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

DSI Data Challenge, April 2020

Team members:

Qinxia Wang, **Shanghong Xie**, **Yuanjia Wang** (Department of Biostatistics, Columbia University), **Donglin Zeng** (Department of Biostatistics, University of North Carolina)

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH
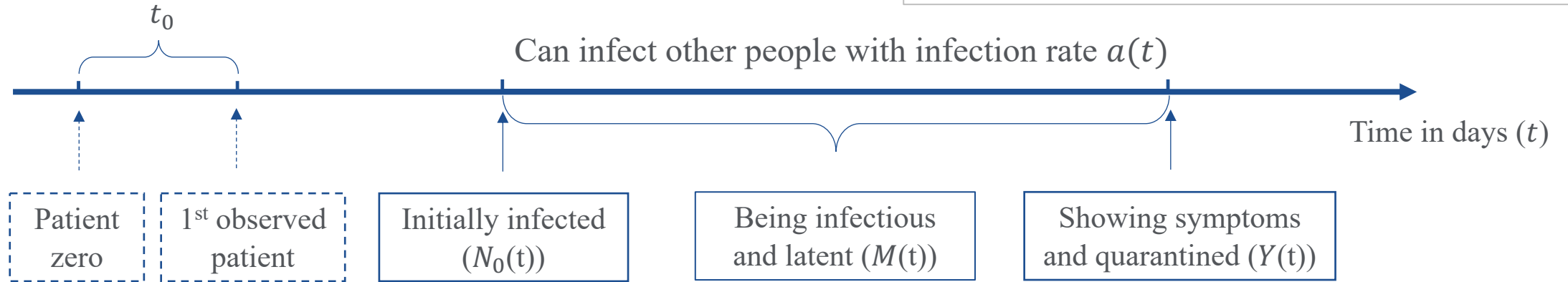
Introduction

- The coronavirus disease COVID-19 has created major health crisis around the world. It is imperative to predict the disease epidemic, investigate the impacts of containment and mitigation measures on infection rates, and compare between countries.

- Existing methods for infectious disease modeling are SEIR models that rely on many untestable prior assumptions (e.g., fitting past influenza data) and unreliable with wide prediction intervals.

- We propose a robust survival-convolution model with few parameters that incorporates the date of unknown patient zero, latent incubation periods, and time-varying reproduction numbers.

- Project goals

  - Build a survival-convolution model to predict daily new confirmed cases, latent cases, the epidemic end date and the final total number of cases.

  - Estimate and compare intervention effects across countries.

  - Predict the future trend for countries (e.g., US) amid the epidemic under different hypothetical intervention scenarios.

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH
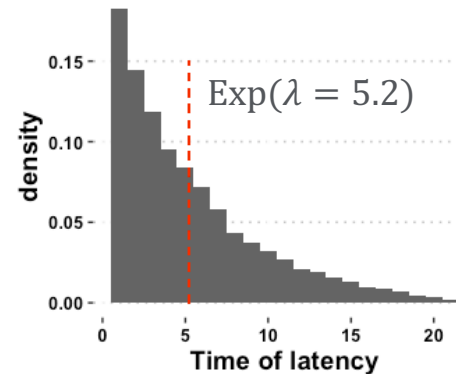
Survival-Convolution Model

$a(t)$ is non-negative, piecewise linear function with knot placing at the day of intervention.

$a(t)$ is related to the reproduction number as

$$R_t = \sum_{k=0}^{\infty} a(t+k)S(k).$$

t_0

Can infect other people with infection rate $a(t)$

Time in days (t)

Patient zero

1st observed patient

Initially infected $(N_0(t))$

Being infectious and latent $(M(t))$

Showing symptoms and quarantined $(Y(t))$

- $M(t) = \sum_{k=0}^{\infty} N_0(t-k)S(k)$
- $Y(t) = \sum_{k=0}^{\infty} N_0(t-k)[S(k) - S(k+1)]$
- $N_0(t+1) = a(t)[M(t) - Y(t)]$

Exp($\lambda = 5.2$)

density

0.15

0.10

0.05

0.00

0   5   10   15   20

Time of latency

- Loss function: $\sum_{t=0}^{\infty} \left(\sqrt{Y(t)} - \sqrt{\hat{Y}(t)} \right)^2$
- Optimizer: Adam (learning rate 0.01-0.02)
- Statistical inference: 500 permutations using standardized residuals.

Results (Italy)

- The curve of predicted daily new cases fits the test data well in the recent 13 days.
- The national lockdown on Mar 11 did not further accelerate the speed at which the infection rate decreases.
- Due to the incubation period, there is a large number of latent cases before being symptomatic.

Figure 1: Observed and predicted daily new cases and 95% prediction interval (shaded). First dashed line indicates the national lockdown (Mar 11). Second dashed line indicates two weeks after. Training data: Feb 15 to Apr 10; Testing data: Apr 11 to Apr 23.
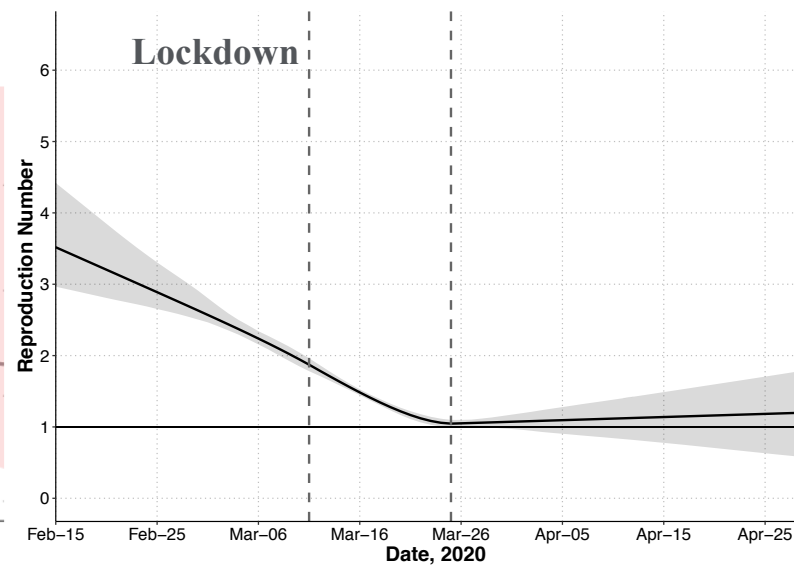
Figure 2: Reproduction number R_t as the average number of secondary infections generated by a primary case accounting for the incubation period.
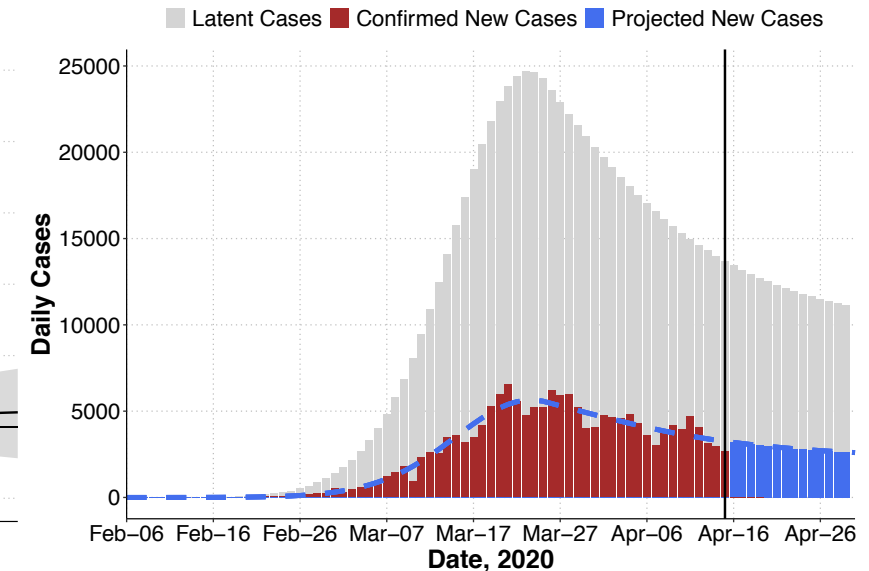
Figure 3: Latent and confirmed cases on each day. Number of latent cases includes all pre-symptomatic cases infected k days before that have not been detected yet.

Results (US)

- R_t declines significantly faster over a two-week period after the declaration of national emergency on Mar 13.
- After the peak (Apr 10), the rate of decrease in daily new cases has slowed down, which drives the current data close to scenario 4.
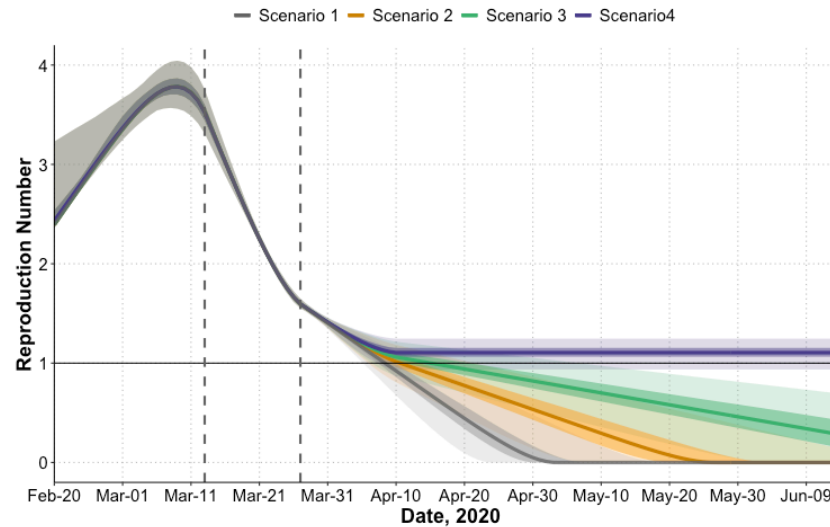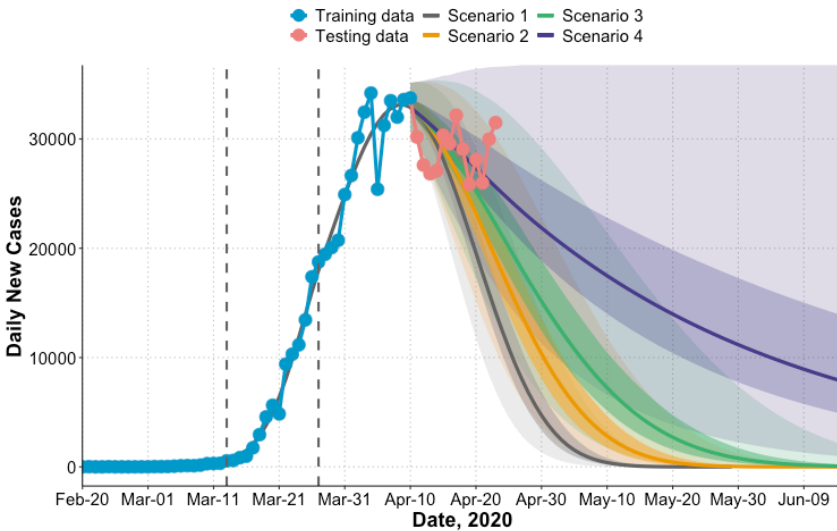

Figure 4: Observed and predicted daily new cases and 95% prediction interval (lighter shaded) and 50% prediction intervals (darker shaded) under four scenarios. Scenario 1: infection rate $a(t)$ follows the same trend after Apr 10 as observed between Mar 27 and Apr 10. Scenario 2 - 4: rate of decrease of $a(t)$ slows by 50%, 75% and 100% after Apr 10. First dashed line indicates the national lockdown (Mar 13). Second dashed line indicates two weeks after. Training data: Feb 20 to Apr 10; Testing data: Apr 11 to Apr 23.

Figure 5: Latent and confirmed cases on each day.

Conclusions and Discussion

- We propose a parsimonious and robust survival-convolution model to predict daily new cases and investigate the effects of mitigation strategies. The model can estimate the reproduction number R_t for transmission over time.

- The mitigation measures of US on Mar 13 is more effective for slowing down R_t compared with the mitigation of Italy on Mar 11. However, both countries have recent R_t remained above 1, suggesting that the epidemic is not yet under control. Continuing the current trend, the end date (when daily new cases decreases to 0) will not arrive before June in US.

- For China and South Korea, we used 2 weeks data to predict the entire epidemic (Appendix).

- We have extended the model to incorporate daily new deaths to estimate the underlying death rate and improve the prediction of cases (in progress).

- The method can be extended to state-level analyses and incorporate between-state heterogeneity and spatial information.

Appendix (China and South Korea)