

Supplementary Materials for

**On the Spread of Coronavirus Infection. A Mechanistic Model to Rate Strategies for
Disease Management**

Shiyang Wang and Doraiswami Ramkrishna

Correspondence to: ramkrish@purdue.edu

This PDF file includes:

Materials and Methods

Supplementary Text

Figs. S1 to S3

Tables S1 to S3

References and Notes

Materials and Methods

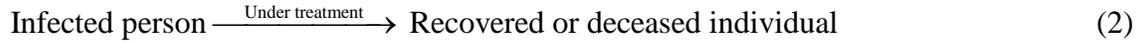
Data. Daily coronavirus infection data (county by county) in United States came from New York Times GitHub source (17). We extracted daily positive infection numbers from 21st January 2020 to 25th April 2020. To estimate the effect of weather (county by county) on the transmissibility of the coronavirus, we used the weather data for 2020 from NOAA Global Surface Summary of the Day (GSOD) (18). To study the state orders and policies for addressing the coronavirus (state by state in US), we extracted data from Kaiser Family Foundation (19).

Mechanics Model. We consider a geometric domain D in which there is a total population density (number per unit area) N_o , containing n individuals (per unit area) diseased with the coronavirus. These population densities must be clearly viewed as averaged over the whole domain, which may be a county, city, state or even an entire country. One may expect steep variations of the actual number densities about their average values. Both the averages and fluctuations about them may be expected to vary with time and location, and a model that includes such dynamic effects would require a complex mathematical framework riddled with parameters not easily identified even if detailed observations can be made. Our goal here is to find a suitably simple framework to produce a mathematical model that contains a limited number of parameters which can be readily identified from gross observations. Furthermore, they should relate in some way to various strategies that may be envisaged to control the spread of infection.

The basic mechanism of spreading infection is assumed to be exposure to a viral population in the environment (exhaled droplets, aerosol, contaminated surfaces, and possibly fecal-oral contamination) contributed by all those infected with the coronavirus. Obviously, the viral population in the environment will notably vary with location but over a period of time during which individuals in motion would encounter viruses when they are in the vicinity of the infected and are thus open to catching the infection which is represented below.

$$\text{Uninfected individual} \xrightarrow{\text{Exposure to Viral Environment}} \text{Infected individual} \quad (1)$$

The rate at which this process of infection occurs would clearly depend on the local viral population. Indeed, the viral virulence would vary significantly with location but the assumed simplicity of the model is derived from sufficient mixing of both the infected and the uninfected individuals that they see the average viral population over a period of time. The fate of infected individuals is represented by



Mathematically, we represent the rate in Eq. (1) by kV where V is the viral population averaged over the domain D . Alternately, $(1/kV)$ may be viewed as the average time for an individual to be infected. Since the transfer of infection implied in Eq. (1) occurs in multiple ways, the average time for infection must represent the mean of the average times for the different mechanisms. Thus the rate at which the infected population increases is then given by $k(N_o - n)V$, where $(N_o - n)$ represents the uninfected population density. The rate constant associated with the process of removed population is denoted as k_r and $k_r n$ represents the rate at which infected people are removed by death or recovering from treatment. We should note that while there are numerous reports on reinfection of COVID-19, majority of recovered patients retain certain immunity against the virus.

We can now write a balance equation for the infected population given by

$$\frac{dn}{dt} = kV(N_o - n) - k_r n \quad (3)$$

Eq. (3) is coupled to a balance equation for V , the viral population which accounts for their multiplication and their random death. This is given by

$$\frac{dV}{dt} = k_v n - k'_v V \quad (4)$$

The first term on the right-hand side accounts for viral multiplication from all diseased individuals while the second term is their first order disappearance. The model at this stage is represented by 4 parameters, k, k_r, k_v, k'_v . Table S1 collates their broad interpretations representing the spreading characteristics. It should be apparent that constants k_r and k_v are not as sensitive to disease management as the constants k and k'_v are. For instance, k and k'_v could be conceivably reduced by the use of masks and social distancing.

Insofar as the infection curve results from different processes, it is more meaningful to convert the model variables to dimensionless variables as below.

$$x \equiv \frac{n}{N_o} = \text{fraction infected}, \quad y \equiv \frac{V}{V_o} = \text{Viral population density scaled to } V_o, \quad \tau \equiv kV_o t$$

As a scaled variable, y is a more convenient representative of the viral population as their absolute numbers are very large. The scaled time relates absolute time to the average time it takes to infect a single person.

$$\frac{dx}{d\tau} = (1-x)y - \gamma x \tag{5}$$

$$\frac{dy}{d\tau} = \alpha x - \beta y \tag{6}$$

Three dimensionless parameters emerge from the above differential equations that are presented in Table 1 of the manuscript, along with their physical interpretations. It is of interest to note that the ratio α / β is identically free of population density (since it is proportional to the ratio of densities N_o / V_o). It will turn out, however, that fitting the infection data to the model shows a strong statistical correlation with the population.

Discussion of R0. In the Fig.1(A) of the manuscript, the P1 duration reveals the period of pathogen transmission with limited prevention in the United States. The early state of virus transmissibility

can be characterized by 'R-naught' (R0), which is the basic reproduction number for the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection, $n(t) = (0) \cdot R0^{t/s}$; s is the serial interval (for COVID-19, its median $\bar{s} \approx 3.96$ days (10)). Our model estimates $R0 \sim e^{\bar{s}/T_{inf}}$ where $\lambda = [-(\beta + \gamma) + \sqrt{(\beta - \gamma)^2 + 4\alpha}] / 2$ is the largest eigenvalue using the approach of next generation matrix (6). Therefore, our estimate of R0 is about 2.8 (the median from data is 2.75; our model is 2.90) whose transmission is stronger than influenza (R0:1.4-1.6) (11) and weaker than Measles (R0:12-18) (12).

Peak Infection Rate. The peak infection rate occurs when the second derivative $d^2x/d\tau^2$ vanishes along its path. It is readily shown that this occurs when

$$y = \frac{1}{2} \left(\frac{\gamma x}{1-x} - \beta - \gamma \right) + \frac{1}{2} \sqrt{\left(\frac{\gamma x}{1-x} - \beta - \gamma \right)^2 + 4x \left(\frac{\gamma^2}{1-x} + \alpha \right)} \quad (7)$$

The peak infection rate is obtained when the curve represented in (7) intersects the curve $\{x(\tau), y(\tau)\}$. This represents the stage beyond which the infection rate will drop. It is now possible to address the much debated strategy of "flattening the curve" at as low an infected fraction as possible. Since the infection rate is given by the right-hand side of (5) we infer that the peak infection rate occurs as

$$\Gamma = \frac{1}{2}(1-x) \left\{ \left(\frac{\gamma x}{1-x} - \beta - \gamma \right) + \sqrt{\left(\frac{\gamma x}{1-x} - \beta - \gamma \right)^2 + 4x \left(\frac{\gamma^2}{1-x} + \alpha \right)} \right\} - \gamma x \quad (8)$$

It will then be of interest to specify either the maximum infection rate or the fraction infected at which there is noticeable slowing of the infection rate (Figs. 2(C, D) of the manuscript).

Modeling Lockdown of New York City. Here, we consider the influence of lockdown policy at New York City, where the isolation is determined by the geographic constraint of five boroughs (the Bronx, Brooklyn, Manhattan, Queens, and Staten Island) within the New York City. In the

strict isolation, the entire New York City D is divided into five boroughs ($D_i, i=1, 2, \dots, 5$) and there being no interaction between the populations in different boroughs (Fig. 3(A) of the manuscript). We may model the transmission dynamics in domain D_i by using Eqs. (5) and (6) with the parameters of

$$\alpha_i = \alpha \left(\frac{N_{io}}{N_o} \right), \quad \beta_i = \beta, \quad \gamma_i = \gamma \left(\frac{N_{io}}{N_o} \right) \quad (9)$$

where (α, β, γ) are the parameters for New York City; N_{io} is the total residing population in borough i and N_o is the total population in New York City. In Fig. 3(B) of the manuscript, we evaluate the lockdown strategy by comparing $x(\tau)$ obtained for domain D with $\sum_i^5 x_i$.

Correlation of Weather vs. \tilde{x} and α/β . Fig. S1(A) shown below indicates that the correlation between \tilde{x} and mean temperature (mean) is insignificant ($p=0.56>0.05$). Figs. S1(B) shows the relationship between \tilde{x} and mean dew point (humidity). For the virulence (α/β), both temperature and dew point is lack of significance in relevance (Figs. S1(C, D)).

During the machine learning, we aim to include enough significant independent variables. We do not incorporate both temperature and dew point in our regression model because of the insignificant correlation. On the other hand, based on our analysis on the population and population density in Figs. 2(A, B) of the manuscript, which shows strong correlation to α/β and \tilde{x} , respectively. We have incorporate both population and population density in the machine learning model.

Machine Learning Model: Evaluation of Policy. In this section, we elaborate our procedure of applying regression model to 2 important indicators \tilde{x} (projected total infection fraction) and α/β (virulence) using machine learning. This section includes: (I) data definition, (II) coding system for categorical values, (III) Feature reduction, (IV) regression and cross validation.

(I) Data Definition: Since March 16th when the US administration declares the '15 days to slow the spread' policy, local government in each state have declared more than 300 executive orders, where the general policy measures can be categorized into 7 aspects:

- 1) Stay at home order:
 - a. Statewide
 - b. High-risk groups
 - c. Rolled back to high-risk groups
 - d. Lifted
 - e. N/A
- 2) Easing social distancing measure:
 - a. No
 - b. Yes
- 3) Mandatory quarantine for travelers:
 - a. All travelers
 - b. All air travelers
 - c. From certain states
 - d. Rolled back to certain states
 - e. Other
 - f. Lifted
 - g. N/A
- 4) Non-essential Business closure:
 - a. All non-essential business closed
 - b. Some non-essential business closed
 - c. Some non-essential businesses permitted to reopen with reduced capacity
 - d. Some non-essential businesses permitted to reopen
 - e. All non-essential businesses permitted to reopen with reduced capacity
 - f. N/A
- 5) Gathering ban:
 - a. All gathers prohibited
 - b. Expanded to 25+ people prohibited
 - c. Expanded to 20+ people prohibited

- d. 10+ people prohibited
 - e. Expanded to 10+ people prohibited
 - f. Other
 - g. Lifted
 - h. N/A
- 6) School closure:
- a. Closed for school year
 - b. Recommended closure for school year
 - c. Closed
 - d. N/A
- 7) Restaurant limits:
- a. Closed except for takeout/delivery
 - b. Limited dine-in service
 - c. Reopened to dine-in service with capacity limits
 - d. Reopened to dine-in service
 - e. N/A

Note that 'N/A' represents that no action is taken. Our data include 50 states and Washington D.C.

(II) Coding System for Categorical Variables in Regression Analysis: For each policy as a categorical variable, we using *simple coding approach* to convert the categorical variable into a series of variables for regression analysis. In simple coding method, we define a reference level for each categorical variable. For instance, in variable 'Stay at home order', we treat the item 'N/A' as the reference level. Therefore, using simple coding system, the parameters can be defined in Table S2. In the Table S2, the level of the categorical variable that is coded as '-1/5' in all of new variables is the reference level. Instead of categorical variable, we enter new variables into regression. We should note that during the regression, the weight of variable 'x1' would be the weight for 'Statewide' minus the weight for 'N/A'. Similarly, other categorical variables can be transformed to new variables. For every categorical variable, we always choose the last option as the reference level.

(III) Feature Selection: P-value & VIF: In linear regression, the predictor is the projected total infection percentage (\tilde{x}). We perform the linear regression (ordinary least squares regression) against the individual categorical variable and calculate P-value and VIF (Variance inflation factor). P-value represent the null-hypothesis and we require $p < 0.05$ to be significant. On the other hand, VIF quantifies the severity of the collinearity during regression analysis ($VIF < 5$). For each categorical variable, we list their performance in Table S3. Based on above analysis, for regression of \tilde{x} , variables (i), (ii), (iii), (vii) are insignificant correlation. Variables (iv), (v) and (vi) are significant. Similarly, we can perform regression of individual variables against α/β .

(IV) Linear Regression & Cross Validation: Next, we include all significant variables in the linear regression model, where the variables are associated with categorical variables (iv), (v) and (vi). Total number of variables associated with categorical variables are 15. By performing the regression with all fifteen variables, we found that for categorical variable (iv), the only significant variable is 'x1'. By excluding all other variables with categorical variable (iv), the overall performance of the model remains unchanged with total 11 variables. Additionally, we include another significant continuous variable 'population density'. With a total of 12 variables, we performed cross validation (10-fold) to examine the significance of our model. In general, all 12 variables are significant in determining \tilde{x} (see Fig. S2(B) for p-value). By examine the weight of each variable, we know that all features are important to determine \tilde{x} . The model has a R-square value around 0.7, indicating 70% accuracy in prediction. On the other hand, for α/β , we only include the population and variables (iv) (see Fig. S3).

Supplementary Text

Mechanics Model Implementation. We solve the Eqs. (5-6) using ode23tb subroutine from Matlab because of the stiffness of the nonlinear equations. To fit parameters $\alpha / \beta, kV_o = 1/T_{inf}$, and γ . We consider data from leading county of fifty states and Washington D.C. Next, we defined an objective function and minimized it using fminsearch subroutine.

To consider the uncertainty of the parameter fitting, we perform Monte Carlo simulations (sample size is 50) to improve the confidence of our fitting. We should note that: in Figs. 1(B, C), the plot of the model used the median of all realization; In Fig. 1(D), all parameters used in the model are median of the distributions; In Figs. 2(A, B), all parameters used in the model are median of the distributions; In Figs. 3(B), the plot of the model used the median of all realization; In Fig. 3(C), the solid line is the median of the distribution and the upper and lower limits of the shaded area are the maximum and minimum of the distribution.

In our mechanistic model, we consider the both infection dynamics of infected population and virulence environment, where the similar mechanism has been applied to study Malaria (i.g., Malaria Model (20,21)). On the other hand, the transmissibility dynamics estimated by the SIR model considers the factor of infected population. Secondly, our model aims to minimize the fitting parameters against the infection curve; other compartmental models in epidemiology (e.g., SEIR), they could result in many parameters fitting, which may lead to the overfitting of the data (22).

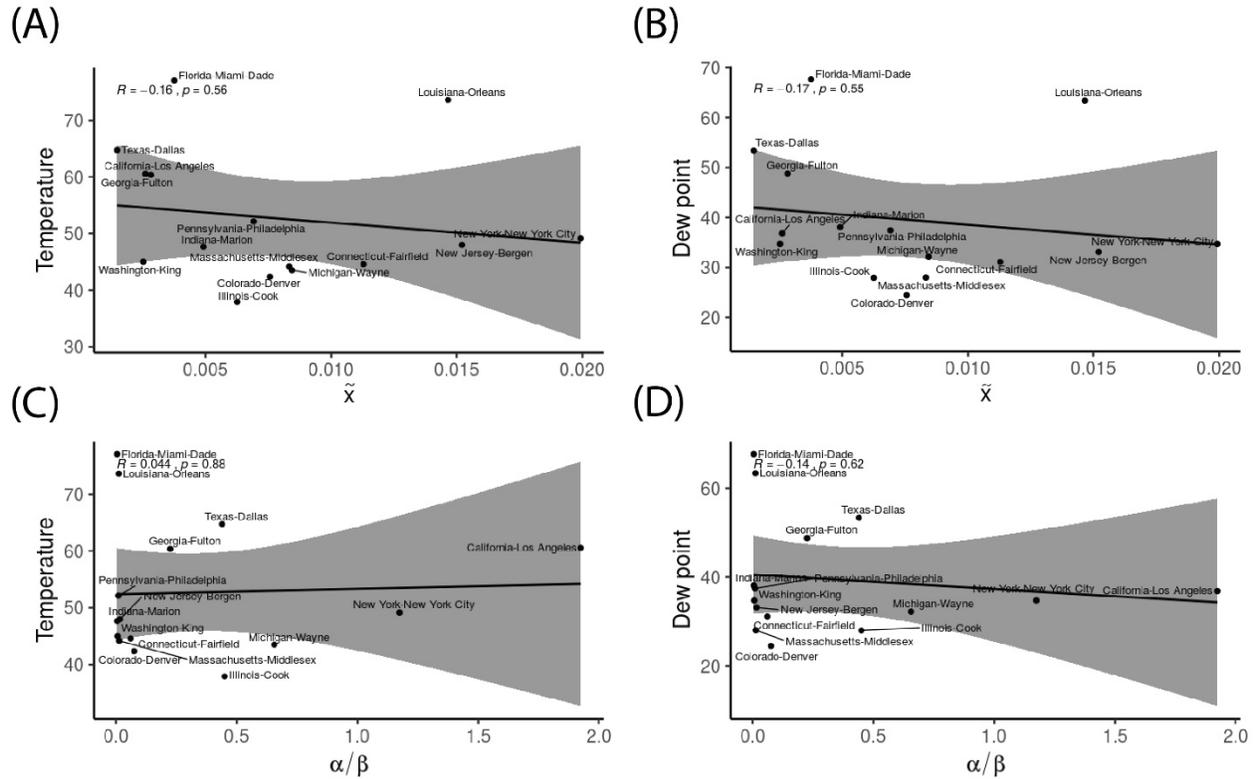


Fig. S1. Impact of weather on the model parameters. (A and B) The correlation between (A) mean temperature vs. the projected infection fraction (\tilde{x}), and (B) dew point vs. the projected infection fraction (\tilde{x}). (C and D) The correlation between (C) mean temperature vs. the virulence parameter (α/β), and (D) dew point vs. the virulence parameter (α/β). The sample size is 15.

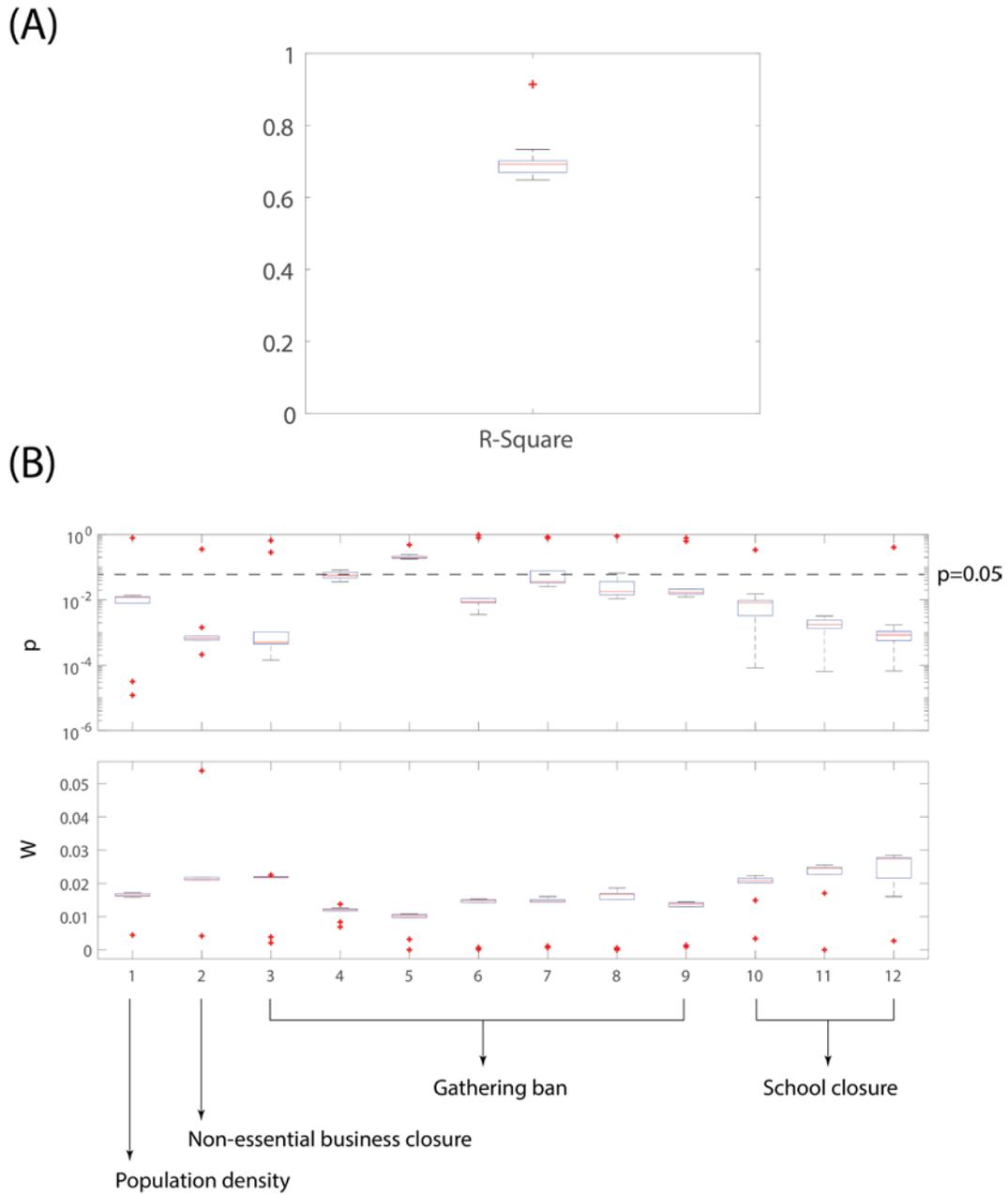


Fig. S2. Performance of regression of \tilde{x} using cross validation. (A) The boxplot of R-Square using cross validation. (B) The boxplot for p-value and weight of 12 variables using cross validation.

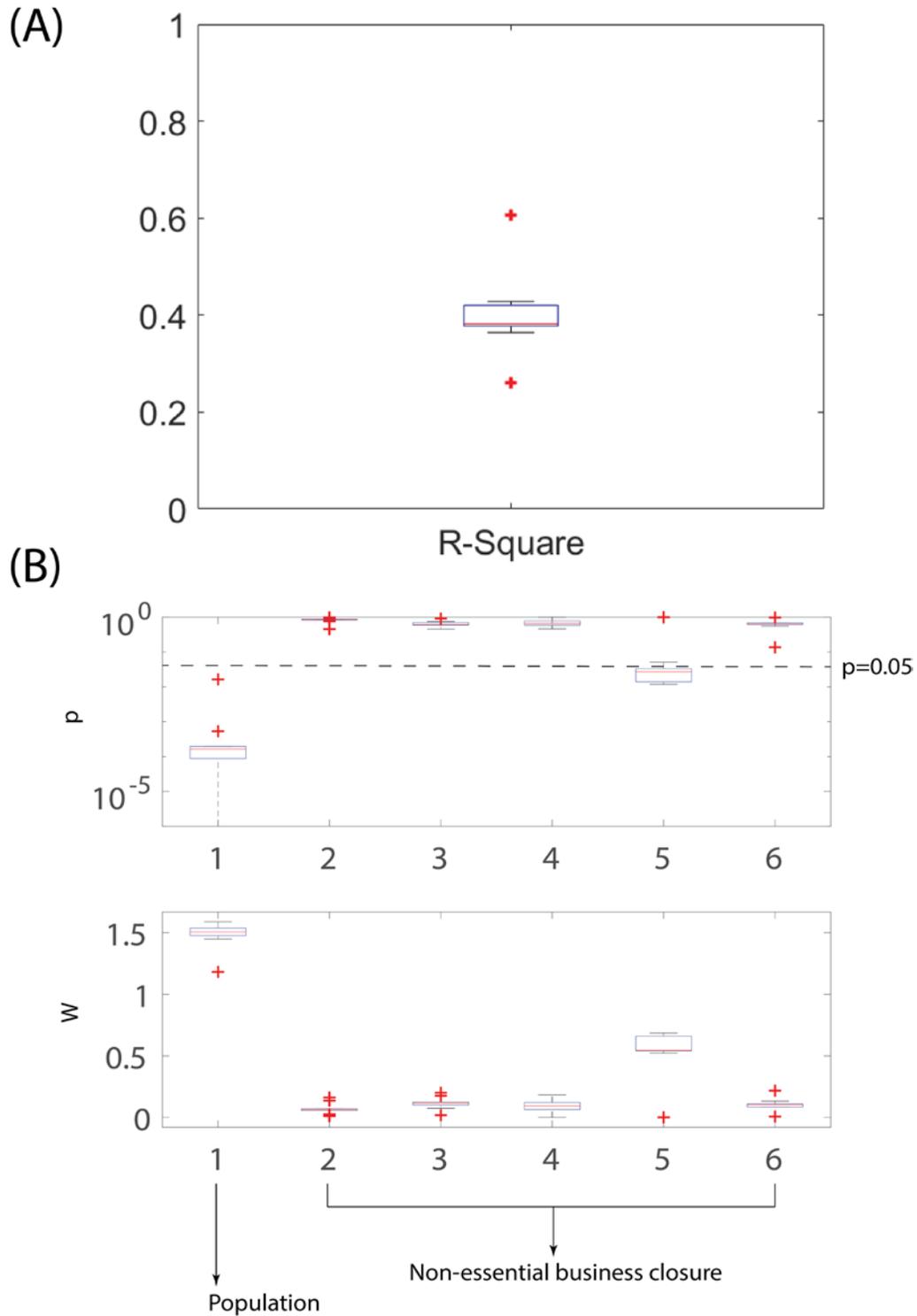


Fig. S3. Performance of regression of α/β using cross validation. (A) The boxplot of R-Square using cross validation. **(B)** The boxplot for p-value and weight of 12 variables using cross validation.

Table S1. Variable definitions in model

Notation	Variables
n	Infected population density at any time t
N_o	Population density in the domain
V_o	Characteristic viral population density
V	Viral population density at any time t
k	Rate constant for transfer of infection k^{-1} is the average time for infection
k_v	Rate constant for production of virus from the infected k_v^{-1} is the average time for viral multiplication
k'_v	Rate of death of virus
k_r	Rate of removed population (either by death or recovery)

Table S2. Coding System for Categorical Variable ‘Stay at home order’

Feature ‘Stay at home order’	New Variable (x1)	New Variable (x2)	New Variable (x3)	New Variable (x4)
‘Statewide’	4/5	-1/5	-1/5	-1/5
‘High-risk groups’	-1/5	4/5	-1/5	-1/5
‘Rolled back to high-risk groups’	-1/5	-1/5	4/5	-1/5
‘Lifted’	-1/5	-1/5	-1/5	4/5
‘N/A’	-1/5	-1/5	-1/5	-1/5

Table S3-1. P-values and VIF coefficients for \tilde{x} against all categorical values

(i) 'Stay at home order'

Variables	P-value	VIF
x1	0.0438	2.41
x2	0.3069	1.28
x3	0.2336	1.14
x4	0.0818	2.55
P-value for regression model: 0.335		

(ii) 'Easing social distancing measure'

Variables	P-value	VIF
x1	0.879	1
P-value for regression model: 0.879		

(iii) 'Mandatory quarantine for travelers'

Variables	P-value	VIF
x1	0.6360	1.0154
x2	0.7244	1.0154
x3	0.7277	1.0154
x4	0.3957	1.0714
x5	0.6351	1.0154
x6	0.2189	1.0910
P-value for regression model: 0.695		

Table S3-2. P-values and VIF coefficients for \tilde{x} against all categorical values

(iv) 'Non-essential Business closure'

Variables	P-value	VIF
x1	8.7586e-05	1.2810
x2	0.62019	1.9673
x3	0.85621	2.2941
x4	0.2553	1.7647
x5	0.66362	2.5556
P-value for regression model: 0.00131		

(i) 'Gathering ban'

Variables	P-value	VIF
x1	0.0019359	2.7647
x2	0.0043698	2.3529
x3	0.017636	1.4706
x4	1.3066e-05	6.8824
x5	0.00038911	1.9216
x6	0.00073714	1.9216
x7	8.2682e-05	5.5882
P-value for regression model: 0.00253		

(ii) 'School closure'

Variables	P-value	VIF
x1	0.24445	8.8431
x2	0.05413	6.1765
x3	0.041329	11.6667
P-value for regression model: 0.0261		

Table S3-3. P-values and VIF coefficients for \tilde{x} against all categorical values

(iii) 'Restaurant limits'

Variables	P-value	VIF
x1	0.55178	1.9608
x2	0.67832	10.8824
x3	0.4261	2.8824
x4	0.65758	12.0000
P-value for regression model: 0.92		

References and Notes

1. M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini Jr., A. Vespignani, The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science* **368**, 395 (2020).
2. L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, C. Fraser, Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, 6491 (2020).
3. L. Bourouiba, Turbulent gas clouds and respiratory pathogen emissions: potential implications for reducing transmission of covid-19. *Jama* (2020).
4. Z.-D. Guo, Z.-Y. Wang, S.-F. Zhang, X. Li, L. Li, C. Li, Y. Cui, R.-B. Fu, Y.-Z. Dong, X.-Y. Chi, M.-Y. Zhang, K. Liu, C. Cao, B. Liu, K. Zhang, Y.-W. Gao, B. Lu, W. Chen, Aerosol and surface distribution of severe acute respiratory syndrome coronavirus2 in hospital wards, Wuhan, China, 2020. *Emerg. Infect. Dis.* **26**, 7(2020).
5. S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, J. Lessler, The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med.* **172**(9), 577-582 (2020).
6. M. Lipsitch, T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, G. Gopalakrishna, S. K. Chew, C. C. Tan, M. H. Samore, D. Fisman, M. Murray, Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**, 1966 (2003).
7. A. Repici, R. Maselli, M. Colombo, R. Gabbiadini, M. Spadaccini, A. Anderloni, S. Carrara, A. Fugazza, M. Di Leo, P. A. Galtieri, G. Pellegatta, E. C. Ferrara, E. Azzolini, M. Lagioia, Coronavirus (covid-19) outbreak: what the department of endoscopy should know. *Gastrointest. Endosc.* (2020).
8. L. Lan, D. Xu, G. Ye, C. Xia, S. Wang, Y. Li, and H. Xu, Positive rt-pcr test results in patients recovered from covid-19. *Jama* **323**, 1502 (2020).
9. C. Fraser, C. A. Donnelly, S. Cauchemez, W. P. Hanage, M. D. Van Kerkhove, T. D. Hollingsworth, J. Griffin, R. F. Baggaley, H. E. Jenkins, E. J. Lyons, T. Jombart, W. R. Hinsley, N. C. Grassly, F. Balloux, A. C. Ghani, N. M. Ferguson, A. Rambaut, O. G. Pybus, H. Lopez-Gatell, C. M. Alpujch-Aranda, I. Bojorquez Chapela, E. Palacios Zavala, D. Ma. Espejo Guevara, F. Checchi, E. Garcia, S. Hugonnet, C. Roth, The WHO Rapid Pandemic Assessment Collaboration, Pandemic potential of a strain of influenza a (h1n1): early findings. *Science* **324**, 1557 (2009).
10. F. M. Guerra, S. Bolotin, G. Lim, J. Heffernan, S. L. Deeks, Y. Li, and N. S. Crowcroft, The basic reproduction number (r_0) of measles: a systematic review. *Lancet Infect. Dis.* **17**, e420 (2017).
11. J. Grein, N. Ohmagari, D. Shin, G. Diaz, E. Asperges, A. Castagna, T. Feldt, G. Green, M. L. Green, F. X. Lescure, E. Nicastri, R. Oda, K. Yo, E. Quiros-Roldan, A. Studemeister, J. Redinski, S. Ahmed, J. Burnett, D. Chelliah, D. Chen, S. Chihara, S. H. Cohen, J. Cunningham, A. D'Arminio Monforte, S. Ismail, H. Kato, G. Lapadula, E. L'Her, T. Maeno, S. Majumder, M.

- Massari, M. Mora-Rillo, Y. Mutoh, D. Nguyen, E. Verweij, A. Zoufaly, A. O. Osinusi, A. DeZure, Y. Zhao, L. Zhong, A. Chokkalingam, E. Elboudwarej, L. Telep, L. Timbs, I. Henne, S. Sellers, H. Cao, S. K. Tan, L. Winterbourne, P. Desai, R. Mera, A. Gaggar, R. P. Myers, D. M. Brainard, R. Childs, T. Flanigan, Compassionate use of remdesivir for patients with severe covid-19. *N. Engl. J. Med.* (2020).
12. R. E. Baker, W. Yang, G. A. Vecchi, C. J. E. Metcalf, and B. T. Grenfell, Susceptible supply limits the role of climate in the early sars-cov-2 pandemic. *Science* (2020).
13. M. S. Eichenbaum, S. Rebelo, and M. Trabandt, The macroeconomics of epidemics. *Tech. Rep.* (National Bureau of Economic Research, 2020).
14. S. Tian, N. Hu, J. Lou, K. Chen, X. Kang, Z. Xiang, H. Chen, D. Wang, N. Liu, D. Liu, G. Chen, Y. Zhang, D. Li, J. Li, H. Lian, S. Niu, L. Zhang, J. Zhang, Characteristics of covid-19 infection in Beijing. *J. Infect.* (2020).
15. M. Day, Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. *BMJ* **368**, m1165 (2020).
16. Z. Du, X. Xu, Y. Wu, L. Wang, B. J. Cowling, and L. A. Meyers, Serial interval of covid-19 among publicly reported confirmed cases. *medRxiv* (2020).
17. New York Times github source, *US coronavirus data* (2020); <https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv>.
18. NOAA Global Surface Summary of the Day, *Weather data for 2020* (2020); <https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00516>.
19. Kaiser Family Foundation, *State Data and Policy Actions to Address Coronavirus* (2020); <https://www.kff.org/health-costs/issue-brief/state-data-and-policy-actions-to-address-coronavirus/>.
20. J. H. Jones, Notes on R0. *California: Department of Anthropological Sciences* **323** (2007).
21. D. L. Smith, F. E. McKenzie, R. W. Snow, S. I. Hay, Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS biology* **5**(3) (2007).
22. M. Pollicott, H. Wang, H. Weiss, Extracting the time-dependent transmission rate from infection data via solution of an inverse ODE problem. *J. Boil. Dynam.* **6**(2), 509-523 (2012).