Analyzing Trends of Covid Outcomes in Five NYC Boroughs

Fitting Richard Curve Model Using Newton-Raphson Algorithm

Tianshu Liu, Jiong Ma, Jiayi Shi, Zhengwei Song, Ziqing Wang

Outline

- COVID Prediction Background
- Richard’s Growth Model
- Newton-Raphson algorithm for model fitting
    - Method for choosing starting values
- Result
    - Fitted parameters
    - Predictions
        - Implications of the results
        - Potential vaccine distribution strategy
- Discussion and Conclusion

Background

The COVID-19 pandemic has affected many aspects of our lives and the distribution of COVID-19 vaccines is a critical step to control the pandemic. The public available data at all levels empowered COVID-19 research and helped policymakers make informed decisions.

Objective: **Fit pandemic curves for each NYC borough** and **make a distribution plan for COVID-19 vaccines** across the city based on the predicted trends.

Dataset: the citywide and borough-specific daily counts of COVID cases, hospitalizations, and deaths from February 29, 2020 to December 11, 2020 published by the New York Department of Health (NYDOH).

Richard growth function

$$N(t) = \frac{a}{\{1+d\exp\{-k(t-t_0)\}\}^{1/d}}$$

- Widely used for prediction of infectious disease transmission dynamics, including COVID-19
- S-shaped, extension of logistic function, doesn't need to be symmetric around the inflection point
- t: the time since the beginning of the pandemic
- a: carrying capacity, i.e. the maximum size or population that could be potentially infected
- k: the parameter that controls the growth rate at the inflection point
- d: the shape parameter
- t_0: the time at an inflection

A Newton-Raphson algorithm for curve fitting

$$Y_i = N\left(t_i, \boldsymbol{\theta}\right) + \epsilon_i$$

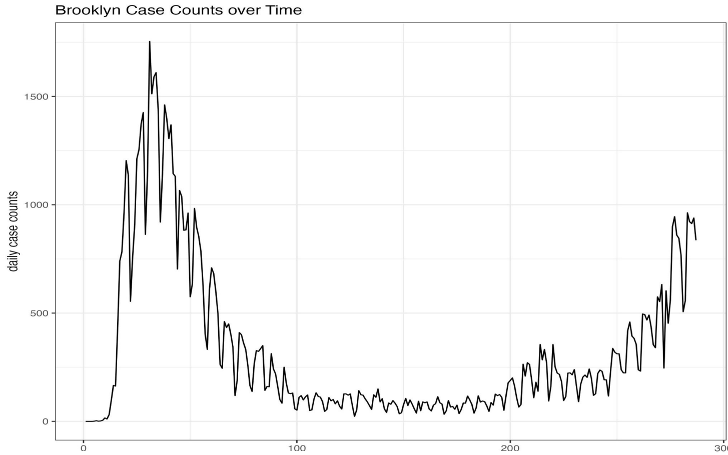
- Minimize $RSS = \sum_{i=1}^{n}(Y_i - N(t_i, \theta))^2$
- Gradient $\nabla f(\boldsymbol{\theta}) = \left(\frac{\partial f(\boldsymbol{\theta})}{\partial a}, \frac{\partial f(\boldsymbol{\theta})}{\partial k}, \frac{\partial f(\boldsymbol{\theta})}{\partial d}, \frac{\partial f(\boldsymbol{\theta})}{\partial t_0}\right)'$
- Hessian $\nabla^2 f(\boldsymbol{\theta})$ with i, j element equal to $\frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$
- i^th^ Iterative step $\boldsymbol{\theta_i} = \boldsymbol{\theta_{i-1}} - \lambda(\nabla^2 f(\boldsymbol{\theta_{i-1}}))^{-1} \nabla f(\boldsymbol{\theta_{i-1}})$

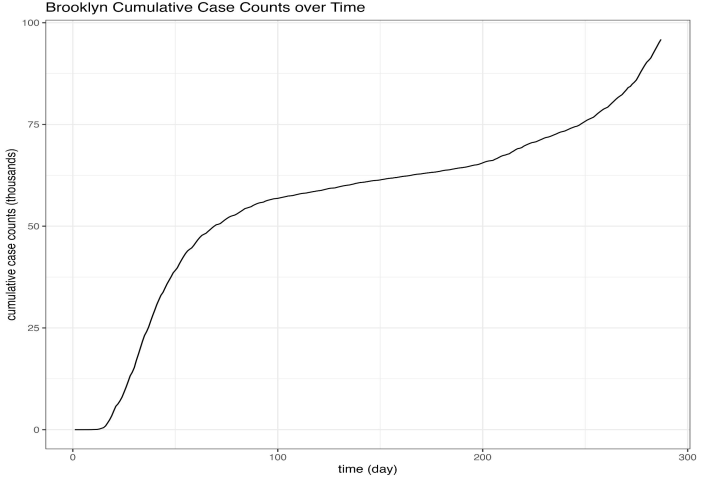
Newton Raphson Algorithm Steps

1. Initialization (i = 0)
    1.1. Objective: Find parameters that maximize -RSS
    1.2. Pass in initial parameter values to compute initial gradient and Hessian
    1.3. Initialize the starting RSS as Inf
2. Iteration steps (i = 1, 2, 3……)
    2.1. While $|RSS_{i-1} - RSS_i| > tol$
        2.1.1. Direction check
            2.1.1.1. If current step not in ascent direction, make the Hessian negative definite by subtracting (max eigenvalue + 1) from Hessian diagonal
        2.1.2. Update parameter values
        2.1.3. Bound check
            2.1.3.1. If any of the updated parameter value is negative, set to the value in previous step
        2.1.4. While $RSS_{i-1} - RSS_i < 0$
            2.1.4.1. Reduce the step size by a half and update the parameter values
            2.1.4.2. Check parameter bounds
3. Convergence criteria: $|RSS_{i-1} - RSS_i| \leq tol$
    3.1. Return the optimal parameter values when the convergence criteria is satisfied

Method for choosing starting values

- Divide into two pandemic waves and choose starting values separately:
  - 1st pandemic wave (w1) : t from 1 to 150
  - 2nd pandemic wave (w2) : t from 151 to 287
- a: chosen based on the maximum cumulative case/hospitalization/death count for each wave
  - # cumulative cases at t=150 for w1;
  - 2 × # cumulative cases at t=287 for w2 (middle point)
- d: a number between 0 and 1. (d=0 Gompertz growth curve [1]; d=1 logistic growth curve [2]). We chose d=0.8 for all scenarios.
- t_0: the time at which the number of daily counts reaches the highest
- k: estimated by fitting a linear regression model





Trend of daily case count (top) and cumulative case count (bottom) in Brooklyn

Choosing starting k using linear regression

$$m = N(t) = \frac{a}{(1+de^{-k(t-t_0)})}$$

$$\frac{a}{m} = 1 + de^{-k(t-t_0)}$$

$$\log(\frac{a}{m}) - 1 = \frac{1}{d}\log(de^{-k(t-t_0)})$$

$$\log((\frac{a}{m} - 1)^d) = \log(d) - k(t - t_0)$$

$$\log(\frac{(\frac{a}{m}-1)^d}{d}) = -k(t - t_0)$$

- Manipulate the Richard's curve into a linear relationship
- Plug in the starting values for a, d, t_0
- Plug in the cumulative case/hospitalization/death count data for m
- Perform linear regression to obtain the starting value for k

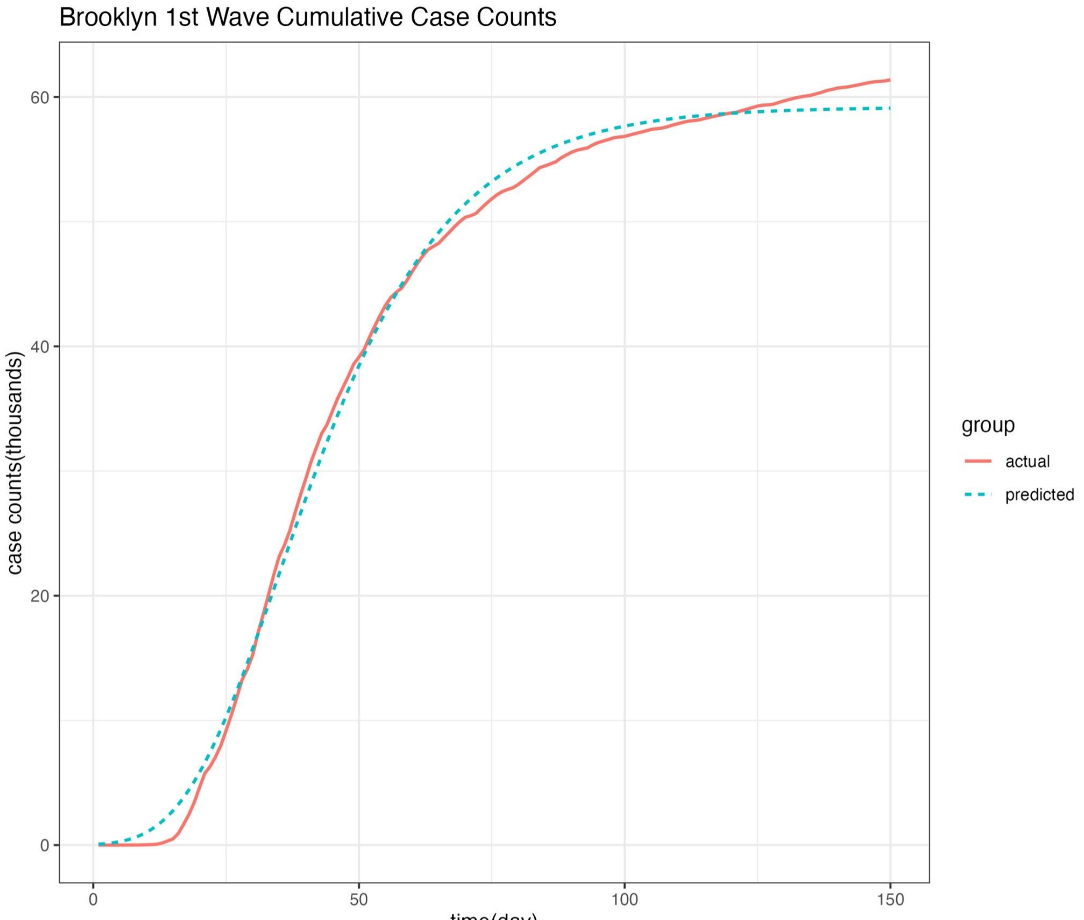
Results: 1st Wave Cumulative Cases in Brooklyn

- Starting value:

a	k	d	t_0
65	0.0432	0.8	40

- Fitted value:

a	k	d	t_0
59.1925	0.0560	4.1822e-11	35.0033



Results: 1st Wave Cumulative Cases by Borough

	Starting value				Result			
	a	k	d	t_0	a	k	d	t_0
BK	65	0.0432	0.8	40	59.1925	0.0560	4.1822e-11	35.0033
BX	70	0.0483	0.8	40	65.3846	0.0617	1.5317e-08	34.8640
MN	30	0.0450	0.8	40	27.0454	0.0511	1.5387e-12	33.7299
QN	70	0.0483	0.8	40	65.3846	0.0617	1.5317e-08	34.8640
SI	15	0.0444	0.8	40	13.9874	0.0723	1.9554e-11	32.9436

Rank of growth rate: Staten Island > Bronx ≈ Queens > Brooklyn > Manhattan

Results: 1st Wave Cumulative Hospitalizations

	Starting value				Result			
	a	k	d	t_0	a	k	d	t_0
BK	17	0.0407	0.8	40	15.7101	0.0618	1.5020e-11	33.5572
BX	15	0.0364	0.8	40	11.882	0.0665	3.201e-11	33.835
MN	10	0.0306	0.8	40	7.6964	0.0667	1.8989e-11	32.5439
QN	20	0.0349	0.8	40	16.3018	0.0882	2.1557e-11	31.8764
SI	5	0.0249	0.8	40	2.4750	0.0574	1.4090e-12	34.5511

Rank of growth rate: Queens > Manhattan > Bronx > Brooklyn > Staten Island

Results: 1st Wave Cumulative Deaths

	Starting value				Result			
	a	k	d	t_0	a	k	d	t_0
BK	6	0.0414	0.8	40	5.6007	0.0752	1.2862e-11	40.8587
BX	5	0.0317	0.8	40	3.911	0.069	2.232e-11	41.267
MN	4	0.0271	0.8	40	2.4564	0.0655	7.5658e-12	41.5707
QN	7	0.0395	0.8	40	5.7782	0.0767	2.2050e-11	41.0972
SI	2	0.0192	0.8	40	0.8855	0.0639	2.5125e-10	40.8286

Rank of growth rate: Queens > Brooklyn > Bronx > Manhattan > Staten Island

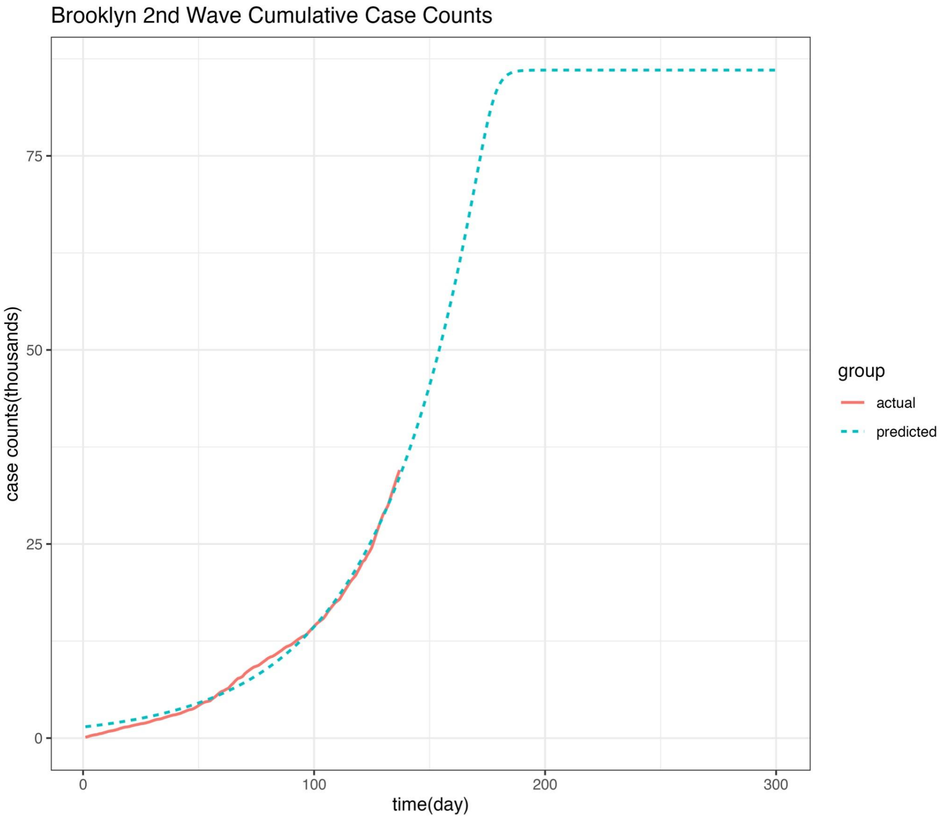
Results: 2nd Wave Cumulative Cases in Brooklyn

- Starting value:

a	k	d	t_0
80	0.2660	0.8	137

- Fitted value:

a	k	d	t_0
85.5067	0.3379	14.6456	169.4569



Results: 2nd Wave Cumulative Cases by Borough

	Starting value				Result			
	a	k	d	t_0	a	k	d	t_0
BK	80	0.0266	0.8	137	85.5067	0.3379	14.6456	169.4569
BX	42	0.0235	0.8	137	49.789	0.365	14.574	169.813
MN	42	0.0253	0.8	137	41.9403	0.4047	15.8552	163.5019
QN	65	0.0263	0.8	137	59.5907	0.5719	21.3935	159.2813
SI	30	0.0292	0.8	137	34.2817	0.381	11.0869	162.7768

Rank of growth rate: Queens > Manhattan > Staten Island > Bronx > Brooklyn

Results: 2nd Wave Cumulative Hospitalizations

	Starting value				Result			
	a	k	d	t_0	a	k	d	t_0
BK	10	0.0221	0.8	137	9.4513	0.0185	0.6664	169.543
BX	5	0.0205	0.8	137	6.736	0.166	8.615	198.097
MN	4	0.0216	0.8	137	2.5723	0.3801	17.2515	169.7831
QN	5	0.0242	0.8	137	9.1255	0.1956	8.7343	186.8791
SI	2	0.0247	0.8	137	1.9815	0.5729	16.915	156.6111

Rank of growth rate: Staten Island > Manhattan > Queens > Bronx > Brooklyn

Results: 2nd Wave Cumulative Deaths

	Starting value				Result			
	a	k	d	t_0	a	k	d	t_0
BK	1.2	0.0232	0.8	137	2.6286	0.0089	0.1412	238.2228
BX	0.5	0.0180	0.8	137	0.524	0.0081	2.695e-09	154.253
MN	0.4	0.0177	0.8	137	0.3741	0.2022	11.1460	189.5283
QN	0.6	0.0199	0.8	137	0.5203	0.3125	17.1367	173.8689
SI	0.3	0.0271	0.8	137	0.3938	0.4208	9.7594	160.3436

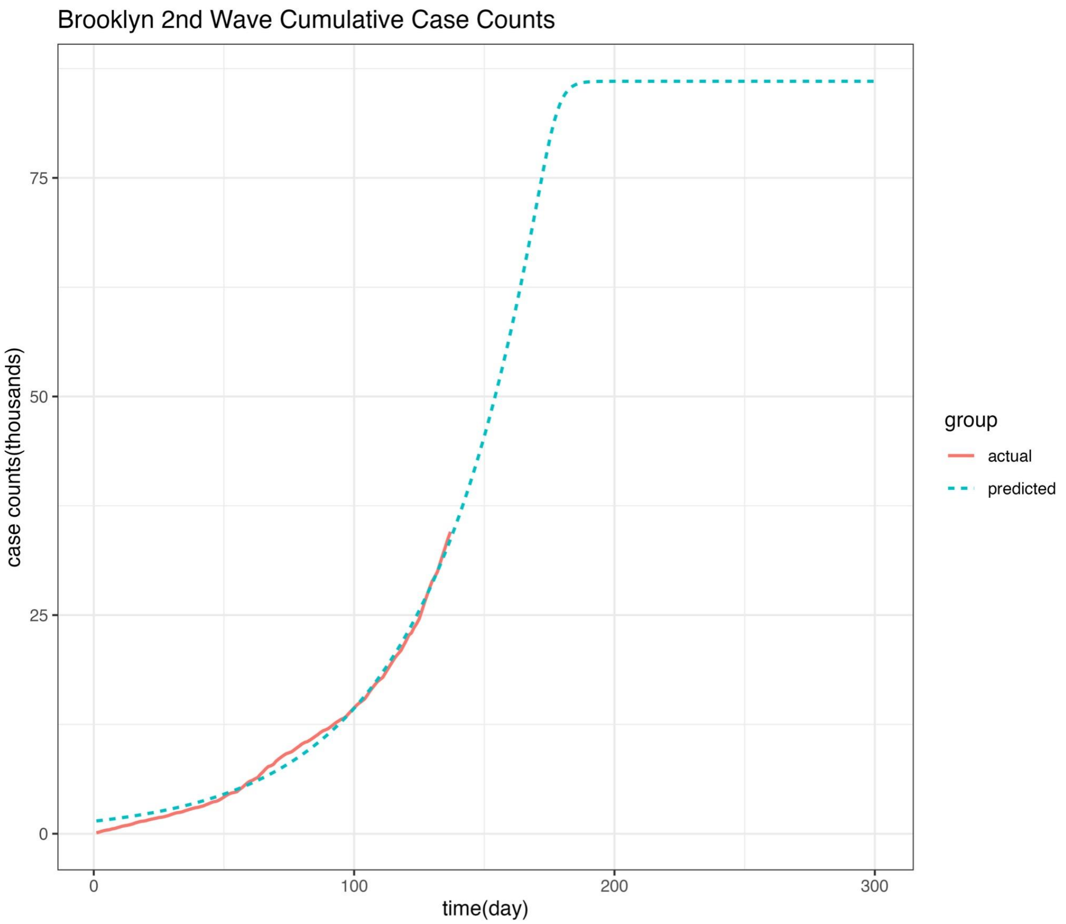
Rank of growth rate: Staten Island > Queens > Manhattan > Brooklyn > Bronx

Results: 1st Wave vs 2nd Wave growth rate

	Cases		Hospitalizations		Deaths	
	Wave 1	Wave 2	Wave 1	Wave 2	Wave 1	Wave 2
BK	0.0560	0.3379 ⬆️	0.0618	0.0185 ⬇️	0.0752	0.0089 ⬇️
BX	0.0617	0.365 ⬆️	0.0665	0.166 ⬆️	0.069	0.0081 ⬇️
MN	0.0511	0.4047 ⬆️	0.0667	0.3801 ⬆️	0.0655	0.2022 ⬆️
QN	0.0617	0.5719 ⬆️	0.0882	0.1956 ⬆️	0.0767	0.3125 ⬆️
SI	0.0723	0.381 ⬆️	0.0574	0.5729 ⬇️	0.0639	0.4208 ⬆️

Predictions and Suggestions

- Predict the number of cumulative cases, hospitalizations and deaths on December 18th, 2020 based on the fitted richard growth functions.



Predictions and Suggestions

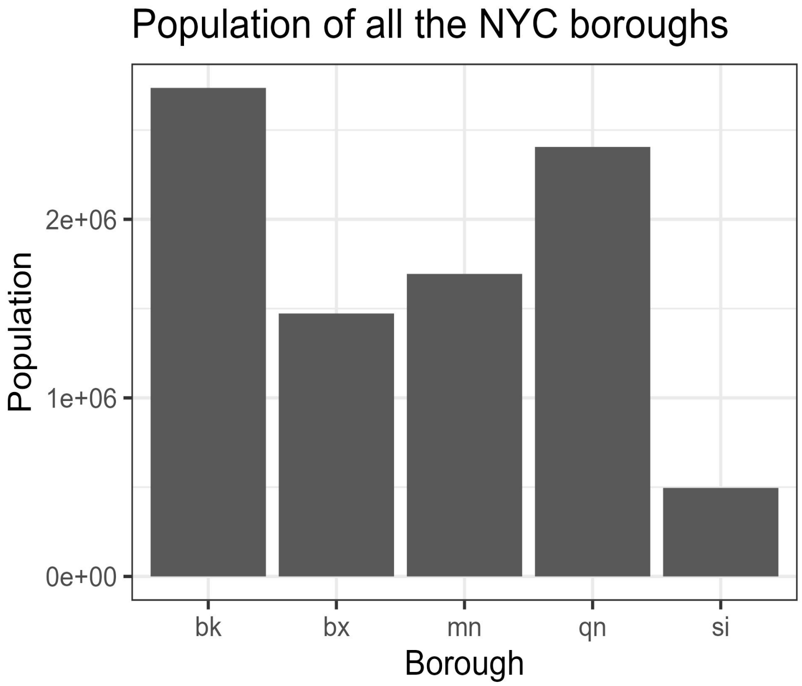
For each borough:

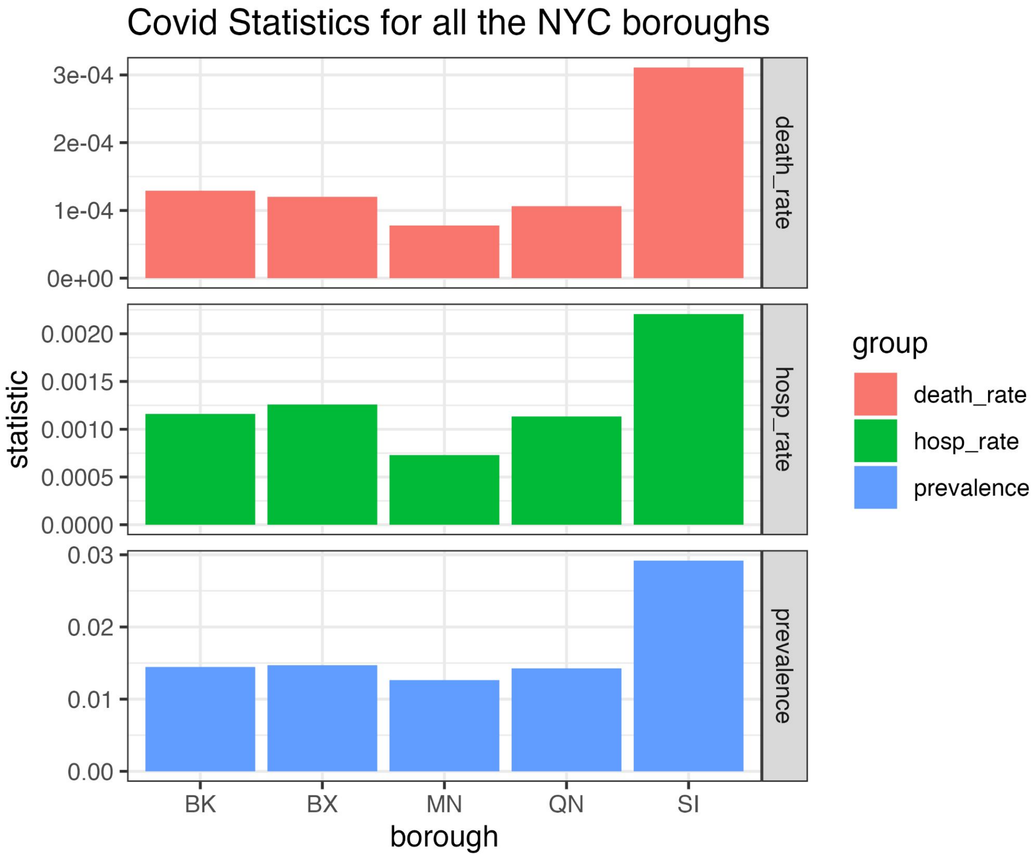
- Populations
- Prevalence: # cumulative cases / # populations
- Hospitalization rate: # cumulative hospitalizations / # cumulative cases
- Death rate: # cumulative deaths / # cumulative cases

Suggestions for vaccination roll-out plan based on different purposes:

1. Prioritize boroughs with highest prevalence: population * prevalence
2. Prioritize boroughs with highest hospitalization rate: population * hospitalization rate
3. Prioritize boroughs with highest death rate: population * death rate

Predictions and Suggestions





Predictions and Suggestions

Standardized statistics for distribution of vaccines

boro	prevalence	k rank	hospitalization	k rank	death	k rank
BK	30%	5	31%	5	33%	4
BX	16%	4	18%	4	17%	5
MN	16%	2	12%	2	12%	3
QN	26%	1	27%	3	24%	2
SI	11%	3	11%	1	14%	1

Discussion and Conclusion

Conclusion:

- Newton-Raphson algorithm helps to fit Richard’s curve for COVID data.
- The selection of starting value is crucial in fitting Richard’s curves. When data for all time point is available, the fitting result will be more accurate and the curve is more likely to be S-shaped.
- The 2nd wave shows a greater growth rate in case than the 1st wave among all 5 boroughs, implying the 2nd wave is more severe than the 1st one. Our fitting result for the 2nd wave can be used to predict the future count.
- The vaccine roll-out plan can be decided on both census data and the measure of interest. A combination of multiple rates is also a good choice to calculate the proportion of vaccine allocated to boroughs.

Discussion and Conclusion

Discussion:

- Model application
- Selection of starting value
  - Trying different starting d, or selecting d based on theory, may improve the model fit.
- The fitting result of 2nd wave is sensitive to the starting value
  - t_0 = 137, the count will be decreasing after that day
  - a is almost 2 times the count in Day 137
- Vaccine roll-out plan
  - A combination of measures of interest can be used
  - Socioeconomic factors: Old population (proportion, growth rate), Race, Income…

Reference

[1] Gompertz B. XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c. Philosophical transactions of the Royal Society of London. 1825;(115):513–583.

[2] Tsoularis A, Wallace J. Analysis of logistic growth models. Mathematical biosciences. 2002; 179(1):21– 55. https://doi.org/10.1016/S0025-5564(02)00096-2 PMID: 12047920

Q&A