COVID in New York City: A Model-Based Perspective

Niket Thakkar, Daniel J. Klein, Prashanth Selvaraj, and Mike Famulare*
Institution for Disease Modeling; *mfamulare@idmod.org

Results as of 2pm PST on March 14th, 2020

Correction on 3/17/2020: *In the original report, mortality estimates did not correctly incorporate the 3-week lag from infection to death. This has been corrected in the text.*

This information was shared with New York public health authorities. We are working with them and our modeling partners on additional reports to help inform decision-making.

Summary

The first case of COVID-19 in New York City was diagnosed on March 1st, 2020. During the following week, the number of cases diagnosed in the city increased to seven, while a cluster in Westchester grew rapidly to 28 cases on account of a superspreading event at a synagogue. At the time of writing, a total of 154 cases have been reported in the city, and 421 in the state as a whole.

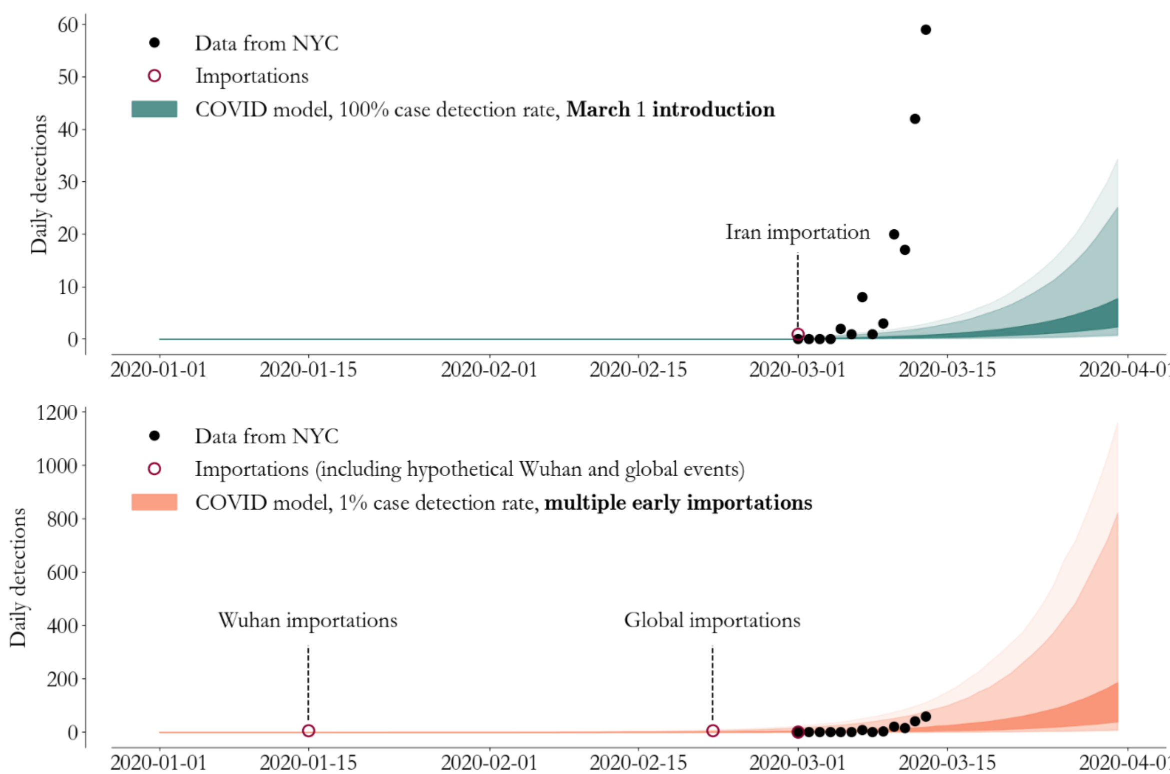

Figure 1. The rapid rise in NYC case counts is inconsistent with the assumption that COVID-19 was introduced on March 1st, even if all infections are being detected. Better alignment is achieved by assuming multiple, earlier importations and with a 1% case detection rate. (Dark region is the model's 50% interval, lighter is 95%, and lightest is 99%).

As with modeling we recently conducted for the Seattle area, we compared a COVID-19-specific model to the case count data in New York City. Under the assumptions of an epidemic doubling time of approximately six days, as consistently found in other geographies, we find that **multiple importations to the city likely happened significantly earlier than March 1st**. New York City's case detection data is more consistent with earlier importation dates, **similar to those observed in Seoul** or Seattle, cities which have **already introduced considerable policy measures** including school closures, social distancing, and restrictions on large gatherings.

We estimate that there are currently roughly 2,200 active infections (with 95% uncertainty interval 160 to 10,000) in New York City. While the probability of death for each COVID-19 infection varies with age and other co-morbidities, approximately 1.6% percent of infections result in death at an approximately 3-week delay from the infection date. With these assumptions in our model, the expected number of deaths in New York City attributable to COVID-19 as of today is 4 with a 95% confidence interval of 0 to 16. Currently, New York City has one death officially attributed to COVID-19.

We recommend immediate and significant policy action in New York City and the surrounding areas to mitigate significant adverse health and economic impacts.

Introduction

The novel coronavirus SARS-CoV-2 emerged in Wuhan, China, in late Nov or early Dec 2019. As of March 9th 2020, it is responsible for 109,577 confirmed cases and 3,809 deaths of the disease COVID-19 (WHO). After initial emergence in China, travel associated cases started to appear in other parts of the world with strong travel connections to Wuhan (http://rocs.hu-berlin.de/corona/). The first confirmed case in the US was a travel-associated case in Snohomish County, WA, screened on January 19th 2020. In the 6 weeks following to late February, a second presumptive case was identified roughly 10 miles away from where the first case was treated. As of the afternoon of March 10th, Washington State reports 267 confirmed cases and 24 confirmed deaths associated with COVID-19 with the majority from King and Snohomish counties.

The first case identified in the state of New York was a 39 year-old health care worker who had recently returned from Iran. COVID-19 appears to have spread rapidly from this initial diagnosis. It is possible, but unlikely, that COVID-19 is more infectious in NYC than elsewhere. Rather, **we believe that COVID-19 was introduced earlier, multiple times**, to the NYC community.

Further evidence of early initial importation comes from directly comparing New York City's case detection data to that of Seoul, South Korea. Seoul has experienced 15 importations, the first of which arrived on January 19th. As of the 9th of March, a total of 131 cases had been diagnosed there, a city of similar size to New York city.

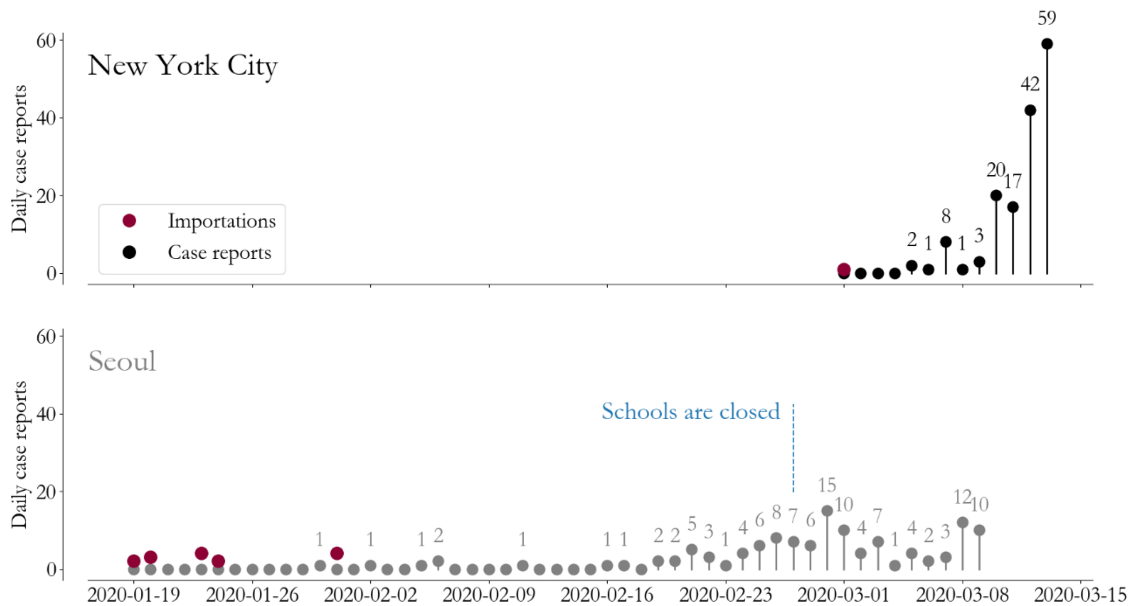

Figure 2. Multiple importations starting in mid-January introduced COVID to Seoul, and current daily case counts are lower than in NYC with significantly more testing, likely due to numerous policy interventions. (Estimated importation date, red; cases reported, black and grey.)

Genetic epidemiology is another tool to address questions of disease introduction and circulation. However, there is only one viral genome sequence available for NY. We cannot draw stronger conclusions without additional sequence data.

Key inputs and assumptions

Our model requires daily case counts as inputs and makes a number of standard assumptions. The geography we are considering is New York City. The key modeling assumption is that individuals can be grouped into one of four disease states: susceptible, exposed but non-infectious, infectious, and recovered. The infectivity of COVID-19, durations of various disease stages, and the approximate reporting rate can be assumed or estimated from time-series data. Importantly:

- COVID-19 has an asymptomatic period that lasts 4-5 days.
- Further delays are expected between symptom onset and diagnosis.
- We assume 1% of infections are diagnosed as cases due to the large number of asymptomatic and sub-clinical infections. This number is significantly higher than estimates for flu reporting in Seattle (0.1%), reflective of current urgency, but lower than estimates for COVID-19 in Seoul, reflective of the scale of testing in NYC.
- The time it takes the number of cumulative infectious to double is 6.2 days.

Transmission modeling

Our ability to model COVID-19 is evolving rapidly. For this analysis, we are using an extension of a published mathematical transmission model in order to rapidly explore and understand case count data as it relates to importation dates and disease parameters.

Healthcare system

In other work, we are beginning to forecast the impact of COVID-19 on the health system. A recurring theme is that **COVID-19 will quickly overwhelm health system capacity if left unchecked**. To extend these analyses to New York City or the state as a whole, we would seek specific data on the system overall, including the number of beds and size of the healthcare workforce.

Conclusions to date

COVID-19 is a serious and deadly virus that has spread rapidly worldwide and in a number of American cities. The main conclusions from this early finding are:

- NYC case counts appear incompatible with a March 1st introduction.
- Models are compatible with the rise in NYC case detections when multiple importations are introduced at timings consistent with other global epidemics
- This general hypothesis is supported by the data, but additional viral sequence data is required to identify specific importation pathways
- With a mature epidemic and cases doubling rapidly, there is little time to waste.
- NYC should immediately institute aggressive policies to avoid significant health and economic consequences.

Results today include model-based findings from a historical look at the case counts in the city, and comparison to other locations experiencing similar burden. We are an institute that specializes in modeling, and we are rapidly developing our tools to provide COVID-19 modeling and scenario analysis, but our capacity is limited. Please reach out through established contact points with any additional requests.

Appendix 1: Detailed transmission model methods

The transmission model is a discrete-time SEIR model based on the Lancet paper of Wu et al. The model can be described concisely by the following equations:

$$S_t = S_{t-1} - \beta S_{t-1}(I_{t-1} + z_{t-1})\varepsilon_t$$

$$E_t = \beta S_{t-1}(I_{t-1} + z_{t-1})\varepsilon_t + \left(1 - \frac{1}{D_e}\right) E_{t-1}$$

$$I_t = \frac{1}{D_e} E_{t-1} + \left(1 - \frac{1}{D_i}\right) I_{t-1}$$

$$C_t \sim \text{Binomial}(I_t, p)$$

where D_e is the latency duration (4 days), D_i is the infectious duration (8 days), β is the attack rate (based on the 6 day doubling time), and z_t is the number of importations on day t. In Figure 1's lower panel, z_t is non-zero on January 15th, February 22nd, and March 1st. The first two non-zero entries are set to 7, corresponding to clusters of importations associated with the initial outbreak in Wuhan and the lagged, global outbreaks that followed, while the final non-zero entry is the initial case found in NYC and linked to Iran. The total, 15 importations, is based on the 15 importations found in Seoul, S. Korea.

In the model, transmission is a log normal stochastic process, with $\text{Var}[\ln \varepsilon_t] = \sigma_\varepsilon^2$ while case detection is a binomial process with reporting rate p. In the results presented above, $\sigma_\varepsilon = 0.722$, a value determined by fitting daily case-detection time series from Seattle's 2018-19 influenza season.