Survival-Convolution Models for Predicting COVID-19 Cases and Assessing Effects of Mitigation Strategies [1]

Qinxia Wang[a], Shanghong Xie[a], Yuanjia Wang[a], and Donglin Zeng[b]

a: Department of Biostatistics, Mailman School of Public Health, Columbia University

b: Department of Biostatistics, Gillings School of Public Health, University of North Carolina at Chapal Hill

---

[1] Wang Q, Xie S et al. (2020). Survival-Convolution Models for Predicting COVID-19 Cases and Assessing Effects of Mitigation Strategies. *Frontiers in Public Health*. 8:325. Github site. *Supported by a COVID-19 supplemental grant*.

Acknowledgments

- Authors: Qinxia Wang, Shanghong Xie, Yuanjia Wang, Donglin Zeng

frontiers in Public Health

BRIEF RESEARCH REPORT
published: 03 July 2020
doi: 10.3389/fpubh.2020.00325

Check for updates

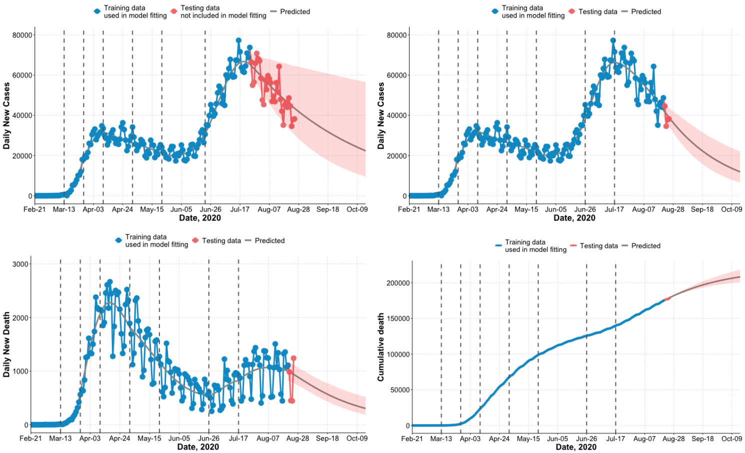
Survival-Convolution Models for Predicting COVID-19 Cases and Assessing Effects of Mitigation Strategies

Qinxia Wang[1], *Shanghong Xie*[1], *Yuanjia Wang*[1*] *and Donglin Zeng*[2*]

[1] *Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY, United States,*
[2] *Department of Biostatistics, Gillings School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States*

- Github webpage with weekly updates:
  `https://github.com/COVID19BIOSTAT/covid19_prediction`
- Funding agency (NIGMS)

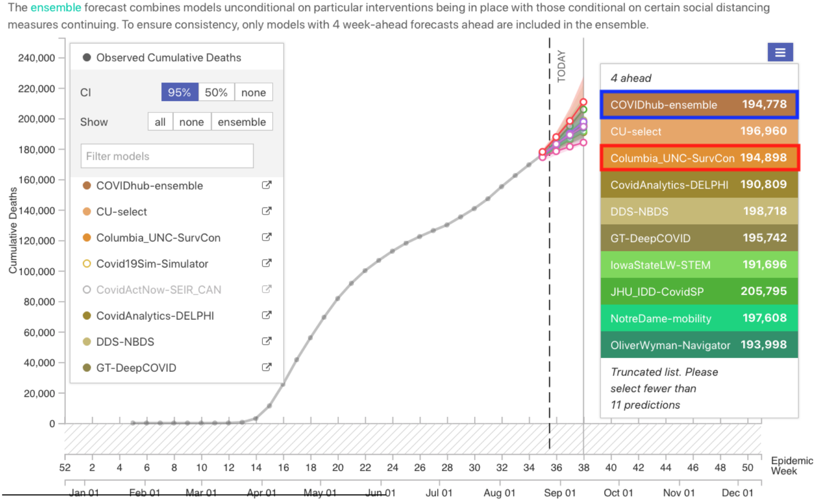
Case and Death Forecasts of COVID-19 in US



We submit our forecasts to COVID Forecast Hub, which is used by the US Centers for Disease Control and Prevention (CDC) and the data journalism site FiveThirtyEight.

CDC Forecasts

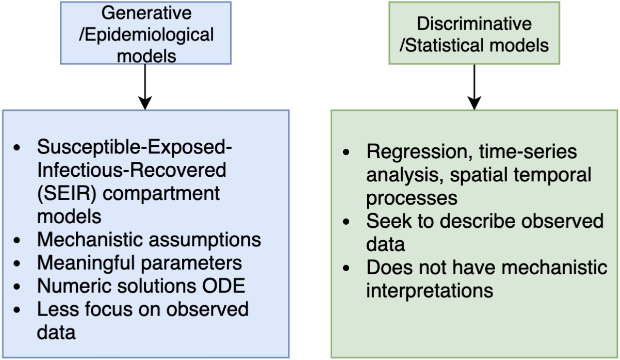
The ensemble forecast predicts that 187,000 to 205,000 total COVID-19 deaths will be reported by September 12[2].



2 https://viz.covid19forecasthub.org

Infectious Disease Modeling

Fig. Epidemiological Models and Statistical Models



Can we strike a balance between the two camps? Provide estimates of important parameters (R_t).

Study Goals

- Demonstrate a simple and robust model to predict COVID-19 epidemic course
- Leverage natural experiment and use a longitudinal pre-post design to evaluate mitigation strategies and compare between countries to inform decision making

Methods

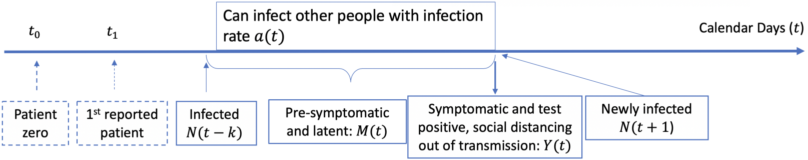
Modeling Considerations

- To make timely decisions, need to capture current trend of epidemic, the best statistics to predict disease spread is daily confirmed cases and daily reported deaths.
- Model national level instead of state level.
- Important factors for modeling:
  - SARS-CoV-2 virus has a long incubation period (up to 14 days, extreme case 21 days)
  - Highly infectious in the pre-symptomatic phase: 40% transmission occurred during this phase (US CDC)
  - Time-varying transmission rate as public health interventions are implemented and societal behavior changes
  - Intervention effect may be time-dependent
- For policy decision making, robustness, transparency, parsimony are important.

Modeling Scheme

Survival-Convolution Model

- $M(t) = \sum_{k=0}^{\infty} N(t-k)S(k)$
- $Y(t) = \sum_{k=0}^{\infty} N(t-k)[S(k) - S(k+1)]$
- $N(t+1) = a(t)[M(t) - Y(t)]$



- ► $M(t)$ total number of latent infected cases in the transmission chain by t.
- ► $N(t)$ number of new infections at t.
- ► $S(k)$ proportion of infected cases remaining infectious and in the transmission chain after k days of exposure (the discrete survival function of time to out of transmission).
- ► $Y(t)$ number of subjects out of transmission.
- ► $N(t+1) = a(t)[M(t) - Y(t)] = a(t)\sum_{k=0}^{\infty} N(t-k)S(k+1)$ gives a convolution update for the new cases using the past infections.

Modeling Transmission Rate

Model $a(t)$ as non-negative, piece-wise linear functions with knots placed at meaningful event times:

- Before report of first case t_1, transmission rate is a constant a_0.
- Once the first positive case was reported, the public starts to respond, so model the transmission rate with a linear (decreasing) trend.
- When a massive public health intervention (e.g., nation-wide lockdown) is implemented, introduce an additional linear function with a new slope parameter.
- The simplest model has only 2 parameters (a_0, a_1)!

Time-varying Effective Reproduction Number

Effective reproduction number (R_t): the expected average number of secondary cases infected by a primary case in a population at time t while accounting for the entire incubation period (Cori et al. 2013)

$$R_t = \sum_{k=0}^{C} a(t+k)G(k),$$

$G(k)$ distribution of serial intervals between primary and secondary cases.

R_t captures the temporal changes in the disease spread.

Evaluation of Public Health Intervention Effect

Quasi-experiments longitudinal pre-post intervention design. Often used to study health policies when randomized trials are not feasible.

Assumptions:

- Local randomization: subjects infected before or after intervention are similar within a short period of time
- Continuity: the trend before implementation continues had the intervention not been implemented

Intervention effects estimated as the difference in the rate of change of $a(t)$ before and after an intervention takes place. Corrects for the natural decline of the infection rate function over time.

Estimation Using Confirmed Daily Cases

Daily reported new cases $Y_o(t_1), Y_o(t_1+1), Y_o(t_0+2), \ldots.$

- Observation model:

$$Y_o(t_i) = Y(t_i) + \sqrt{Y(t_i)}\epsilon(t_i),$$

- Squared loss function of the predicted confirmed numbers after a square-root transformation:

$$\sum_{t_1 \le t \le t_*} \left[\sqrt{Y_o(t)} - \sqrt{Y(t;\theta)}\right]^2$$

- Estimation: gradient descent implemented in Tensorflow.
- Inference: permutation of predicted standardized residuals over time $\tilde{\epsilon}(t) = \left[Y_o(t) - Y(t;\widehat{\theta})\right] / \sqrt{Y(t;\widehat{\theta})}$.

Analysis Details

Data

Numbers of daily confirmed new cases and deaths can be obtained from many public sources.

- JHU Center for System Science and Engineering (CSSE) `https://github.com/CSSEGISandData/COVID-19`
- WHO, New York Times
- We obtain data from a publicly available database that curates and validates multiple sources on COVID-19 statistics: `www.worldometers.info/coronavirus`

Analysis Goals

Countries to analyze: China, South Korea, Italy, US

- China, South Korea: prediction of future daily new cases, expected number of cases and latent cases, when R_t will be reduced to below 1.0, and when the epidemic will be controlled.
- Italy: estimate the effects of mitigation strategies.
- US: evaluate the impact of an intervention at a future date (e.g., lifting mitigation measures) that will change the epidemic.

Model Setups

China and South Korea: a single piece for $a(t)$. About two weeks data for training and the rest of data up to May 10 as testing data. Infection rate:

$$a(t) = \begin{cases} a_0^+ & t < t_1 \\ (a_0 + a_1(t - t_1))^+ & t \geq t_1 \end{cases}$$

3 parameters: t_0, a_0, a_1.

Goal: examine prediction performance.

Model Setups

Italy: 4 pieces. A knot at nation-wide lockdown (March 11, t_2) to capture the immediate effect of lockdown. Two knots with two weeks apart afterwards, for short-term and longer-term intervention effect (March 25, t_3; April 8, t_4).

$$a(t) = \begin{cases} a_0^+ & t < t_1, \\ (a_0 + a_1(t - t_1))^+ & t_1 \le t < t_2, \\ (a_0 + a_1(t_2 - t_1) + a_2(t - t_2))^+ & t_2 \le t < t_3, \\ (a_0 + a_1(t_2 - t_1) + a_2(t_3 - t_2) + a_3(t - t_3))^+ & t_3 \le t < t_4 \\ (a_0 + a_1(t_2 - t_1) + a_2(t_3 - t_2) + a_3(t - t_3) + a_4(t - t_4))^+ & t \ge t_4. \end{cases}$$

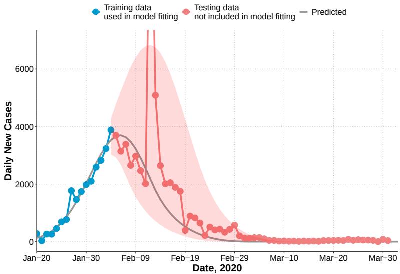
Goal: estimate lockdown effect, i.e., immediate effect (a_2 vs a_1), short-term (a_3 vs a_1), longer-term (a_4 vs a_1).

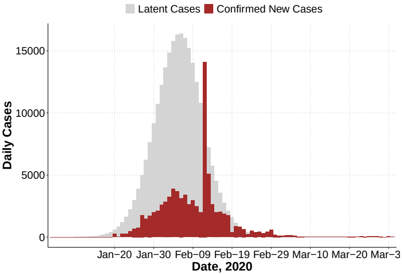
US (10-20 days behind Italy): 5 pieces. A knot at the declaration of national emergency (March 13, t_2) and 3 knots (2-week apart) afterwards (March 27, April 10, April 24).

Results

China

Training data: January 20 to February 4; testing data: February 5 to May 10.

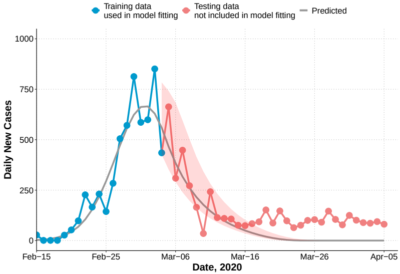


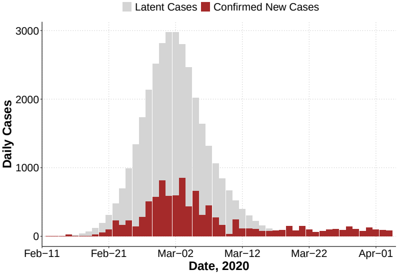


- t_0: Jan 3 (17 days before first report)
- Predicted total: 58,415; 95% CI: (42,516, 133,083)
- Observed total: 82,901. Two outliers on Feb 12, 13. Excluding outliers: 62,356.

South Korea

Training data: February 15 to March 4; testing data: March 5 to May 10.

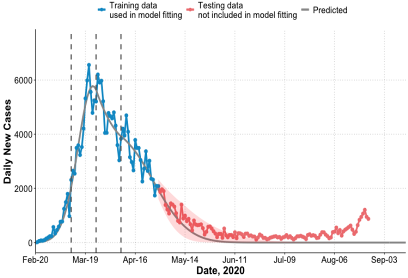


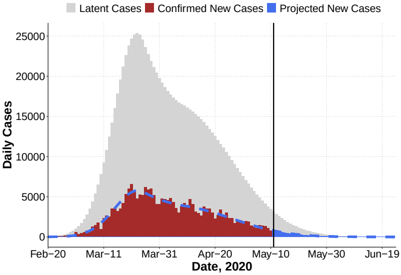


- t_0: Feb 11 (4 days before first report)
- Small outbreak after March 15 not captured
- Predicted total by March 15: 7,816
- Observed total: 8,162.

Italy

Training data: February 20 to April 29 (7 weeks after lockdown); testing data: April 30 to August 25.

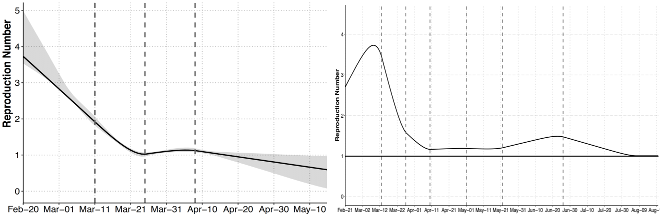




- t_0: Feb 10 (10 days before first report)
- Predicted total by May 31: 223,079 (CI: 202,940, 263,152)
- Observed total: 232,997
- Rate of decrease after the peak is slower than rising (asymmetric)

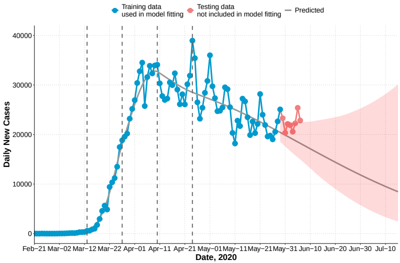
Effective Reproduction Numbers R_t

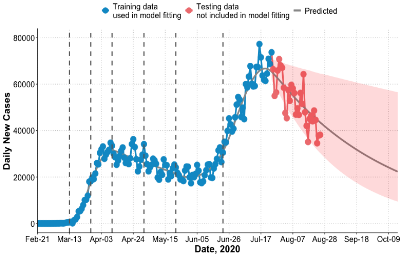
Fig. R_t in Italy (left) and US (right)



- R_t reduced to < 1.0 in 14 days in China and South Korea. R_t reduced to 1.02 in 42 days in Italy (remained 1.0 for 3 weeks). Lockdown in Italy did not significantly further reduce the rate of decrease ($p > 0.05$). US R_t reduced to 1.1 in 50 days, flat for 6 weeks before increasing again.

US





Left: Training data: February 21 to May 29. Total cases: 2.7 million, total deaths: 157K. Date with < 100 cases: Nov 9; Duration: 262. Already observed an uptick by early June.

Right: Training data: February 21 to July 24. Already predicted the decreasing trend.

Current Forecasts

Death model: Jointly fit incidence cases and incidence deaths.

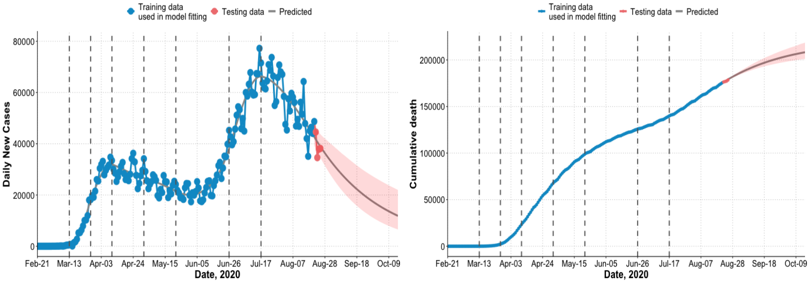


Figure. Training data: February 21 to August 16.
Total cases: 6.8 million, total deaths: 214K. Date with $<$ 100 cases: Mid December.

Discussion

Discussion

Propose a simple model for predicting daily new cases, latent cases, R_t, and comparison of mitigation strategies. Works well for short term but not long term without assumptions.

For model building, may not need big data or any sophisticated predictive analytics, but require:

- reliable data sources (official reports);
- informative statistics (daily number of new cases);
- scientific knowledge (virus incubation distribution);
- simpler model and approach can be more useful and robust for population science.

Discussion

Limitations: not accounting for differences between countries.

Current work:

- Modeling state-level data and compare response strategies.
- Evaluate effect of public health interventions (timing of imposing or lifting restrictions, mask wearing, school reopening).
- Account for covariates: differences in demographics of infected populations, testing capacity, state-level health risk measures, hospital resources, mobility data.

Discussion

How can statisticians help?

- Modeling disease transmission, study prognostic factors, design and analysis of treatment and vaccine trials.
- Societal responses and behavioral changes are major tools for preventing another outbreak without a vaccine.
- Behavioral changes can be shaped by public view towards risks.
- Communicating risks with the public
  - Low mortality rate? Consider intensity and excess mortality.
  - Only reporting the low absolute risks to certain individuals (e.g., younger adults) does not convey their risks to vulnerable populations.

THANK YOU !