Analyzing the impacts of public policy on COVID-19 transmission in Indiana: The role of model and dataset selection

George Mohler[*] Martin B. Short[†] Frederic Schoenberg[‡]
Daniel Sledge[§]

Abstract

Dynamic estimation of the reproduction number of COVID-19 is important for assessing the impact of public health measures on virus transmission. State and local decisions about whether to relax or strengthen mitigation measures are being made in part based on whether the reproduction number, R_t, falls below the self-sustaining value of 1. Using branching point process models and Indiana COVID-19 data as a case study, we show that estimates of the current value of R_t, and whether it is above or below 1, depend critically on particular choices of data selection and model specification and estimation. In particular, we find a range of R_t values from 0.47 to 1.20 as we vary the the type of estimator and input dataset. We present methods for model comparison and evaluation and then discuss the policy implications of our findings.

1 Introduction

During the first months of 2020, nations responded to the COVID-19 pandemic by adopting a variety of public health interventions, including contact

---

[*] Department of Computer and Information Science, Indiana University - Purdue University Indianapolis

[†] Department of Mathematics, Georgia Institute of Technology

[‡] Department of Statistics, University of California Los Angeles

[§] Department of Political Science, University of Texas Arlington

tracing, disease surveillance, and mandated social distancing [34, 35, 44]. Within the United States, the ongoing transmission of COVID-19 represents a serious public health threat and a significant strain on local, state, and federal resources. In the US, public health authority is largely vested in states and localities, with local decision-makers playing a critical role in shaping public health responses and in deploying resources during times of crisis. The federal government, meanwhile, plays a coordinating role through the CDC, funds research through agencies such as NIH, and helps shape the regulatory environment through the FDA and other agencies [41, 17, 8].

While this system has the potential to be highly responsive and adaptive, it is prone to problems including divergent outcomes across political jurisdictions and difficulty coordinating responses to emergent events. The lack of widespread testing in the early stages of transmission in the US forced policymakers to make decisions without high-quality data and foreclosed the possibility of contact tracing and disease surveillance, which might under different circumstances have proved a powerful public policy tool [21]. In the absence of testing and of pharmaceutical interventions such as vaccine or anti-viral therapies, social distancing measures (including shelter-in-place orders and mandated closure of non-essential businesses) emerged as the primary tool at the disposal of state and local decision-makers [18].

Despite the very real public health benefits of such interventions, they have potentially large economic and social costs, which are distributed unevenly across society. Over time, meanwhile, public and political pressure to relax public health interventions has increased. Policymakers, as a result, confront a complex set of problems and high levels of uncertainty [4, 12, 20]

Given these circumstances, the swift assessment of how differing public health strategies impact the transmission of COVID-19 is critical to fostering data-driven policy-making [16]. One means of measuring the impact of public health interventions is through the effective reproduction number of a virus, e.g. the average number of individuals an infected person directly infects. The initial reproduction number in the absence of interventions, R_0, for COVID-19 has been estimated across several studies to be around 3.28 (1.4, 6.5) [27]. A study conducted with data in China through mid February estimated that, as a result of public health measures, the effective time-varying reproduction number, $R(t)$, was reduced from 2 to 1 [46]. In Singapore, the impact of social distancing on R was estimated to be between 78.2% and 99.3% [26]. Recent research on interventions in Europe observed similar effects resulting from school closings, bans on mass gatherings, and other social distancing measures [14].

1.1 A need for model comparison and evaluation

While many of the forecasting models guiding policy makers on COVID-19 capture uncertainty in parameter estimates, a large number of these analyses are presented via stand-alone models; model comparison, evaluation and goodness-of-fit tests are often not presented. In addition, there are a variety of data sources available from data aggregation websites [11, 1] to local government data portals [2]. In this article, we show how estimates of the impact of policy interventions can vary depending on modeling and estimation choices, as well as the dataset used as input. While a variety of frameworks are in use for modeling COVID-19 [5], including agent based, compartmental models and branching point processes, here we focus on the latter point process type of model given our expertise and their wide use in estimating the reproduction number of a virus [29, 13, 38, 9, 45]. Within this framework we compare three choices for modeling the impact of interventions on COVID-19: 1) a step function modeling an immediate impact on R_t at key policy change dates that is employed in the highly cited paper [14], 2) a constant R_t up until a key policy change date followed by exponential decay [22, 24], and 3) a non-parametric histogram estimator that adapts to changes in the reproduction number over time. These choices are not meant to be exhaustive, but illustrative of the variation in estimates that can arise.

We then apply these models to both daily case and mortality COVID-19 data in Indiana from three different sources: a COVID-19 data portal in wide use hosted at Johns Hopkins University [11] (abbreviated as “jhu” throughout), the Covid Tracking project [1] (abbreviated as “covidtracking” throughout), and the local Indiana data portal hosted at [2] (abbreviated as “in.gov” throughout). One issue we investigate is the reporting lag of new cases and deaths posted to the Indiana state department of health dashboard, often several days after the testing date (with reporting sometimes paused on the weekend). While on the Indiana state health department website the data are retrospectively updated and corrected, data on aggregation websites like covidtracking and jhu are based on daily updates to cumulative counts and are not retrospectively corrected. Consequently, artificial peaks and valleys are present in the covidtracking and jhu count data. For this reason Indiana provides a unique test case to investigate different choices in data processing and their impact on critical parameters in models for the spread of Covid-19.

In Section 2 we present the branching point process modeling framework and discuss how these models can be estimated from data. Then, in Section 3 we present our results when the models are applied to Indiana COVID-19 data. We show that estimates of the current value of R_t in Indiana, and whether it is above or below 1, depend on the model and dataset used for estimation. We also present several methods that can either be used to compare competing models or used to assess the goodness-of-fit of a particular model. We find a range of R_t values from 0.47 to 1.20 as we vary the the type of estimator and input dataset. In Section 4 we discuss the policy implications of our findings.

2 Methods

We consider a branching point process [30, 19] framework to estimate a time-varying reproduction number $R(t)$ [9, 45, 32, 39]. The conditional intensity (rate) of infections is modeled as

$$\lambda(t) = \mu + \sum_{t>t_i} R(t_i)w(t-t_i), \tag{1}$$

where $w(t)$ and $R(t)$ are the inter-infection time distribution and dynamic reproduction number respectively. We also include an exogeneous rate μ modeling imported infections. The point process governed by Equation 1 can be viewed as an approximation to the common SIR model of infectious diseases during the initial phase of an epidemic when the total infections is small compared to the overall population size and $w(t)$ is specified to be exponential [36]. When $w(t)$ is chosen to be gamma distributed, the Hawkes process also can approximate staged compartment models, like SEIR, if the average waiting time in each compartment is equal [28].

We consider three competing models for $R(t)$:

1. A **step function** that changes at the Indiana stay-at-home order date t_{sh}, effective on March 24, 2020:

$$R(t) = \begin{cases} R_0 & t \leq t_{sh} \\ R_1 & t > t_{sh} \end{cases} \tag{2}$$

This is analogous to the step function used in [14] to assess the impact of public health interventions on COVID-19 in Europe.

2. An **exponential decay** [22, 24] after the stay-at-home date of the form:
$$R(t) = \begin{cases} R_0 & t \leq t_{sh} \\ R_0 \exp(-ct) & t > t_{sh} \end{cases} \tag{3}$$

3. A **histogram estimator** that adapts to dynamic changes in R_t over time
$$R(t) = \sum_{k=1}^{B} r_k 1\{t \in I_k\}. \tag{4}$$
Here the I_k are intervals discretizing time, B is the number of such intervals, and r_k is the estimated reproduction rate in interval k. In the remainder of the paper we use a bin-width of 1 week. We also merge the bins of the first 3 weeks due to the low number of events during that time period.

The branching process can be estimated via an expectation-maximization (EM) algorithm for maximum likelihood inference [43, 31, 25]. Given initial guesses for the model parameters and r_k, the EM algorithm iteratively updates the parameters and branching probabilities by alternating between the **E-step update**:

$$p_{ij} = R(t_j) w(t_i - t_j)/\lambda(t_i) \tag{5}$$

$$p_{ii} = \mu/\lambda(t_i) \tag{6}$$

and **M-step update:**

$$w(t) \sim MLE(\{t_i - t_j; p_{ij}\}) \tag{7}$$

$$\mu = \sum_i p_{ii}/T \tag{8}$$

$$r_k = \sum_{t_i > t_j} p_{ij} 1\{t_j \in I_k\}/N_k \tag{9}$$

where T is the total length of the observation period, N_k is the total number of events in interval k, and $w(t)$ is estimated via weighted MLE using the inter-event times as observations and branching probabilities as weights. While $w(t)$ can be estimated using a Weibull, Gamma, or log-normal distribution, we use a non-parametric histogram estimator with bin-wideth of 1 day to prevent model mis-specification.

Competing models can be compared using the Akaike Information Criterion [3], $AIC = 2p - 2\log(L)$. The AIC balances goodness of fit measured by the log-likelihood, $\log(L)$, and over-parametrization by penalizing the number of parameters, p (lower AIC is better). Alternatively, the goodness of fit of a branching process model can be assessed using residual analysis of rescaled event times [33],

$$\tau_i = \int_0^{t_i} \lambda(t)dt \tag{10}$$

The rescaled times are distributed according to a unit rate Poisson process if the model is correctly specified.

3 Results

We apply the estimation procedure outlined above to Indiana COVID-19 case and mortality daily counts (new cases rather than cumulative) from March 5, 2020 to April 26, 2020. While school and business closings occurred on March 16, 2020, there is limited case data available before this date and we therefore assume R_0 is constant across all models up until the stay-at-home order on March 24, 2020.

We present estimated R_t curves in Figure 1, estimated intensities λ_t (Figure 2) and inter-infection time distributions $\omega(t)$ (Figure 3), and in Table 1 we present the corresponding AIC values for the different models (step, exponential, histogram), different data sources (Johns Hopkins [11], Covid Tracking Project [1], State of Indiana [2]), and different data types (case counts or mortality).

The first observation of note is that the AIC values are lower for all models using the local in.gov data rather than data from aggregation websites. In the case of Indiana, new cases and deaths on a given date are often reported to the state department of health several days after the fact (with reporting often paused on the weekend). While on the state health department website the data are retrospectively updated and corrected, the aggregation website data are based on daily updates to cumulative counts (and they do not go back and correct historical cumulative count data). Consequently, artificial peaks and valleys are present in the covidtracking and jhu estimates (see Figure 2). Our findings are consistent with recent recommendations to use local data whenever possible [23]. Variation in estimates of R_t that arise from using case or death counts is higher earlier in the Indiana epidemic. Estimates of R_t are initially as high as 5 when using case data but are between 2 and 4 using death counts. The high value of R_t early on for cases may be due to the initial lack of testing followed by a rapid growth in testing over a several week period.

We find a larger variation in R_t across models than across data sources. For example, when applied to jhu case data the models provide final R_t values (at the end date of April 26th) of 1.20 (histogram), 0.83 (step) and 0.66 (exponential). Across the datasets the estimated value of R_t at the final time tends to be higher for the histogram as it adapts to changes in reproduction after the stay-at-home order on March 24th in Indiana. While R_t fell below 1 according to the histogram several weeks after the order, it later rose back above 1 (possibly due to lack of adherence to social distancing or the emergence of new clusters in counties outside of the Indianapolis Metropolitan area). The histogram estimator also consistently has the lowest AIC values of the competing models because of its ability to adapt to local changes in time.

While the AIC is useful for comparing competing models, goodness-of-fit of a point process model can be evaluated using residual analysis. In Figure 2 we plot point process intensities fit to Indiana COVID-19 cases and deaths per day from March 5, 2020 to April 26, 2020. We again use a histogram estimate for the inter-infection time distribution $\omega(t)$ which we plot in Figure 3. In Figure 4, we plot the normalized cumulative distribution of rescaled event times and compare them to confidence bounds of the cumulative distribution for a unit rate Poisson process. We find that the estimated intensity using the histogram and exponential decay for R_t provides a good fit to Indiana new deaths per day, whereas the intensity that uses a step function for R_t under-estimates the empirical death rate. This is in comparison to Figure 2, where all of the intensities appear to give plausible fits to the data based upon visual inspection.

4 Discussion

State and local policymakers are using mathematical models and projections to inform decisions about whether and how to shift away from the most stringent forms of social distancing. In many cases, policymakers are informed by a single model that has not been compared to other options that may yield different results. The IHME model [10] that the White House relies upon

model	dataset	datatype	log-likelihood	AIC	R_{final}
histogram	covidtracking	deaths	1965.80	**-3921.60**	1.17
step	covidtracking	deaths	1824.10	-3644.20	0.68
exponential	covidtracking	deaths	1947.40	-3890.90	0.49
histogram	in.gov	deaths	1980.10	**-3950.20**	0.97
step	in.gov	deaths	1951.50	-3899.00	0.95
exponential	in.gov	deaths	1967.50	-3931.00	0.59
histogram	jhu	deaths	1954.60	**-3899.30**	1.12
step	jhu	deaths	1877.30	-3750.60	0.85
exponential	jhu	deaths	1944.60	-3885.20	0.47
histogram	covidtracking	cases	81521.00	**-163030.00**	1.19
step	covidtracking	cases	80809.00	-161610.00	0.83
exponential	covidtracking	cases	80838.00	-161670.00	0.67
histogram	in.gov	cases	81534.00	**-163060.00**	1.19
step	in.gov	cases	80727.00	-161450.00	0.83
exponential	in.gov	cases	80853.00	-161700.00	0.67
histogram	jhu	cases	81516.00	**-163020.00**	1.20
step	jhu	cases	80741.00	-161480.00	0.83
exponential	jhu	cases	80793.00	-161580.00	0.66

Table 1: Comparison of dynamic R_t model fits across estimator type, data sources and data types for Indiana COVID-19 data. Number of bins is $B = 5$ for the histogram estimator. The other two models each have $p = 2$ parameters for the dynamic reproduction number.

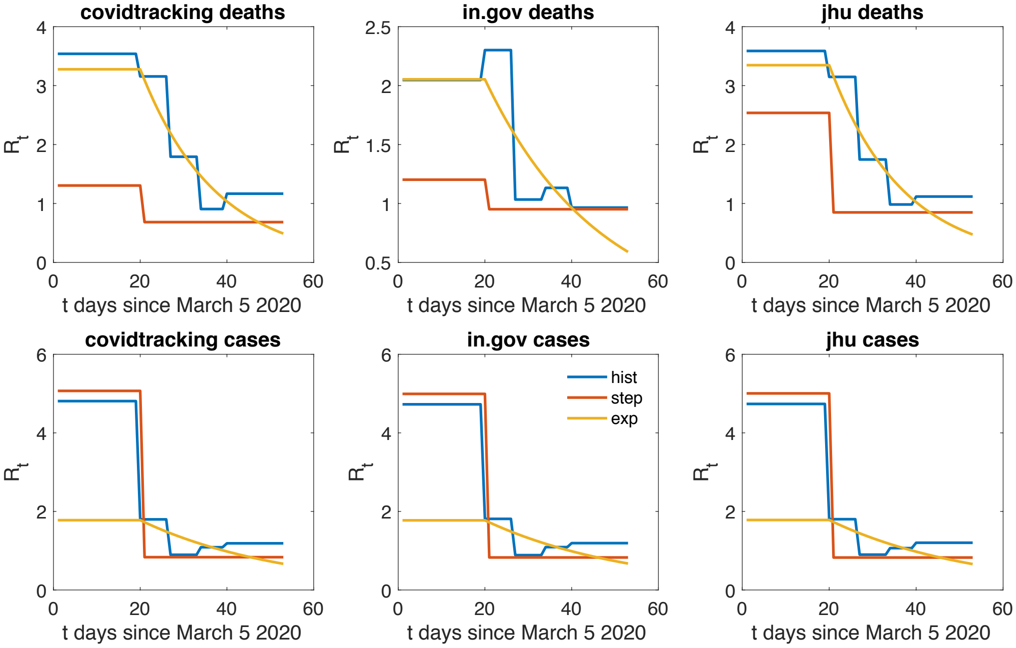

Figure 1: Fitted R_t curves for histogram estimator (blue), step function (red) and exponential decay (yellow) applied to different Indiana COVID-19 data types and data sources.

(according to New York Times reporting [7]) has generated more optimistic projections compared to several other high profile alternatives [7]. In a recent fivethirtyeight article, Nate Silver analyzed U.S. state level COVID-19 cases [40], finding that Indiana "may be past the peak." Here, however, we show that such a finding depends on the type of model and dataset used.

Already, a number of states have started to relax mandated social distancing measures as COVID-19's reproduction number has declined. In Georgia, Gov. Brian Kemp allowed a wide variety of businesses, including gyms, tattoo parlors, and nail salons to reopen, with required social distancing and sanitary measures in place [37]. In Texas, meanwhile, Gov. Greg Abbott issued an order allowing restaurants, movie theaters, and malls to open with limited capacity as of May 1 [42].

A likely scenario moving forward is that social distancing measures in many areas will be relaxed as transmission is reduced and new daily cases

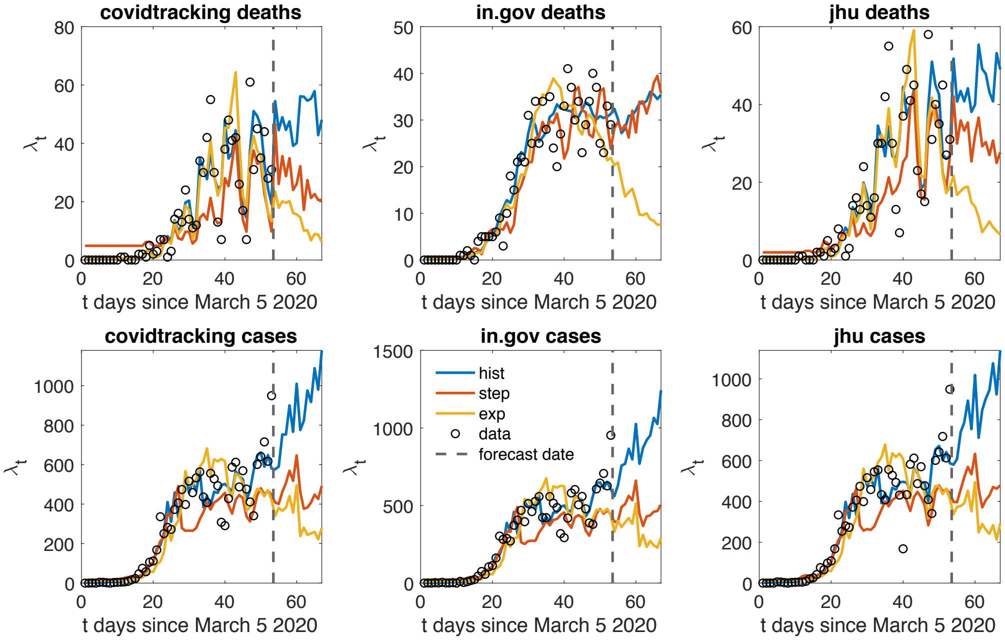

Figure 2: Fitted intensity λ_t curves for histogram estimator (blue), step function (red) and exponential decay (yellow) applied to different Indiana COVID-19 data types and data sources. Example realizations of λ_t simulated for 14 days past the current date (dashed line) show growth or decay depending on whether $R_t > 1$ or $R_t < 1$.

and deaths decrease. Relaxation, in turn, may be followed by periods where R_t again rises above 1, resulting in new cases and perhaps in the decision to engage in subsequent rounds of social distancing. Similar patterns were observed during the 1918 Influenza pandemic [6]. As a result, policy decisions about social distancing measures based on the observed value of R_t may lead to a situation in which the duration of the pandemic is significantly extended [15].

During the early stages of the COVID-19 pandemic, it was important that forecasting models be rapidly developed in order to inform decision-making. Now that these models have been implemented and knowledge of COVID-19 transmission dynamics has matured, research is needed on the trade-offs between competing model and data choices. Here we showed that, when varying the estimator of dynamic R_t between three simple choices, along

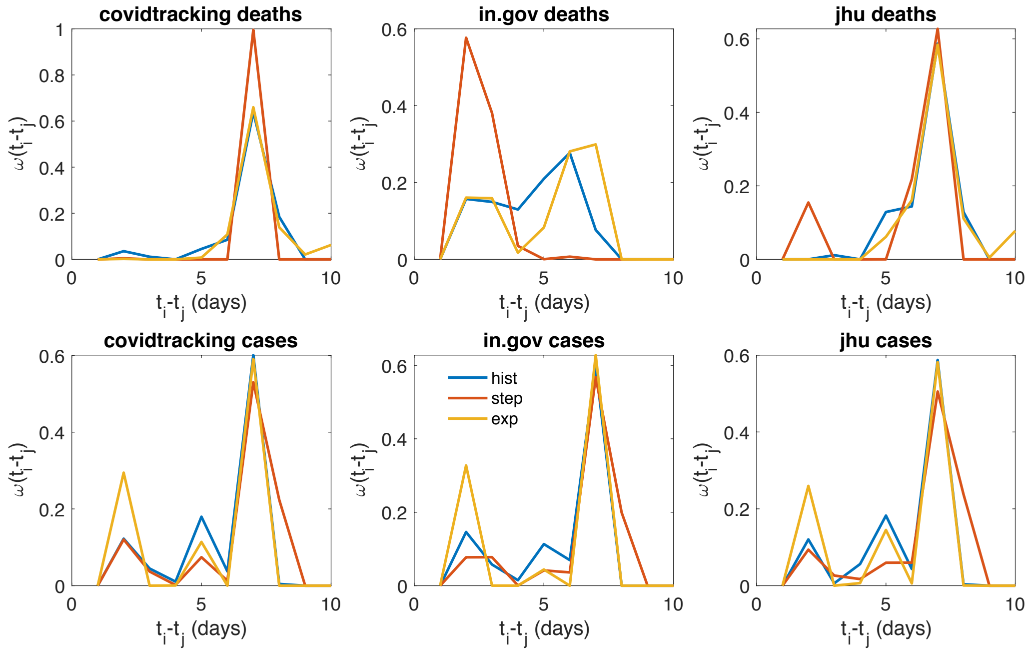

Figure 3: Fitted inter-infection time distribution $\omega(t_i - t_j)$ curves for histogram estimator (blue), step function (red) and exponential decay (yellow) applied to different Indiana COVID-19 data types and data sources.

with 3 different data sources, we get dramatically different answers to the question of whether public health interventions in Indiana have reduced the reproduction number to below one. In helping to inform crucial decisions about public health interventions, this suggests, modelling efforts should be careful to consider model specification, goodness-of-fit, and sensitivity to the choice of input data. For policymakers, these results emphasize both the complex nature of the pandemic and the critical importance of acknowledging that findings about the impacts of public policy interventions will vary as a result of model and data selection.

Acknowledgements

This research was supported by NSF grants SCC-1737585, ATD-1737996 and ATD-1737925.

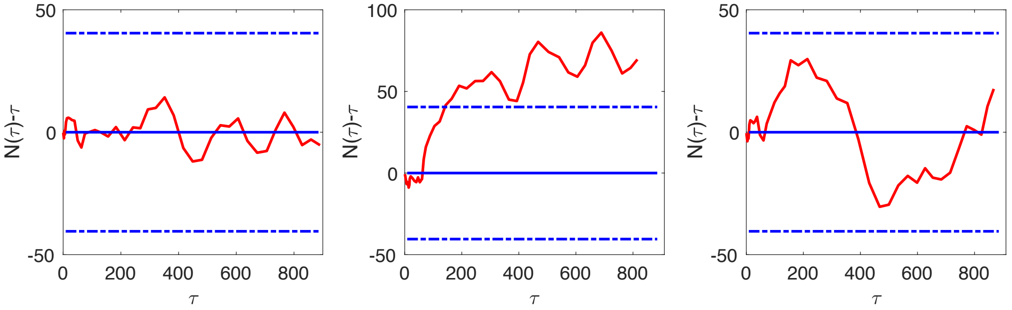

Figure 4: Normalized cumulative distribution of rescaled event times along with 95% error bounds of the Kolmogorov-Smirnov statistic for histogram-based intensity (left), step function (middle), and exponential (right) applied to in.gov daily death counts.

References

[1] https://covidtracking.com/.

[2] https://www.coronavirus.in.gov/.

[3] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[4] Frank R Baumgartner and Bryan D Jones. *The politics of information: Problem definition and the course of public policy in America*. University of Chicago Press, 2015.

[5] Andrea L Bertozzi, Elisa Franco, George Mohler, Martin B Short, and Daniel Sledge. The challenges of modeling and forecasting the spread of covid-19. *arXiv preprint arXiv:2004.04741*, 2020.

[6] Martin CJ Bootsma and Neil M Ferguson. The effect of public health measures on the 1918 influenza pandemic in us cities. *Proceedings of the National Academy of Sciences*, 104(18):7588–7593, 2007.

[7] Quoctrung Bui, Josh Katz, Alicia Parlapiano, and Margot Sanger-Katz. What 5 coronavirus models say the next month will look like. *New York Times*, 2020.

[8] Daniel Carpenter. *Reputation and power: organizational image and pharmaceutical regulation at the FDA*, volume 137. Princeton University Press, 2014.

[9] Simon Cauchemez, Pierre-Yves Boëlle, Christl A Donnelly, Neil M Ferguson, Guy Thomas, Gabriel M Leung, Anthony J Hedley, Roy M Anderson, and Alain-Jacques Valleron. Real-time estimates in early detection of sars. *Emerging infectious diseases*, 12(1):110, 2006.

[10] 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.

[11] Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track covid-19 in real time. *The Lancet Infectious Diseases*, 2020.

[12] Derek A Epp. *The structure of policy change.* University of Chicago Press, 2018.

[13] CP Farrington, MN Kanaan, and NJ Gay. Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics*, 4(2):279–295, 2003.

[14] Seth Flaxman, Swapnil Mishra, Axel Gandy, et al. Estimating the number of infections and the impact of non-pharmaceutical interventions on covid-19 in 11 european countries. *Imperial College COVID-19 Response Team*, 30, 2020.

[15] Elisa Franco. A feedback sir (fsir) model highlights advantages and limitations of infection-based social distancing, 2020.

[16] Thomas R Frieden. Evidence for health decision making—beyond randomized, controlled trials. *New England Journal of Medicine*, 377(5):465–475, 2017.

[17] Lawrence O Gostin and James G Hodge. Us emergency legal responses to novel coronavirus: Balancing public health and civil liberties. *Jama*, 323(12):1131–1132, 2020.

[18] Lawrence O. Gostin and Lindsay F. Wiley. Governmental Public Health Powers During the COVID-19 Pandemic: Stay-at-home Orders, Business Closures, and Travel Restrictions. *JAMA*, 04 2020.

[19] Alan G Hawkes and David Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.

[20] Brian W Head et al. Wicked problems in public policy. *Public policy*, 3(2):101, 2008.

[21] Joel Hellewell, Sam Abbott, Amy Gimma, Nikos I Bosse, Christopher I Jarvis, Timothy W Russell, James D Munday, Adam J Kucharski, W John Edmunds, Fiona Sun, et al. Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, 2020.

[22] B Ivorra, MR Ferrández, M Vela-Pérez, and AM Ramos. Mathematical modeling of the spread of the coronavirus disease 2019 (covid-19) considering its particular characteristics. the case of china. Technical report, Technical report, MOMAT, 03 2020. URL: https://doi-org. usm. idm. oclc. org . . . , 2020.

[23] Nicholas P Jewell, Joseph A Lewnard, and Britta L Jewell. Predictive mathematical models of the covid-19 pandemic: Underlying principles and value of projections. *JAMA*, 2020.

[24] Phenyo E Lekone and Bärbel F Finkenstädt. Statistical inference in a stochastic epidemic seir model with control intervention: Ebola as a case study. *Biometrics*, 62(4):1170–1177, 2006.

[25] Erik Lewis and George Mohler. A nonparametric em algorithm for multiscale hawkes processes. *preprint*, 2011.

[26] Joseph A Lewnard and Nathan C Lo. Scientific and ethical basis for social-distancing interventions against covid-19. *The Lancet. Infectious diseases*, 2020.

[27] Ying Liu, Albert A Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. The reproductive number of covid-19 is higher compared to sars coronavirus. *Journal of travel medicine*, 2020.

[28] Alun L Lloyd. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1470):985–993, 2001.

[29] Sebastian Meyer, Leonhard Held, and Michael Höhle. Spatio-temporal analysis of epidemic phenomena using the R package surveillance. *Journal of Statistical Software*, 77(11), 2017.

[30] Sebastian Meyer, Leonhard Held, and Michael Höhle. Spatio-temporal analysis of epidemic phenomena using the r package surveillance. *Journal of Statistical Software*, 77(11), 2017.

[31] George O Mohler, Martin B Short, P Jeffrey Brantingham, Frederic Paik Schoenberg, and George E Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108, 2011.

[32] Thomas Obadia, Romana Haneef, and Pierre-Yves Boëlle. The r0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC medical informatics and decision making*, 12(1):147, 2012.

[33] Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401):9–27, 1988.

[34] An Pan, Li Liu, Chaolong Wang, Huan Guo, Xingjie Hao, Qi Wang, Jiao Huang, Na He, Hongjie Yu, Xihong Lin, et al. Association of public health interventions with the epidemiology of the covid-19 outbreak in wuhan, china. *Jama*, 2020.

[35] Andrea Remuzzi and Giuseppe Remuzzi. Covid-19 and italy: what next? *The Lancet*, 2020.

[36] Marian-Andrei Rizoiu, Swapnil Mishra, Quyu Kong, Mark Carman, and Lexing Xie. Sir-hawkes: linking epidemic models and hawkes processes to model diffusions in finite populations. In *Proceedings of the 2018 World Wide Web Conference*, pages 419–428, 2018.

[37] Rick Rojas and Michael Cooper. Georgia, tennessee and south carolina say businesses can reopen soon. *New York Times*, 2020.

[38] F.P. Schoenberg, M. Hoffmann, and R. Harrigan. A recursive point process model for infectious diseases. *Annals of the Institute of Statistical Mathematics*, 71(5):1271–1287", 2019.

[39] Frederic Paik Schoenberg, Marc Hoffmann, and Ryan J Harrigan. A recursive point process model for infectious diseases. *Annals of the Institute of Statistical Mathematics*, 71(5):1271–1287, 2019.

[40] Nate Silver. Coronavirus cases are still growing in many u.s. states. *FiveThirtyEight.com*, 2020.

[41] Daniel Sledge. *Health Divided: Public Health and Individual Medicine in the Making of the Modern American State*. University Press of Kansas, 2017.

[42] Konstantinand Toropin and LeBlanc. Texas governor will allow state's stay-at-home order to end on thursday and open businesses friday. *CNN*, 2020.

[43] Alejandro Veen and Frederic P Schoenberg. Estimation of space–time branching process models in seismology using an em–type algorithm. *Journal of the American Statistical Association*, 103(482):614–624, 2008.

[44] Rochelle P. Walensky and Carlos del Rio. From Mitigation to Containment of the COVID-19 Pandemic: Putting the SARS-CoV-2 Genie Back in the Bottle. *JAMA*, 04 2020.

[45] Jacco Wallinga and Peter Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of epidemiology*, 160(6):509–516, 2004.

[46] Chong You, Yuhao Deng, Wenjie Hu, Jiarui Sun, Qiushi Lin, Feng Zhou, Cheng Heng Pang, Yuan Zhang, Zhengchao Chen, and Xiao-Hua Zhou. Estimation of the time-varying reproduction number of covid-19 outbreak in china. *Available at SSRN 3539694*, 2020.