematic review of the emerging literature. *Critical Care*, 24:1–4, 2020.

[31] Haley E Randolph and Luis B Barreiro. Herd immunity: Understanding covid-19. *Immunity*, 52(5):737–741, 2020.

[32] Safiya Richardson, Jamie S. Hirsch, Mangala Narasimhan, James M. Crawford, Thomas McGinn, Karina W. Davidson, , and the Northwell COVID-19 Research Consortium. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA*, 323(20):2052–2059, 05 2020.

[33] I Michael Ross. *A primer on Pontryagin's principle in optimal control*. Collegiate publishers, 2015.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[34] Nick Warren Ruktanonchai, JR Floyd, Shengjie Lai, Corrine Warren Ruktanonchai, Adam Sadilek, Pedro Rente-Lourenco, Xue Ben, Alessandra Carioli, Joshua Gwinn, JE Steele, et al. Assessing the impact of coordinated covid-19 exit strategies across europe. *Science*, 2020.

[35] Aki Sakurai, Toshiharu Sasaki, Shigeo Kato, Masamichi Hayashi, Sei-ichiro Tsuzuki, Takuma Ishihara, Mitsunaga Iwata, Zenichi Morise, and Yohei Doi. Natural History of Asymptomatic SARS-CoV-2 Infection. *New England Journal of Medicine*, 0(0):null.

[36] Steven Sanche, Yen Ting Lin, Chonggang Xu, Ethan Romero-Severson, Nicolas W Hengartner, and Ruian Ke. The novel coronavirus, 2019-ncov, is highly contagious and more infectious than initially estimated. *arXiv preprint arXiv:2002.03268*, 2020.

[37] Allan Saul, Nick Scott, Brendan S Crabb, Suman S Majundar, Benjamin Coghlan, and Margaret E Hellard. Victoria's response to a resurgence of covid-19 has averted 9,000-37,000 cases in july 2020. *The Medical Journal of Australia*, page 1, 2020.

[38] Shawn Radcliffe. 2020. "Here's Exactly Where we Are with Vaccines and Treatments for COVID-19". Healthline. `https://www.healthline.com/health-news/heres-exactly-where-were-at-with-vaccines-and-treatments-for-covid-19`. August 2020.

[39] Afroza Shirin, Fabio Della Rossa, Isaac Klickstein, John Russell, and Francesco Sorrentino. Optimal regulation of blood glucose level in type i diabetes using insulin and glucagon. *PloS one*, 14(3):e0213665, 2019.

[40] Afroza Shirin, Feng Song, Yen-Ting Lin, William S Hlavacek, and Sorrentino S. Prediction of optimal drug schedules for controlling autophagy. *Scientific Reports*, 9(1428), 2019.

[41] Lorenzo Tondo. Italy at crossroads as fears grow of covid-19 second wave. *The Guardian*, 2020.

[42] Nguyen Van Vinh Chau, Vo Thanh Lam, Nguyen Thanh Dung, Lam Minh Yen, Ngo Ngoc Quang Minh, Le Manh Hung, Nghiem My Ngoc, Nguyen Tri Dung, Dinh Nguyen Huy Man, Lam Anh Nguyet, Le Thanh Hoang Nhat, Le Nguyen Truc Nhu, Nguyen Thi Han Ny, Nguyen Thi Thu Hong, Evelyne Kestelyn, Nguyen Thi Phuong Dung, Tran Chanh Xuan, Tran Tinh Hien, Nguyen Thanh Phong, Tran Nguyen Hoang Tu, Ronald B Geskus, Tran Tan Thanh, Nguyen Thanh Truong, Nguyen Tan Binh, Tang Chi Thuong, Guy Thwaites, Le Van Tan, and Oxford University Clinical Research Unit COVID-19 Research Group. The Natural History and Transmission Potential of Asymptomatic Severe Acute Respiratory Syndrome Coronavirus 2 Infection. *Clinical Infectious Diseases*, 2020.

[43] Roman Wölfel, Victor M Corman, Wolfgang Guggemos, Michael Seilmaier, Sabine Zange, Marcel A Müller, Daniela Niemeyer, Terry C Jones, Patrick Vollmar, Camilla Rothe, et al. Virological assessment of hospitalized patients with covid-2019. *Nature*, 581(7809):465–469, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Methods

Estimation of $I_{\max}$

The $I_{\max}$ values can be approximated with simple assumptions as seen in Table 3 for a selection of major U.S. cities. These $I_{\max}$ values are a ratio of ICU beds to the population of people which require them. Following [21], we estimate that the probability of death conditioned on symptomatic infection is equal to $f_H \times (1 - f_R) = 0.01134$, where the two parameters f_H and f_R were independently computed in [29] and [32], respectively. Data [30, 14] shows that the mortality rate for patients sent to ICU is between 30% and 40%, thus, it is reasonable to assume that an overall fraction of infected people equal to $0.01134/0.35 = 0.0324$ needs ICU beds. In the hypothetical situation that the population of an entire state contracts COVID-19 3.24% will require an ICU bed. The ρ term is a modifier which denotes how many ICU beds are available to COVID-19 patients as some beds could be used for other reasons. Reasonably, the value of ρ ranges from 2/3 to 1 in Table 3. $I_{\max}$ is then calculated as the number of available ICU beds (including the ρ assumption) divided by number people which contract COVID-19 and also require an ICU bed (3.24% assumption).

Table 3: $I_{\max}$ values for U.S. Cities

City	$I_{\max}$ ($\rho = 2/3$)	$I_{\max}$ ($\rho = 1$)
New York, NY	0.88%(5704/648000)	1.32% (8556/648000)
Los Angeles, CA	0.64% (8241/1280124)	0.97%(12361/1280124)
Chicago, IL	1.00%(4094/410508)	1.50%(6141/410508)
Dallas, TX	0.86%(8109/939600)	1.29%(12163/939600)
Houston, TX	0.86%(8109/939600)	1.29%(12163/939600)
Riverside, CA	0.64% (8241/1280124)	0.97%(12361/1280124)
Miami, FL	0.98% (6829/695952)	1.47%(10244/695952)
Philadelphia, PA	1.047%(4342/414720)	1.57%(6513/414720)
Atlanta, GA	0.65%(2249/344088)	0.98%(3374/344088)
Phoenix, AZ	0.55%(1309/2358450)	0.83%(1963/2358450)
Boston, MA	0.4%(893/223330)	0.60%(1340/223330)
San Francisco, CA	0.64% (8241/1280124)	0.97%(12361/1280124)
Seattle, WA	0.46%(1143/246888)	0.69%(1714/24688)
Indianapolis, IN	0.83%(1811/218120)	1.25%(2716/218120)
Detroit, MI	0.71%(2315/32358)	1.07%(3472/32358)
Baltimore, MD	0.40%(893/223330)	0.60%(1340/223330)
Denver, CO	0.56%(1039/186590)	0.84%(1559/186590)
Portland, OR	0.56%(768/136660)	0.84%(1152/136660
Las Vegas, NV	0.75%(752/99792)	1.13%(1129/99792)

medRxiv preprint doi: https://doi.org/10.1101/2020.08.27.20183574; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

For all our numerical experiments, we consider two values for $I_{\max}$, one corresponding to $\rho = 2/3$ and another one corresponding to $\rho = 1$.

Timescale for T_{herd}

We now attempt to answer the following question. By enforcing satisfaction of the constraint with the equal sign $I_s(t) + I_{tp}(t) = I_{\max}$, how long would it take before herd immunity is achieved? By assuming long-term immunity of those recovered from the virus, we can expect herd immunity to arise when roughly 80% of the population has been exposed [31]. Consider for example the case of NYC for which from Table 3 we see that $I_{\max}$ is between 0.88 % and 1.32 %. We assume an average hospitalization of 20 days [23]. That means that the time to achieve herd immunity T_{herd} varies between 606 days=$(80 \times 20/(1.32 \times 2))$ and 909 days=$(80 \times 12/(0.88 \times 2))$. The factor of 2 accounts for the fact that roughly one exposed person out of two develops symptoms. Analogously, for Chicago, we estimate T_{herd} to vary between 533 days=$(80 \times 20/(1.5 \times 2))$ and 800 days=$(80 \times 20/(1 \times 2))$. From these back-of-the-envelope calculations we see that the timescale over which herd immunity can be achieved without violating the $I_{\max}$ constraint appears to be quite long and comparable to if not longer than the timescale over which we expect availability of a vaccine.

Acknowledgements

The authors would like to thank Bill Hlavacek, Isaac Klickstein, Neethi Thevan T., and Enrico Del Frate for insightful discussions. YTL thanks the support from the Laboratory Directed Research and Development Program at Los Alamos National Laboratory (Project XX01).