

Title: Proactive social distancing mitigates COVID-19 outbreaks within a month across 58 mainland China cities

Authors: Zhanwei Du¹, Xiaoke Xu², Lin Wang³, Spencer J. Fox¹, Benjamin J. Cowling⁴, Alison P. Galvani⁵ and Lauren Ancel Meyers^{1,6*}

Affiliations:

1. The University of Texas at Austin, Austin, Texas 78712, The United States of America
2. Dalian Minzu University, Dalian 116600, China.
3. University of Cambridge, Cambridge CB2 3EH, UK
4. The University of Hong Kong, 7 Sassoon Rd, Hong Kong SAR, China
5. Center for Infectious Disease Modeling and Analysis, Yale School of Public Health, New Haven, Connecticut, USA;
6. Santa Fe Institute, Santa Fe, New Mexico, The United States of America

* Correspondence: Lauren Ancel Meyers (laurenmeyers@austin.utexas.edu)

Appendix

We collected data from online reports published by the China CDC and health commissions. The data comprised 8,410 confirmed cases prior to February 15, 2020 across 271 cities in mainland China with known dates of symptom onset and classified as imported or locally infected. In any case, with missing symptom onset timings, we allocate a random value post the official announcement following the gamma distribution of the period from the onset of symptoms¹. We also aggregated data on the timing of seven different classes of social distancing measures in 58 cities outside Hubei province (**Appendix Table S1**).

Estimation of R_t

We estimate the time-varying reproduction numbers for the outbreaks in 58 cities using the R package ‘EpiEstim’² based on the method of Ref.³. Given time series data for incident imported and locally-infected cases and the distribution of serial intervals, the algorithm produces the time series of R_t with medians and 95% CI. We assume the serial interval

follows the Gamma distribution with a mean of 5.11 days and standard deviation of 2.68 days¹ and the length of the sliding time window for estimation of 7 days⁴.

Regression analysis

To assess the impact of intervention type and timing on the speed of containment, we applied a simple regression and variable selection method (**Table S2**). Specifically, we fit a linear regression model predicting the time until containment (i.e., days between symptom onset for the first reported case and the estimated 95% CI upper bound of R_t dropping below one) as given by $\chi = \sum_{i=1}^8 \beta_i T_i + \epsilon$, where β_i is the coefficient of the i th variable T_i and ϵ is the intercept. We analyze data from a subset consisting of 36 of the 58 cities in mainland China that implemented all eight interventions and contained the local outbreak by 14 February 2020. We used the best-subsets regression to identify the best fit model for all models containing, k , parameters where k ranges from one to eight, and chose our final model based on the Akaike Information Criterion corrected for small sample sizes (AICc)⁵.

Table S1. Variable definitions (for each city)

Symbol	Definition	Candidate model predictor
t_0	Day of symptom onset for the first reported case	
t_e	Day on which the earliest social distancing measure(s) was enacted	
$t_{\text{control}_{95}}$	Day that the upper 95% CI bound of R_t drops below one (without rebounding)	
χ	Containment period: the number of days between symptom onset of the first reported case (t_0) and the upper 95% CI of R_t dropping below one ($t_{\text{control}_{95}}$)	
T_{SD}	Days between first reported case (t_0) and implementation of the first social distancing measure implemented (t_e)	X
$T_{\text{entertainment}}$	Days between the first reported case (t_0) and ban on entertainment and public gatherings (e.g. bar, café, cinema)	X
T_{service}	Days between the first reported case (t_0) and restrictions on public services, including hospitals, schools, stores, and restaurants	X
$T_{\text{level-1}}$	Days between the first reported case (t_0) and the initiation of urban level-1 response for systematic testing and isolation of confirmed cases	X
$T_{\text{intra}_{\text{trans}}}$	Days between the first reported case (t_0) and suspension of intra-city public transport (bus and subway)	X
$T_{\text{inter}_{\text{trans}}}$	Days between the first reported case (t_0) and suspension of inbound and outbound travel (i.e. inter-city rail, highway, and air travel)	X
T_{report}	Days between the first reported case (t_0) and online posting of confirmed case reports	X

T_{assist}	Days between the first reported case (t_0) and the recruitment of governmental staff and volunteers to assist with quarantine and social distancing	X
---------------------	---	---

Table S2. Results of linear regression relating immediacy of interventions to speed of containment (i.e., days until upper 95% CI of $R_t < 1$) among 36 of 58 cities in mainland China who implemented all eight intervention measures and had sufficient data for prediction available χ by February 14, 2020. The parameters in the table were selected from among all of the candidates in Table S1 using the best-subsets method ⁶ that identifies the most informative combinations of predictors with respect to the Hurvich and Tsai's Information Criterion ⁷. The fitted model has $R^2=0.27$ with p -value < 0.001 .

Predictors	Coefficient	p-value
Intercept	14.37 (9.02, 19.72)	0.000
T_{SD}	2.41 (0.97, 3.86)	0.002
$T_{\text{level-1}}$	-1.87 (-3.14, -0.60)	0.005

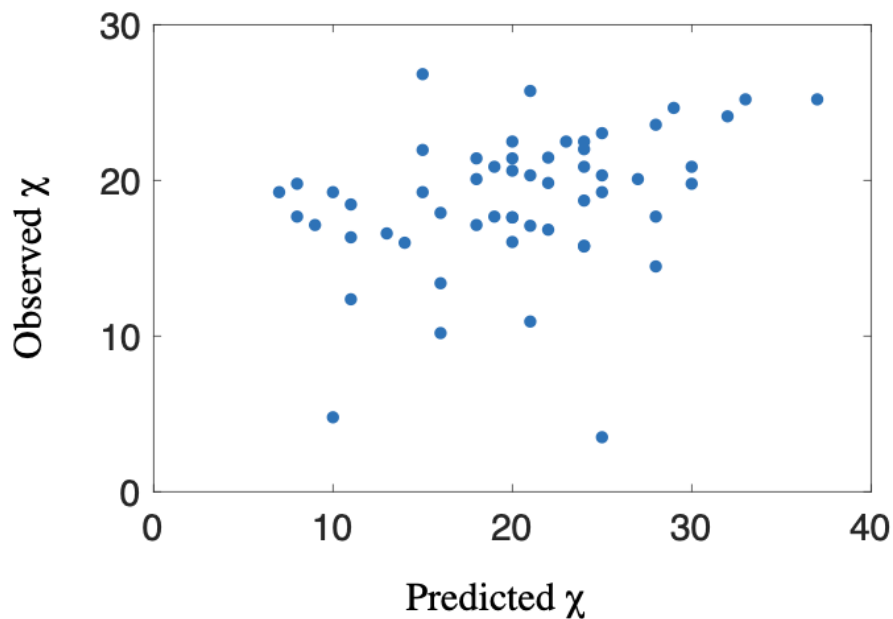


Figure S1. Predicted versus observed speed of containment (χ) for 36 cities in Mainland China based on the fitted regression model (Appendix Table S1). The data have a Pearson correlation coefficient of 0.52 with $P < 0.001$.

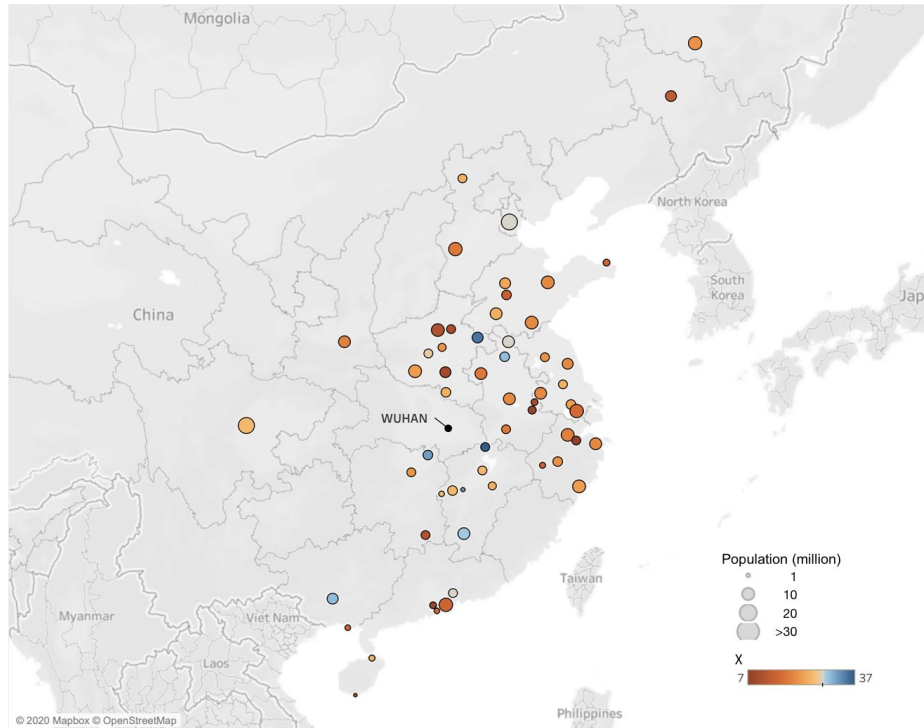


Figure S2. The number of days between the first reported case and containment (χ) for 58 cities in Mainland China that achieved containment prior to February 15, 2020. Outbreaks are considered contained when the estimated upper 95% CI bound of R_t drops below one without rebounding. Orange indicates the 48 cities that achieved containment within four weeks; blue indicates the 10 cities that did not.

Table S3. Distribution fits for the time between the first reported case and containment (χ) and time between the first reported case and the implementation of the first social distancing measure for 58 cities in Mainland China prior to February 15, 2020. Timing is calculated in terms of days from symptom onset of the first reported case in the city. Containment is defined by the first day that the estimated upper 95% CI for R_t permanently drops below one.

	Distribution	Shape [95% CI]	Scale [95% CI]	AIC
χ	Gamma	7.63 [5.343,10.895]	2.694 [1.864,3.893]	394.164
	Lognormal	2.956 [2.854,3.058]	0.389 [0.329,0.476]	398.975
	Weibull	3.282 [2.682,4.016]	22.937 [21.119,24.911]	390.329
T_{SD}	Gamma	6.121 [4.294,8.725]	2.079 [1.437,3.008]	349.986
	Lognormal	2.46 [2.342,2.577]	0.448 [0.379,0.549]	357.869
	Weibull	2.981 [2.436,3.649]	14.244 [13.008,15.597]	344.642

Acknowledgements

We thank Dr. Simon Cauchemez for helpful discussions. We acknowledge the financial support from NIH (U01 GM087719) and the National Natural Science Foundation of China (61773091).

References

- 1 Zhang J, Litvinova M, Wang W, *et al.* Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect Dis* 2020; published online April 2. DOI:10.1016/S1473-3099(20)30230-9.
- 2 Cori A. EpiEstim: Estimate time varying reproduction numbers from epidemic curves. 2019. URL <https://CRAN.R-project.org/package=EpiEstim> R package version; : 2–2.
- 3 Thompson RN, Stockwin JE, van Gaalen RD, *et al.* Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* 2019; **29**: 100356.
- 4 Cowling BJ, Ali ST, Ng TWY, *et al.* Impact assessment of non-pharmaceutical interventions against COVID-19 and influenza in Hong Kong: an observational study. *Epidemiology*. 2020; published online March 16. DOI:10.1101/2020.03.12.20034660.
- 5 Hurvich CM, Tsai C-L. Regression and time series model selection in small samples. *Biometrika* 1989; **76**: 297–307.
- 6 Yang H. The case for being automatic: introducing the automatic linear modeling (LINEAR) procedure in SPSS statistics. *Multiple Linear Regression Viewpoints* 2013; **39**: 27–37.
- 7 IBM Knowledge Center. https://www.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.modeler.help/idh_alm_model_selection.htm (accessed March 30, 2020).