Estimating the infection fatality risk of COVID-19 in New York City, March 1–May 16, 2020

Wan Yang,[1*] Sasikiran Kandula,[2] Mary Huynh,[3] Sharon K. Greene,[4] Gretchen Van Wye,[3] Wenhui Li,[3] Hiu Tai Chan,[3] Emily McGibbon,[4] Alice Yeung,[4] Donald Olson,[5] Anne Fine,[4] Jeffrey Shaman[2]

[1]Department of Epidemiology, Mailman School of Public Health, Columbia University; [2]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University; [3]Bureau of Vital Statistics, New York City Department of Health and Mental Hygiene; [4]Bureau of Communicable Disease, New York City Department of Health and Mental Hygiene; [5]Bureau of Equitable Health Systems, New York City Department of Health and Mental Hygiene.

*correspondence to: wy2202@cumc.columbia.edu (WY)

Abstract

During March 1-May 16, 2020, 191,392 laboratory-confirmed COVID-19 cases were diagnosed and reported and 20,141 confirmed and probable COVID-19 deaths occurred among New York City (NYC) residents. We applied a network model-inference system developed to support the City's pandemic response to estimate underlying SARS-CoV-2 infection rates. Based on these estimates, we further estimated the infection fatality risk (IFR) for 5 age groups (i.e. <25, 25-44, 45-64, 65-74, and 75+ years) and all ages overall, during March 1–May 16, 2020. We estimated an overall IFR of 1.45% (95% Credible Interval: 1.09-1.87%) in NYC. In particular, weekly IFR was estimated as high as 6.1% for 65-74 year-olds and 17.0% for 75+ year-olds. These results are based on more complete ascertainment of COVID-19-related deaths in NYC and thus likely more accurately reflect the true, higher burden of death due to COVID-19 than previously reported elsewhere. It is thus crucial that officials account for and closely monitor the infection rate and population health outcomes and enact prompt public health responses accordingly as the pandemic unfolds.

Key words: COVID-19; infection fatality risk; age-specific; reporting rate; New York City

Introduction

The novel coronavirus SARS-CoV-2 emerged in late 2019 in China and subsequently spread to 200+ other countries. As of June 26, 2020, there were over 9.47 million reported COVID-19

cases and over 484.2 thousand deaths worldwide.[1] As the pandemic continues to unfold and populations in many places worldwide largely remain susceptible, understanding the severity, in particular, the infection fatality risk (IFR), is crucial for gauging the full impact of COVID-19 in the coming months or years. However, estimating the IFR of COVID-19 is challenging due to the large number of undocumented infections, fluctuating case detection rates, and inconsistent reporting of fatalities. Further, the IFR of COVID-19 could vary by location, given differences in demographics, healthcare systems, and social construct (e.g., intergenerational households are the norm in some societies whereas older adults commonly reside and congregate in long-term care and adult care facilities in others). Most IFR estimates thus far have come from data recorded in China, the Diamond Princess cruise ship, and France.[2-5] Yet the IFR in the United States—the country currently reporting the largest number of cases—remains unclear.

New York City (NYC) reported its first case on March 1, 2020, in a traveler, and quickly became the epicenter in the United States. By May 16, 2020, there were 191,392 diagnosed cases and 20,131 deaths reported in NYC (Table 1). During the pandemic, the NYC Department of Health and Mental Hygiene (DOHMH) and the Mailman School of Public Health at Columbia University have been collaborating in generating real-time model projections in support of the City's pandemic response. Our latest model-inference system uses a network model to simulate SARS-CoV-2 transmission in the City's 42 United Hospital Fund neighborhoods.[6] The model is run in conjunction with the Ensemble Adjustment Kalman Filter (EAKF)[7] and fit simultaneously to case and mortality data for each of the 42 neighborhoods while accounting for under-detection, delay from infection to case reporting and death, and changing interventions (e.g., social distancing). In this study, we apply this network model-inference system to estimate the IFR for 5 age groups (i.e. <25, 25-44, 45-64, 65-74, and 75+ years) and all ages overall, from March 1 to May 16, 2020. In the process, we also estimate reporting rates—i.e. the fraction of infections documented as confirmed cases—and the cumulative infection rate by May 16, 2020.

Methods

Data

Laboratory-confirmed COVID-19 cases reported to the NYC DOHMH were aggregated by week of diagnosis and age group (<1, 1-4, 5-14, 15-24, 25-44, 45-64, 65-74, and 75+ years) for each of the 42 United Hospital Fund neighborhoods[6] in NYC, according to the patient's residential address at time of report. The mortality data, from deaths registered and analyzed by the NYC DOHMH, combined confirmed and probable COVID-19-associated deaths. Confirmed COVID-19-associated deaths were defined as those occurring in persons with laboratory-confirmed SARS-CoV-2 infection; and probable COVID-19 deaths were defined as those with COVID-19, SARS-CoV-2, or a similar term listed on the death certificate as an immediate, underlying, or contributing cause of death but did not have laboratory-confirmation of COVID-19.[8] Due to privacy concerns, mortality data were aggregated to 5 coarser age groups (<18, 18-44, 45-64, 65-74, and 75+ years) for each neighborhood by week of death. To match with the age grouping for case data, we used the citywide fraction of deaths occurring in each of the five finer age groups (i.e. <1, 1-4, 5-14, 15-24, 25-44) to apportion deaths in the <18 and 18-44 year age categories. For this study, case and mortality data were both retrieved on May 22, 2020.

The mobility data, used to model changes in COVID-19 transmission rate due to public health interventions implemented during the pandemic (e.g., social distancing), came from SafeGraph[9, 10] and contained counts of visitors to locations in each zip code based on mobile device locations. The released data were anonymized and aggregated in weekly intervals. We spatially aggregated these data to the neighborhood level.

This study was classified as public health surveillance and exempt from ethical review and informed consent by the Institutional Review Boards of both Columbia University and NYC DOHMH.

Network transmission model

The network model simulated intra- and inter neighborhood transmission of COVID-19 and assumed susceptible-exposed-infectious-removed (SEIR) dynamics, per the following equations:

87
$$\begin{cases} \frac{dS_i}{dt} = -S_i \sum_{j=1}^{j=42} b_j \beta_{city} c_{ij} I_j / N_j \\ \frac{dE_i}{dt} = S_i \sum_{j=1}^{j=42} b_j \beta_{city} c_{ij} I_j / N_j - \frac{E_i}{Z} \\ \frac{dI_i}{dt} = \frac{E_i}{Z} - \frac{I_i}{D} \\ \frac{dR_i}{dt} = \frac{I_i}{D} \end{cases}$$

88

89 where S_i, E_i, I_i, R_i, and N_i are the numbers of susceptible, exposed (but not yet infectious),
90 infectious, and removed (either recovered or deceased) individuals and the total population,
91 respectively, from a given age group (described below) in neighborhood *i*. β_{city} is the citywide
92 transmission rate, which incorporated seasonal variation as observed for OC43, a beta-
93 coronavirus in humans from the same genus as SARS-CoV-2.[11] To allow differential transmission
94 in each neighborhood, we included a multiplicative factor, b_i, to scale neighborhood local
95 transmission rates. *Z* and *D* are the latency and infectious periods, respectively (Table S1).
96
97 The matrix $[c_{ij}]$ represents changes in contact rates over time and connectivity among
98 neighborhoods and was computed based on mobility data. Briefly, changes in contact rates
99 (either intra or inter neighborhoods) for week-*t* were computed as a ratio of the number of
100 visitors during week-*t* to that during the week of March 1, 2020 (the first week of the pandemic
101 in NYC when there were no interventions in place), and further scaled by a multiplicative factor
102 m_1; m_1 was estimated along with other parameters. To compute the connectivity among the
103 neighborhoods, we first divided the inter-neighborhood mobility by the local mobility (this gave
104 a relative measure of connectivity; e.g., if two neighborhoods are highly connected with lots of
105 individuals traveling between them, inter-neighborhood mobility would be closer to 1 and
106 much lower than 1 otherwise); we then scaled these relative rates by a multiplicative factor m_2,
107 which was also estimated along with other parameters.
108
109 Observation model

To account for delays in diagnosis and reporting, we included a time-from-infectious-to-case-reporting (i.e., diagnosis) lag, drawn from a gamma distribution with a mean of T_m and standard deviation (SD) of T_{sd} days. To account for under-detection, we included a case reporting rate (r), i.e. the fraction of infections (including subclinical or asymptomatic infections) reported as cases. To compute the model-simulated number of new cases per week, we multiplied the model-simulated number of infections per day (including those from the previous weeks) by the reporting rate, and further distributed these simulated cases in time per the distribution of time-from-infectious-to-case-reporting. We then aggregated the daily lagged, reported cases to weekly totals for model inference (see below). Similarly, to compute the model-simulated deaths per week, we multiplied the simulated-infections by the IFR and then distributed these simulated deaths in time per the distribution of time-from-infectious-to-death, and aggregated these daily numbers to weekly totals. For each week, the reporting rate (r), the mean (T_m) and standard deviation (T_{sd}) of time-from-infectious-case-reporting, and the IFR were estimated based on weekly case and mortality data. The distribution of time-from-diagnosis-to-death was based on observations of n=15,686 COVID-19 confirmed deaths in NYC (gamma distribution with mean = 9.36 days and SD = 9.76 days; Table S1).

Parameter estimation

To estimate model parameters (b_i, β_{city}, Z, D, m_1, m_2, T_m, T_{sd}, r, and IFR, for i=1,…,42) and state variables (S_i, E_i, and I_i, for i=1,…,42) for each week, we ran the network-model stochastically with a daily time step in conjunction with the EAKF and fit to weekly case and mortality data from the week starting March 1 to the week ending May 16, 2020. The EAKF uses an ensemble of model realizations (n=500 here), each with initial parameters and variables randomly drawn from a prior range (see Table S1). After model initialization, the model ensemble was integrated forward in time for a week to compute the model-simulated number of cases and deaths for that week; these prior estimates were then combined with the observed cases and deaths for the same week to compute the posterior per Bayes' theorem.[7] The posterior distribution of each model parameter/variable was updated for that week at the same time.[7] This parameter estimation process was done separately for each of the eight age groups (i.e. <1, 1-4, …, and

75+). To account for stochasticity in model initiation, we ran the parameter estimation process independently 10 times. Results for each age group were combined from these 10 runs (each with 500 realizations). To combine estimates of reporting rate and IFR for <25 year-olds or all ages overall, we weighted the age-group specific estimates by the fraction of estimated infections from each related group.

Results

The model-inference system was able to recreate the case and mortality time series for each age group and all ages overall (Fig. 1). For most age groups, confirmed cases peaked during the week of March 29 and the mortality rate peaked about one week later than the case rate, due to the time-lag from severe infection to death (Fig. 1).

There were, however, substantial under-detection of infections, variations by age group, and fluctuations of case reporting rates over time, in part due to changing testing criteria (e.g., testing was restricted to severely ill patients in the early phase due to material shortages in testing equipment and personal protective equipment). The estimated reporting rate for all ages overall started at a low level of 2.3% [median; 95% credible interval (CrI): 0.4–4.5%; same below] in the week of March 1; it increased to 21.4% (95% CrI: 14.4–31.3%) in the week of March 15 and stayed at similar levels afterwards (Fig 2F). The estimated reporting rate was highest for the two oldest age groups and substantially lower for younger age groups (Fig 2 D and E vs. Fig 2 A-C). During the last week of this study (i.e., May 10, 2020), we estimated that 25.9% (95% CrI: 15.1–41.8%) of infections among 65-74 year-olds and 34.2% (95% CrI: 23.0–49.9%) among 75+ year-olds were reported; in comparison, only 10.6% (95% CrI: 7.4–17.7%) of infections among <25 year-olds and 16.9% (95% CrI: 13.1–26.8%) among 25-44 year-olds were reported.

After accounting for the case reporting rate, the epidemic peak for new infections occurred one week sooner during the week of March 22, 2020 for <45 year-olds and all ages combined (Fig 2 A-B and F). This was coincident with the timing of public health interventions in

NYC – public schools in NYC were closed on March 16, 2020 and a citywide stay-at-home order was imposed starting the week of March 22, 2020.[12] Tallied over the entire study period, the estimated overall cumulative infection rate was 15.8% (95% CrI: 11.4–24.4%) by May 16, 2020. However, the estimated cumulative infection rate varied substantially by age group. Specifically, 25-44 and 45-64 year-olds had the highest cumulative infection rates, at 20.1% (95% CrI: 14.4–30.0%) and 20.7% (95% CrI: 15.7–28.0%), respectively; 65-74 and 75+ year-olds had the second highest cumulative infection rates, at 14.9% (95% CrI: 11.2–22.8%) and 12.9% (95% CrI: 9.9–19.9%); and <25 year-olds had the lowest cumulative infection rate (8.0%; 95% CrI: 5.1–16.7%). Of note, these estimates, albeit with large uncertainties, are in line with reported measures from serology surveys (e.g., 19.9% positive in NYC, as of May 1, 2020, likely from testing of 25-64 year-olds[13, 14]). In addition, the spatial variation estimated by our model-inference system[15] was in line with reported measures (i.e., highest in the Bronx and lowest in Manhattan[13, 14]). This consistency with independent serology survey data provides some independent validation of our model estimates.

During March 1 - May 16, 2020, a total of 20,141 COVID-19 deaths (15,723 confirmed and 4,418 probable) and 191,392 COVID-confirmed cases were reported in NYC. The crude confirmed case fatality risk was thus 8.22%. After accounting for changing case reporting rates and excluding the first three weeks (i.e., March 1-21, 2020) with zero or few reported deaths for which model estimates were less accurate, we estimate that the overall IFR, including both confirmed and probable deaths, was 1.45% (95% CrI: 1.09–1.87%) during March 22 – May 16, 2020.

Examining estimates by age group, estimated IFR was lowest in young age groups. The average IFR was 0.011% (95% CrI: 0.005–0.016%) for <25 year-olds, increased by ~10 fold to 0.12% (95% CrI: 0.077–0.15%) for 25-44 year-olds, and by another 7 fold to 0.94% (95% CrI: 0.74–1.21%) for 45-64 year-olds (Fig. 3 A-C). These estimates were similar to IFRs reported for China for corresponding age groups.[3] However, the estimated IFR for the two oldest age groups was much higher than the younger age groups and about twice as high as rates reported

for these age groups in China.[3, 4] The average IFR was 4.67% (95% CrI: 3.21–6.66%) for 65-74 year-olds and 13.83% (95% CrI: 9.65–17.78%) for 75+ year-olds. In addition, the estimated IFR fluctuated substantially over time for these two elderly groups. For 65-74 year-olds, estimated IFR was 6.10% (95% CrI: 4.90–7.57%) during the week of April 5, 2020 but decreased to 3.79% (95% CrI: 1.68–6.90%) during the week of May 10, 2020 (Fig 3D). For 75+ year-olds, IFR was estimated to be 16.99% (95% CrI: 13.15–20.11%) during the week of April 5, 2020 but decreased to 9.77% (95% CrI: 4.53–14.81%) during the week of May 10, 2020 (Fig 3F).

Discussion

In light of the large uncertainties in IFRs for COVID-19 due to under-ascertainment of cases, we have used a model-inference system, developed to support the pandemic response in NYC, to estimate local IFRs. During March 1–May 16, 2020, NYC recorded the largest numbers of COVID-19 cases and deaths in the US and perhaps worldwide. Despite public health efforts to slow the pandemic, e.g. by social distancing, and to ramp up healthcare capacity, over 20,000 lives were lost from COVID-19 in a short span of two months. Based on this large number of deaths, the estimated overall IFR was 1.45% in NYC. This estimate included both confirmed and probable COVID deaths. If only COVID confirmed deaths were included, given that 78.0% of deaths among the total were laboratory-confirmed, the estimated overall IFR would be around 1.1%. Both estimates were higher than previously reported for elsewhere (e.g., about 0.7% in both China[3] and France[5]). Importantly, NYC has nosologists who review all death certificates and record deaths into a unified electronic reporting system rapidly. This mortality surveillance infrastructure and enhanced nosology thus allow more rapid and complete death reporting in NYC. As such, our estimates here likely more accurately reflect the underlying fatality risk of COVID-19 infection. Further, given the likely stronger public health infrastructure and healthcare systems in NYC than many other places,[16] the higher IFR estimated here suggests that mortality risk from COVID-19 may be higher in the United States and likely other countries as well than previously reported. Of note, despite the large surge in cases and hospitalizations, through quick expansion of healthcare systems, most hospitals in NYC were able to meet patient care demand during the two-month period. As many jurisdictions in the United States

226 are considering re-opening after months of social distancing, it is crucial that officials account
227 for and closely monitor the infection rate and health outcomes including hospitalizations and
228 mortality and take prompt public health responses accordingly.
229
230 While the IFR estimated here was similar to those previously reported elsewhere for
231 younger age groups, we found that IFRs for individuals 65 years and older in NYC were about
232 twice as high as prior reports.[3] These higher IFRs may be in part due to differences in
233 population characteristics, in particular, the prevalence of underlying medical conditions such
234 as diabetes mellitus, chronic lung disease, and cardiovascular disease.[17, 18] Regardless,
235 estimated weekly IFR was as high as 6.1% for 65-74 year-olds and 17.0% for 75+ year-olds.
236 These dire estimates highlight the severity of COVID-19 in elderly populations and the
237 importance of infection prevention in congregate settings. Thus, early detection and adherence
238 to infection control guidance in long-term care and adult care facilities should be a priority for
239 COVID-19 response as the pandemic continues to unfold.
240
241 In this study, we incorporated multiple data sources, including age-grouped, spatially
242 resolved case and mortality data as well as mobility data, to calibrate our model-inference
243 system. Of note, the timing of the COVID-19 pandemic varied substantially among NYC
244 neighborhoods. For instance, peak mortality rates occurred up to 4 weeks apart among the 42
245 neighborhoods. Fitting the model-inference system simultaneously to these diverse case and
246 mortality time series thus enabled better constraint of key model parameters (e.g., case
247 reporting rate and IFR). However, we note there remain large uncertainties in model estimates.
248 A full assessment of COVID-19 severity will require comprehensive serology surveys of the
249 population by age group and neighborhood, given the large heterogeneity of infection rates
250 across population segments and space. In addition, we only included deaths that were lab-
251 confirmed or explicitly coded as related to COVID-19. A recent study reported that excess
252 deaths in NYC during about the same period could be more than 24,000.[8] Further, recent
253 studies have reported severe sequelae of COVID-19 in children, i.e. Multi-system Inflammatory

Syndrome in Children. Thus, it is important to monitor health outcomes in younger age groups post-infection as the pandemic unfolds, despite the low IFRs noted to date.

Acknowledgments:

This study was supported by the National Institute of Allergy and Infectious Diseases (AI145883), the National Science Foundation Rapid Response Research Program (RAPID; 2027369), and the NYC DOHMH. We thank the NYC DOHMH Bureau of Vital Statistics team and staff members in the NYC DOHMH Incident Command System Surveillance and Epidemiology Section for data management. We thank Miranda S. Moore at NYC DOHMH, who wrote and maintained the code to generate weekly extracts of COVID-19 case data for provision to Columbia University under a data use agreement. We thank Jaimie Shaff at NYC DOHMH and Helen Alesbury and Kate Klein at the Office of Chief Medical Examiner of the City of New York for initiating and coordinating discussions on modeling COVID-19 mortality in NYC. We also thank Columbia University Mailman School of Public Health for high performance computing and Safe Graph (safegraph.com) for providing the mobility data used in this study.

Conflict of Interest:

JS and Columbia University disclose partial ownership of SK Analytics. JS discloses consulting for BNI. Other authors declare no conflict of interest.

References:

1. World Health Organization. Coronavirus disease (COVID-2019) situation reports 2020 [updated 6/26/2020; cited 2020 6/26]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/.
2. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, Cowling BJ, Lipsitch M, Leung GM. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nat Med. 2020;26(4):506-10. PubMed PMID: 32284616; PMCID: PMC7094929.

3. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, Cuomo-Dannenburg G, Thompson H, Walker PGT, Fu H, Dighe A, Griffin JT, Baguelin M, Bhatia S, Boonyasiri A, Cori A, Cucunuba Z, FitzJohn R, Gaythorpe K, Green W, Hamlet A, Hinsley W, Laydon D, Nedjati-Gilani G, Riley S, van Elsland S, Volz E, Wang H, Wang Y, Xi X, Donnelly CA, Ghani AC, Ferguson NM. Estimates of the severity of coronavirus disease 2019: A model-based analysis. The Lancet Infectious diseases. 2020. PubMed PMID: 32240634; PMCID: PMC7158570.

4. Russell TW, Hellewell J, Jarvis CI, van Zandvoort K, Abbott S, Ratnayake R, group CC-w, Flasche S, Eggo RM, Edmunds WJ, Kucharski AJ. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the diamond princess cruise ship, february 2020. Eurosurveillance. 2020;25(12):2000256.

5. Salje H, Tran Kiem C, Lefrancq N, Courtejoie N, Bosetti P, Paireau J, Andronico A, Hoze N, Richet J, Dubost CL, Le Strat Y, Lessler J, Levy-Bruhl D, Fontanet A, Opatowski L, Boelle PY, Cauchemez S. Estimating the burden of SARS-cov-2 in France. Science. 2020. PubMed PMID: 32404476.

6. NewYork City Department of Health and Mental Hygiene. NYC uhf 42 neighborhoods. Available from: http://a816-dohbesp.nyc.gov/IndicatorPublic/EPHTPDF/uhf42.pdf.

7. Anderson JL. An ensemble adjustment Kalman filter for data assimilation. Mon Weather Rev. 2001;129(12):2884-903.

8. New York City Department of Health and Mental Hygiene (DOHMH) COVID-19 Response Team. Preliminary estimate of excess mortality during the COVID-19 outbreak — New York City, march 11–may 2, 2020. MMWR Morbidity and mortality weekly report. 2020;69:603-5.

9. SafeGraph. Weekly patterns: Foot traffic data to understand the COVID-19 pandemic 2020. Available from: https://www.safegraph.com/weekly-foot-traffic-patterns.

10. Lasry A, Kidder D, Hast M, Poovey J, Sunshine G, Winglee K, Zviedrite N, Ahmed F, Ethier KA, Program CDCPHL, New York City Department of H, Mental H, Louisiana Department of H, Public Health S, King C, San Francisco C-RT, Alameda County Public Health D, San Mateo County Health D, Marin County Division of Public H. Timing of community mitigation and

changes in reported COVID-19 and community mobility - four U.S. Metropolitan areas, february 26-april 1, 2020. MMWR Morbidity and mortality weekly report. 2020;69(15):451-7. PubMed PMID: 32298245.

11. Yang W, Kandula S, Shaman J. Eight-week model projections of COVID-19 in New York City 2020 [updated March 16, 2020]. Available from: https://github.com/wan-yang/COLUMBIA-COVID19-PROJECTIONS-FOR-NYC/blob/master/summary_nyc.projection200316_CU_updated.pdf.
12. New York State Department of Health. New York state on pause 2020 [cited 2020 5/16/2020]. Available from: https://coronavirus.health.ny.gov/new-york-state-pause.
13. www.governor.ny.gov. Amid ongoing COVID-19 pandemic, governor cuomo announces results of completed antibody testing study of 15,000 people showing 12.3 percent of population has COVID-19 antibodies 2020 [updated MAY 2, 2020; cited 2020 5/12/2020]. Available from: https://www.governor.ny.gov/news/amid-ongoing-covid-19-pandemic-governor-cuomo-announces-results-completed-antibody-testing.
14. Goodman JD, Rothfeld M. 1 in 5 new yorkers may have had covid-19, antibody tests suggest 2020 [updated 4/23/2020; cited 2020 5/12/2020]. Available from: https://www.nytimes.com/2020/04/23/nyregion/coronavirus-antibodies-test-ny.html.
15. Yang W, Kandula S, Shaman J. Projections of covid19 epidemic outcomes and healthcare demands for New York City (NYC) 2020 [cited 2020 5/12/2020]. Available from: https://github.com/wan-yang/COLUMBIA-COVID19-PROJECTIONS-FOR-NYC.
16. Medbelle. Best hospital cities ranking 2019 2019. Available from: https://www.medbelle.com/best-hospital-cities-usa.
17. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang HH, Mercer SW, Sanderson C, McKee M, Troeger C, Ong KI, Checchi F, Perel P, Joseph S, Gibbs HP, Banerjee A, Eggo RM. How many are at increased risk of severe COVID-19 disease? Rapid global, regional and national estimates for 2020. medRxiv. 2020:2020.04.18.20064774.
18. CDC COVID-19 Response Team. Preliminary estimates of the prevalence of selected underlying health conditions among patients with coronavirus disease 2019—United

States, february 12–march 28, 2020. Morbidity and Mortality Weekly Report. 2020;69(13):382.

Table 1. Summary estimates. Cases and deaths were reported during March 1 – May 16, 2020. Crude case fatality risk (CFR) was computed as the proportion of persons with confirmed COVID-19 illness who died. Cumulative infection rates, median (95% CrI), show percentages of population, for each age group or all ages overall, estimated to have been infected by May 16, 2020. IFR, median (95% CrI), was estimated here, and averaged over March 22 - May 16, 2020; we excluded estimates during March 1-21, 2020, because estimates were less accurate for these earliest weeks when zero or few deaths were reported.

Age	Confirmed Cases	Confirmed and Probable Deaths	Estimated cumulative infection rate (%)	Estimated IFR (%)
<25	14692	40	8 (5.1, 16.7)	0.011 (0.005, 0.016)
25-44	60474	688	20.1 (14.4, 30)	0.12 (0.077, 0.15)
45-64	69839	4457	20.7 (15.7, 28)	0.94 (0.74, 1.21)
65-74	23875	4866	14.9 (11.2, 22.8)	4.67 (3.22, 6.66)
75+	22512	10090	12.9 (9.9, 19.9)	13.83 (9.65, 17.78)
all	191392	20141	15.8 (11.4, 24.4)	1.45 (1.09, 1.87)

Figures

Figure 1. Model fit. Black boxes show model estimates of cases per 100,000 population and grey boxes show model estimates of mortality rates; thick horizontal lines and box edges show the median, 25th, and 75th percentiles; vertical lines extending from each box show 95% CrI. Blue dots indicate observed incidence rates and red dots show observed mortality rates.

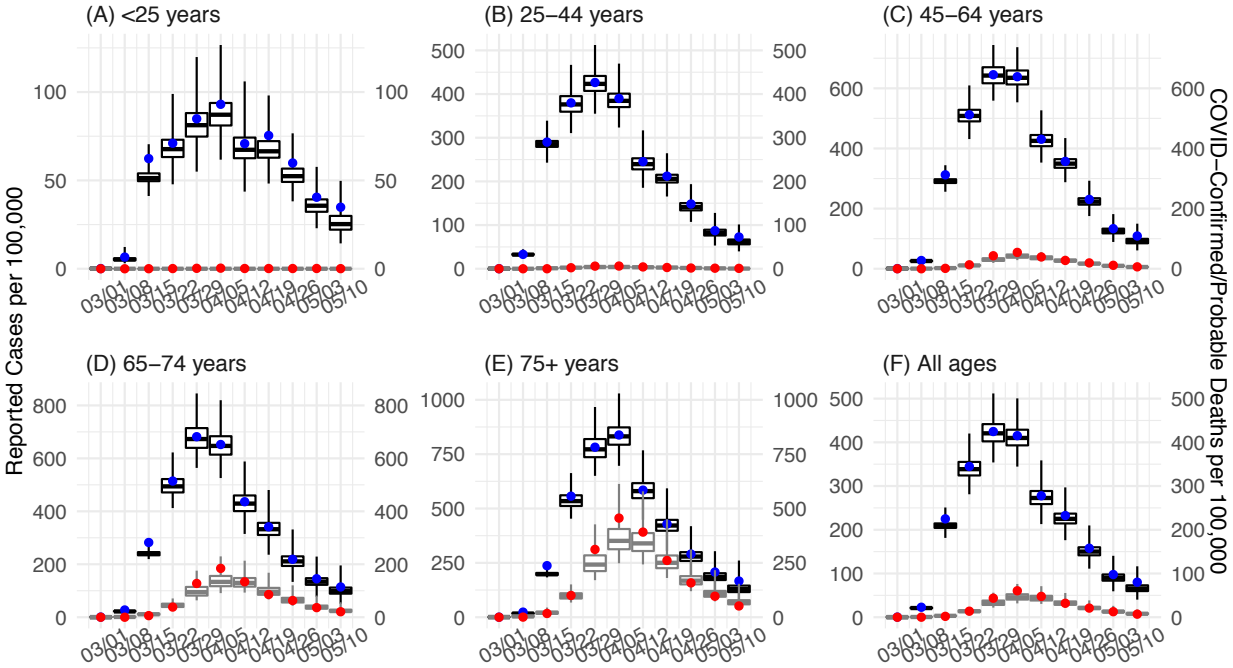


Figure 2. Estimated infection and case reporting rates over time. Boxplots show estimated infection rates. Red lines show the estimated median case reporting rate with surrounding areas show the 50% and 95% CrI. x-axis shows the first day of each week (mm/dd) from the week of March 1 to the week of March 10, 2020.

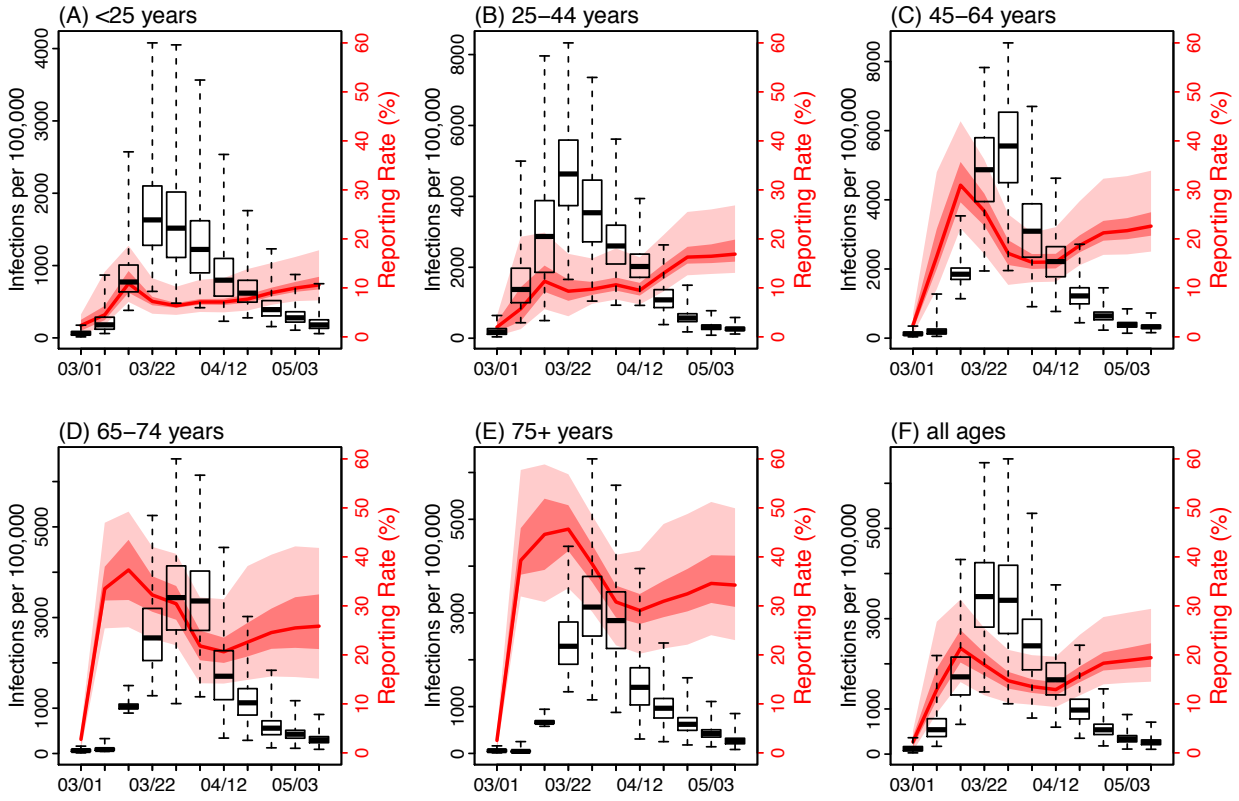
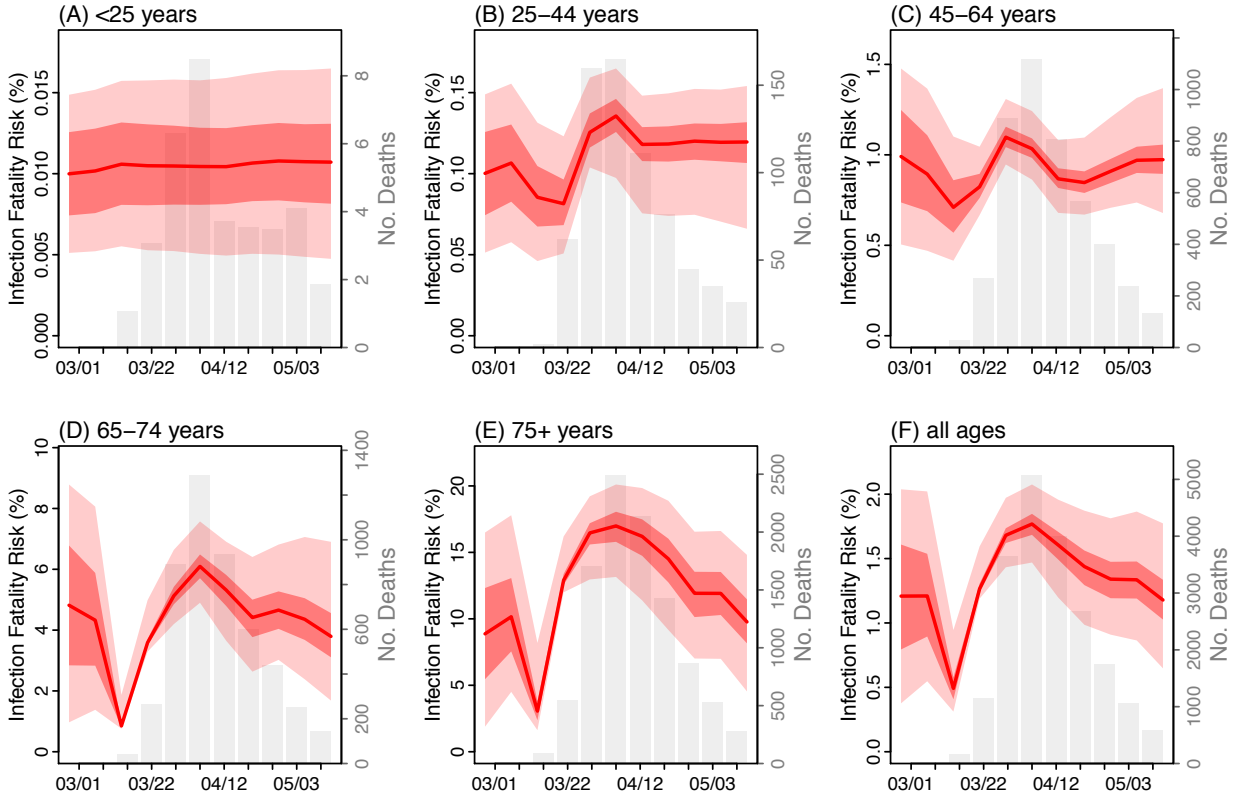


Figure 3. Estimated infection fatality risk. Red lines show the estimated median IFR with surrounding areas indicating the 50% and 95% CrI. For comparison, the grey bars show the number of deaths reported for each week from the week of March 1 to May 10, 2020.



Supplementary Material

Table S1. Prior ranges for main model parameters and variables. The spatial, temporal, and age resolution of each parameter or variable, estimated in the model-inference system, is specified in the column "Resolution". Note posterior parameter estimates can extend outside the specified prior ranges.

Parameter/variable	Symbol	Resolution	Prior range	Source/rationale
Initial exposed	$E(t=0)$	neighborhood- and age-group specific, estimated for the beginning of the Week of March 1, 2020	300 – 8000 total citywide, scaled by population size for each age group and neighborhood	Large uncertainties, used very wide range
Initial infectious	$I(t=0)$	neighborhood- and age-group specific, , estimated for the beginning of the Week of March 1, 2020	150 – 4000 total citywide, scaled by population size for each age group and neighborhood	Assumed to be half the initial exposed
Initial susceptible	$S(t=0)$	neighborhood- and age-group specific, estimated for the beginning of the Week of March 1, 2020	$N - E - I$	Assumed all were susceptible except for those initially exposed/infectious

Population size in each age group and neighborhood	N	neighborhood- and age-group specific	N/A	NYC intercensal population estimates for 2018[1]
Citywide transmission rate	β_{city}	Citywide, age-group specific, estimated for each week	[0.5, 1] per day overall; scaled by contact rate for each age group based on contact data from the POLYMOD study[2]	Based on R_0 estimates of around 1.5-4 for SARS-CoV-2[3-5]
Scaling of neighborhood transmission rate	b_i	neighborhood- and age-group specific, estimated for each week	[0.8, 1.2] for age groups under 65 years; [0.5, 1.5] for age groups 65 or older	Around 1; larger variation for elderly groups based on data
Latency period	Z	Citywide, age-group specific, estimated for each week	[2, 5] days	Incubation period: 5.2 days (95% CI: 4.1, 7)[3]; latency period is likely shorter than the incubation period
Infectious period	D	Citywide, age-group specific, estimated for each week	[2, 5] days	Time from symptom onset to hospitalization: 3.8 days (95% CI: 0, 12.0) in China,[6] plus 1-2 days viral shedding before symptom

				onset. We did not distinguish symptomatic/asymptomatic infections.
Multiplicative factor for mobility	m_1	Citywide, age-group specific, estimated for each week	[1, 2] for <1 year; [0.5, 1.5] for three age groups 1-24 years; [0.1, 1.5] for age group 25-44; [1, 2.5] for age groups 45 or older	Initial model testing showed transmission rates for younger age groups were more sensitive to changes in mobility whereas the two oldest age groups were not sensitive to mobility. For age groups with contact rates lower than the average (based on the POLYMOD study[2]), we raised the diagonal elements in the mobility matrix to the power of the relative contact rate (<1) to account for insensitivity of transmission rate in these age groups to mobility.

Multiplicative factor for neighborhood connectivity	m_2	Citywide, age-group specific, estimated for each week	[0.5, 2]	Likely around 1 but with large uncertainties
Mean of time from viral shedding to diagnosis	T_m	Citywide, age-group specific, estimated for each week	[3, 8] days	From a few days to a week from symptom onset to diagnosis/ reporting,[6] plus 1-2 days of viral shedding (being infectious) before symptom onset
Standard deviation (SD) of time from viral shedding to diagnosis	T_{sd}	Citywide, age-group specific, estimated for each week	[1, 3] days	To allow variation in time to diagnosis/reporting
Reporting rate	r	Citywide, age-group specific, estimated for each week	Starting from [0.001, 0.05] at time 0 and allowed to increase over time using space re-probing[7]	Large uncertainties
Infection fatality risk (IFR)		Citywide, age-group specific, estimated for each week	[5, 15]×10^{-4} for ages under 25; [5, 15]×10^{-3} for ages 25-44; [5, 15]×10^{-2} for ages 45-64; [0.01,	Based on previous estimates[8] but extend to have wider ranges

		0.1] for ages 65-74; [0.02, 0.2] for ages 75+;	
Time from diagnosis to death	Citywide	Gamma distribution with mean of 9.36 days and SD of 9.76 days	Based on *n*=15,686 COVID-19 confirmed deaths in NYC as of May 17, 2020.

References:

1. New York City Department of Health and Mental Hygiene. NYC DOHMH population estimates, modified from US census bureau interpolated intercensal population estimates, 2000-2018. . Updated August 2019. ed.
2. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 2008;5(3):e74. Epub 2008/03/28. PubMed PMID: 18366252; PMCID: 2270306.
3. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Liu M, Tu W, Chen C, Jin L, Yang R, Wang Q, Zhou S, Wang R, Liu H, Luo Y, Liu Y, Shao G, Li H, Tao Z, Yang Y, Deng Z, Liu B, Ma Z, Zhang Y, Shi G, Lam TTY, Wu JT, Gao GF, Cowling BJ, Yang B, Leung GM, Feng Z. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New Engl J Med. 2020.
4. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in Wuhan, China: A modelling study. Lancet. 2020. PubMed PMID: 32014114.
5. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-cov-2). Science. 2020;368(6490):489-93.

6. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, Li M, Zheng W, Yi L, Chen X, Wu Q, Liang Y, Wang X, Yang J, Sun K, Longini IM, Jr., Halloran ME, Wu P, Cowling BJ, Merler S, Viboud C, Vespignani A, Ajelli M, Yu H. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside hubei province, China: A descriptive and modelling study. The Lancet Infectious diseases. 2020. PubMed PMID: 32247326.
7. Yang W, Shaman J. A simple modification for improving inference of non-linear dynamical systems. arXiv. 2014:1403.6804. Epub 26 Mar 2014.
8. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, Cuomo-Dannenburg G, Thompson H, Walker PGT, Fu H, Dighe A, Griffin JT, Baguelin M, Bhatia S, Boonyasiri A, Cori A, Cucunuba Z, FitzJohn R, Gaythorpe K, Green W, Hamlet A, Hinsley W, Laydon D, Nedjati-Gilani G, Riley S, van Elsland S, Volz E, Wang H, Wang Y, Xi X, Donnelly CA, Ghani AC, Ferguson NM. Estimates of the severity of coronavirus disease 2019: A model-based analysis. The Lancet Infectious diseases. 2020. PubMed PMID: 32240634; PMCID: PMC7158570.