Estimating the cumulative incidence of SARS-CoV-2 infection and the infection fatality ratio in light of waning antibodies

Kayoko Shioda[1]*, Max SY Lau[2], Alicia NM Kraay[3], Kristin N Nelson[3], Aaron J Siegler[3], Patrick S Sullivan[3], Matthew H Collins[4], Joshua S Weitz[5], Benjamin A Lopman[3]

[1]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA
[2]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, USA
[3]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA
[4]Division of Infectious Diseases, Department of Medicine, Emory University School of Medicine, Atlanta, GA, USA
[5]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA;
School of Physics, Georgia Institute of Technology, Atlanta, GA, USA
*Corresponding author. Email: kayoko.shioda@emory.edu; kayoko.shioda@aya.yale.edu

Abstract

Background: Serology tests can identify previous infections and facilitate estimation of the number of total infections. However, immunoglobulins targeting severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have been reported to wane below the detectable level of serological assays. We estimate the cumulative incidence of SARS-CoV-2 infection from serology studies, accounting for expected levels of antibody acquisition (seroconversion) and waning (seroreversion), and apply this framework using data from New York City (NYC) and Connecticut.

Methods: We estimated time from seroconversion to seroreversion and infection fatality ratio (IFR) using mortality data from March-October 2020 and population-level cross-sectional seroprevalence data from April-August 2020 in NYC and Connecticut. We then estimated the daily seroprevalence and cumulative incidence of SARS-CoV-2 infection.

Findings: The estimated average time from seroconversion to seroreversion was 3-4 months. The estimated IFR was 1.1% (95% credible interval: 1.0-1.2%) in NYC and 1.4% (1.1-1.7%) in Connecticut. The estimated daily seroprevalence declined after a peak in the spring. The estimated cumulative incidence reached 26.8% (24.2-29.7%) and 8.8% (7.1-11.3%) at the end of September in NYC and Connecticut, higher than maximum seroprevalence measures (22.1% and 6.1%), respectively.

Interpretation: The cumulative incidence of SARS-CoV-2 infection is underestimated using cross-sectional serology data without adjustment for waning antibodies. Our approach can help quantify the magnitude of underestimation and adjust estimates for waning antibodies.

Funding: This study was supported by the US National Science Foundation and the National Institute of Allergy and Infectious Diseases.

Keywords: Antibody; ascertainment bias; case ascertainment ratio; COVID-19; cumulative incidence; infection fatality ratio; SARS-CoV-2; seroconversion; seroprevalence; seroreversion; waning antibody.

1

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes Coronavirus Disease 2019 (COVID-19), rapidly spread across the world in 2020.[1] Globally, there have been 36 million documented cases and more than one million COVID-19 associated fatalities as of October 15, 2020, with case counts continuing to increase.[2] Reliable measurement of infection history in a population is a critical epidemiologic outcome, and is needed to derive several key epidemiologic indices such as the infection fatality ratio (IFR). However, the number of SARS-CoV-2 infections reported through public health surveillance mechanisms is underestimated because of the limited capacity of testing and surveillance systems, an overwhelmed healthcare system, low healthcare-seeking behavior among those with mild disease, imperfect sensitivity of diagnostic tests (especially rapid antigen tests) and a large fraction of infections that are asymtomatic.[3] Assays that detect viral antigen or genomic material cannot identify individuals who were previously infected once viral material is no longer present. In contrast, serology tests measuring the level of immunoglobulins have the potential to identify those previously infected. Serology results can be used to generate an estimate of the cumulative incidence, but this requires the relatively strong assumption that antibodies persist permanently after infection. Because antibody levels for SARS-CoV-2 wane over time and can become undetectable, some previously infected individuals could have already returned to a seronegative status at the time of testing (i.e. seroreversion). Hence, new methods are needed to account for waning antibodies when using cross-sectional serology data to estimate cumulative incidence of SARS-CoV-2 infection. The adjusted cumulative incidence can then be used to obtain more accurate estimates of critical epidemiologic measures, such as the IFR and the case ascertainment ratio.

Population-level serosurveys for SARS-CoV-2 have been conducted across the world, with variation in geographic scale, sample demographics, sampling mechanisms, and testing methods.[4,5] Although the reported incidence of COVID-19 varies widely by location, a consistent

finding in all settings is that total estimated infections vastly outnumber confirmed cases. For example, the Centers for Disease Control and Prevention (CDC) and commercial laboratories conducted large-scale geographic longitudinal serosurveys in 10 sites in the United States in the spring and summer of 2020.[6] Seroprevalence ranged from 1% to 6.9% across sites. That study estimated that the number of total infections was 6-24 times higher than that of documented cases, while acknowledging that these ratios varied widely depending on the timing of sampling or the stage of the epidemic in each location. These early population-level serosurveys provided critical insights on the true burden of COVID-19; however, as we enter the 10th month of the US epidemic, the ability to estimate the cumulative incidence directly from serosurveys is increasingly limited because antibody levels continue to wane and the corresponding serologic 'record' of historical infection is lost. Indeed, Ibarrondo *et al.* estimated that the half-life of anti-SARS-CoV-2 spike receptor-binding domain IgG was 36 days.[7] Patel *et al.* reported that over half (11/19) of health care personnel who tested seropositive in early April became seronegative at a second visit in June (approximately 60 days after the baseline).[8] Taken together, these findings suggest that serosurveys have likely failed to recognize previous infection in those whose antibody levels have already waned below the detectable limit at the time of sampling. An adjustment for waning antibody kinetics must account for the interacting timescales of antibody kinetics, case incidence and the period over which a serosurvey is conducted.

Here, we describe a framework for estimating the cumulative incidence and IFR of SARS-CoV-2 from population-level cross-sectional serology data and mortality data, by adjusting for the timeline of seroconversion (acquisition of the detectable level of antibodies) and seroreversion (loss of detectable antibodies). We apply this framework to data from New York City and Connecticut, because these two sites observed a large wave of COVID-19 cases that lasted for a relatively short period in the spring 2020, followed by low case counts in the summer and early fall (Figure S1). We note that it is critical to distinguish the ability to detect antibodies from immunity to reinfection, which may persist longer than antibodies are

3

detectable.[9] We make no claim that the time from seroconversion to seroreversion estimated from this framework reflects the duration of immunity protection against reinfection.

Methods

Population-level cross-sectional seroprevalence data

Details of the serosurvey conducted by the CDC and commercial laboratories can be found elsewhere.[6,10] Briefly, the survey collected blood samples in 10 US sites (Connecticut, Louisiana, Minneapolis-St Paul-St Cloud metro area, Missouri, New York City metro area, Philadelphia metro area, San Francisco Bay area, South Florida, Utah, and Western Washington State) from March to July 2020, and expanded the serosurvey to all 50 states in August. Blood specimens were originally collected for reasons unrelated to COVID-19, such as for routine medical care or sick visits. Multiple rounds of surveys have been conducted at each site, approximately every 3-4 weeks (Table S1). Each round tested approximately 1,800 samples from each site.[11] Samples were tested by an enzyme-linked immunosorbent assay (ELISA) against the SARS-CoV-2 spike protein that detects the total immunoglobulin response (IgA, IgM, and IgG) with sensitivity 96% (95% confidence interval (CI): 98.3-99.9%) and specificity 99.3% (95% CI: 98.3-99.9%).[6] Publicly available seroprevalence data adjusted for age and sex distributions were downloaded on October 8, 2020.[12] We focused our analysis on New York City and Connecticut because the single, short wave of infection in the spring allowed us to evaluate how the antibodies acquired over a short period subsequently waned during a time when minimal new infections were occurring in these locations (Figure S1). More details can be found in the supplementary methods.

Mortality data and case data

Mortality data are less subject to changes in testing capacity and guidelines over time than case data, so we used mortality data to estimate parameters. We analyzed daily time

series data on COVID-19 associated deaths from March to September 2020 in New York City and Connecticut. New York City data and Connecticut data were downloaded from their government websites on October 2 and 6, 2020, respectively.[13,14] We used the total (i.e., probable and confirmed) deaths for our mortality time series (supplementary methods). To account for a delay distribution between symptom onset and death, we used data on the date of symptom onset and date of death for 6,999 COVID-19 associated deaths in Georgia, USA from February to October 2020.

To estimate case ascertainment ratios (see the "Model" section for detail), we also utilized daily time series data for the number of documented cases in New York City and Connecticut from their government websites (supplementary methods).[13,14]

Model

We estimated the IFR and time from seroconversion to seroreversion using the reported number of deaths and CDC serosurvey data (Figure 1). Using a Markov chain Monte Carlo (MCMC) analysis, we estimated the number of seropositive individuals on day t (S_t) as follows:

$$S_t = \sum_{k=1}^{t-1} I_{t-k} \int_{y=t-k}^{t} g_y \times (1 - Z_{t-y}), 1 \le t \le tmax, 1 \le k \le t-1 \quad \text{(Equation 1)}$$

where I_t is the total number of SARS-CoV-2 infections on day t, which is the number of reported deaths divided by the estimated IFR, after accounting for the delay between symptom onset and death (supplementary methods). We estimated the IFR, assuming that the IFR was constant over time in the main analysis and relaxed this assumption in a sensitivity analysis (see the "sensitivity analysis" section for details). Additional parameters include $tmax$, the total number of days in the daily time series data for COVID-19 deaths, and g, the probability density function of the Weibull distribution for time from symptom onset to seroconversion with the mean 11.5 days and standard deviation 5.7 days,[15] which is consistent with other reports.[16–19] For asymptomatic cases, g represents the time from onset of infectiousness to seroconversion (supplementary

5

methods). Z is the cumulative density function of the Weibull distribution for time from seroconversion to seroreversion. Therefore, $\int_{y=t-k}^{t} g_y \times (1 - Z_{t-y})$ in Equation 1 represents the probability that an individual seroconverted before day t and seroreverted after day t, which in other words means the probability that an individual remains seropositive on day t. We estimated the mean of the Weibull distribution for time from seroconversion to seroreversion, while fixing standard deviation at 50 days (supplementary method). The daily seroprevalence (P_t) was calculated by dividing S_t by the population (8.3 million for New York City and 3.7 million for Connecticut). We compared the estimated P_t with the reported seroprevalence in each round of the CDC commercial laboratory serosurvey, and calculated the log-likelihood assuming the binomial distribution, which was then used in the MCMC analysis for parameter estimation (supplementary methods).

Given an estimated model from above, the number of individuals who seroconverted on day t (C_t) may be calculated as follows:

$$C_t = \sum_{k=1}^{t-1} I_k \times G_{t-k} \quad \text{(Equation 2)}$$

where G is the cumulative density function of the aforementioned Weibull distribution for time from symptom onset to seroconversion. G_{t-k} represents the probability that infected individuals on day k (I_k) had seroconverted by day t. We calculated the cumulative incidence on day t by dividing the cumulative sum of C_t by the population. We also calculated a cumulative case ascertainment ratio (A_t) as follows:

$$A_t = \frac{\sum_{t=1}^{tmax} T_t}{\sum_{t=1}^{tmax} C_t} \quad \text{(Equation 3)}$$

where T_t is the number of documented cases on day t (supplementary methods). The ascertainment bias was calculated as $\frac{1}{A_t}$.

The median of the posterior samples was reported as a point estimate, and 2.5th and 97.5th percentiles were reported as 95% credible intervals (CrIs). All analyses were performed

with R (Vienna, Austria). The code can be found in the following github repository:

https://github.com/lopmanlab/SARS-CoV-2_CumInc_WaningAntibodies.

Sensitivity analysis

We performed sensitivity analyses to evaluate how sensitive our results were to different assumptions. We changed the fixed value of the standard deviation for time from seroconversion to seroreversion from 50 days to 20 and 70 days. We selected these values based on previous findings.[20] We also relaxed the assumption of the constant IFR. It has been reported that the IFR and case fatality ratio for COVID-19 declined over time in New York City[21] and in other countries.[20] Therefore, we assumed that the IFR decreased by 5% per week from mid-March to the end of July 2020, reflecting findings reported by the previous studies,[21] and the model estimated the IFR in mid-March (before the decline started). Lastly, we used data from Wuhan, China to inform delay between symptom onset and death,[22] instead of the Georgia data to evaluate the impact of variation in this distribution on parameter inference.[23]

Ethical considerations

Seroprevalence data, mortality data, and case data in New York City and Connecticut were publicly available, deidentified, and aggregated. The Georgia Department of Public Health Institutional Review Board (IRB) has determined that the project is exempt from the requirement for IRB review and approval.

Results

Estimated timeline of seroreversion, IFR, and case ascertainment ratio

The average time from seroconversion to seroreversion was estimated to be 4.0 months (95% CrI: 3.6-4.6 months) using the New York City data and 3.0 months (95% CrI: 2.3-4.1 months) using the Connecticut data with the standard deviation fixed at 50 days (Figure 2).

More than 85% of the infected individuals were estimated to become seronegative due to waning antibodies within six months after seroconversion (Table 1).

The IFR was estimated to be 1.1% (95% CrI: 1.0-1.2%) for New York City and 1.4% (95% CrI: 1.1-1.7%) for Connecticut. The estimated case ascertainment ratio increased rapidly in the early phase of the pandemic and continued to gradually increase from May to October in both sites (Figure 3). The ascertainment ratio was estimated to reach 13% (95% CrI: 12-14%) in New York City and 18% (95% CrI: 14-23%) in Connecticut at the end of September 2020, suggesting that the number of estimated infections was 7.7 and 5.6 times greater than the number of documented cases in New York City and Connecticut, respectively.

Estimated daily seroprevalence and cumulative incidence

In both New York City and Connecticut, the estimated daily seroprevalence decreased over time after the first wave of COVID-19, diverging from the estimated cumulative incidence adjusted for waning antibodies (Figure 4). The cumulative incidence was estimated to reach 26.8% (95% CrI: 24.2-29.7%) in New York City and 8.8% (95% CrI: 7.1-11.3%) in Connecticut at the end of September 2020. In contrast, the estimated daily seroprevalence peaked at 22.1% (95% CrI: 20.5-23.6%) in mid-May and declined to 4.9% (95% CrI: 3.9-6.9%) by the end of September in New York City. In Connecticut, the estimated daily seroprevalence peaked at 6.1% (95% CrI: 5.3-7.1%) at the end of May and early June and declined to1.3% (95% CrI: 0.9-2.0%) by the end of September.

Sensitivity analysis

We ran the analysis with different values of the standard deviation for time from seroconversion to seroreversion (20, 50, and 70 days). The estimated IFR and average time from seroconversion to seroreversion were robust to this variation for Connecticut, but the estimated mean time increased as the standard deviation increased for the New York City data

(Table S2). Although changes in the standard deviation had small effects on the estimated cumulative incidence, the speed of decline in the daily seroprevalence after the peak varied by standard deviation (Figure S2). When the standard deviation was 20 days, the daily seroprevalence declined rapidly after the peak and became 1.2% (95% CrI: 0.9-2.1%) at the end of September, while it declined more slowly with the standard deviation 70 days (7.4% (95% CrI: 6.1-8.7%) at the end of September).

Under the assumption of the decreasing IFR (Figure S3), the average time from seroconversion to seroreversion was estimated to be 4.0 months (95% CrI: 3.5-4.8 months) with the standard deviation 50 days in New York City, which was consistent with the result using the constant IFR (Table S2). The estimated IFR in mid-March was 2.9% (95% CrI: 2.6-3.3%), which automatically decreased by 5% per week and reached 1.1% (95% CrI: 0.9-1.2%) at the end of July (Figure S3). The average IFR from March to September was 1.4% (95% CrI: 1.2-1.6%). The daily seroprevalence and cumulative incidence did not appreciably change between the models using the constant IFR vs. decreasing IFR (Table S2). We also used data from Wuhan, China to account for variation in the delay between symptom onset and death (Figure S4), and found that results were robust to this change (Table S2).

Discussion

A reliable estimate of the cumulative number of people who have been infected with SARS-CoV-2 is critical for understanding and, ultimately, controlling the COVID-19 pandemic. Measuring severity (in the form of the IFR), monitoring progress towards a herd immunity threshold, and predicting the impact of vaccination all depend on a robust estimate of cumulative incidence of infection. Given evidence that anti-SARS-CoV-2 antibodies wane below the detection limit among a substantial portion of the population, we developed an analytical framework to estimate cumulative incidence of infection from cross-sectional serological surveys. First, our model was able to capture the observed declines in seroprevalence after the

9

first wave of COVID-19 in the spring of 2020 in New York City and Connecticut. Second, we were able to estimate that the cumulative incidence was 26.8% (95% CrI: 24.2-29.7%) in New York City and 8.8% (95% CrI: 7.1-11.3%) in Connecticut by the end of September 2020, which was greater than the peak daily seroprevalence of 22.1% (95% CrI: 20.5-23.6%) in mid-May in New York City and 6.1% (95% CrI: 5.3-7.1%) at the end of May in Connecticut. Cumulative incidence could be underestimated by cross-sectional serosurveys only a few months after the first cases of SARS-CoV-2 in a population, because individuals no longer have a detectable level of antibodies 3-4 months after seroconversion on average. Taken together, our findings suggest that cumulative incidence of SARS-CoV-2 should not be estimated directly from cross-sectional serology data, especially at later stages of an outbreak. Rather, cumulative incidence should be adjusted given impacts of waning of detectable antibodies in serological assays.

The timeline of seroconversion and seroreversion should ideally be determined by individual-level longitudinal studies that follow up each infected individual multiple times, as frequently as possible, for a long enough period to observe seroreversion. The timeline for seroreversion after SARS-CoV-2 infection reported by the longitudinal studies published to date varies. Some studies have reported rapid waning of IgG, with substantial attrition of the seropositive population in as little as 60 days,[7,24] while others reported that antibodies remained above the detectable threshold for at least 82 days after symptom onset[25] or 120 days after qPCR diagnosis of SARS-CoV-2.[26] These differences are likely due to limitations or heterogeneity of the studies reported to date, including short follow-up times, small sample sizes, and different serology testing methods and analytic sensitivities.[27] Most of the longitudinal studies published to date have followed participants for 14-150 days after symptom onset or baseline visits, which is not long enough to understand the complete timeline of seroreversion for all participants, particularly for IgG.[7,8,19,20,24,25] The variation in clinical and demographic characteristics and severity of infection of participants in each study has also likely influenced these different findings.[16,17,20,24,28] Infected individuals with mild or no symptoms who may exhibit

10

lower titer and shorter time to seroreversion are often not included in these longitudinal studies. Therefore, assessing the timeline using population-level data, such as cross-sectional serology data and mortality data, is a viable alternative approach. Using our framework, infected individuals were estimated to remain seropositive for about 3-4 months on average. The average duration estimated by the New York City data and Connecticut data were mostly in agreement, with overlapping credible intervals. Differences in these average durations could be attributable to differences in demographic and clinical characteristics of infected individuals and differences in the testing and reporting practices. It is also important to note that the timeline of seroconversion and seroreversion is dependent on serological assays and targeted immunoglobulins that may have different thresholds to define seropositivity.

We estimated that the IFR was 1.1% (95% CrI: 1.0-1.2%) in New York City, which was consistent with estimates from the routine care group at Mount Sinai Hospital in New York City.[29] Yang *et al.* estimated the time-varying IFR for New York City (1.39%; 95% CrI: 1.04-1.77),[21] matching with our estimate under the assumption of decreasing IFR (1.4% (95% CI: 1.2-1.6%) on average between March and September 2020). Although the sensitivity analyses did not change our central conclusions, we note that different values of the standard deviation for time from seroconversion to seroreversion changed the estimated seroprevalence in the later phase of the study period (August and September 2020; Figure 4 and Figure S2). We ran the model with standard deviation 20, 50, and 70 days, because the estimated standard deviation for time from seroconversion to seroreversion for IgA and IgM was approximately 50 and 25 days, respectively.[20] Our standard deviation could be larger than those estimated for small study populations, as we used population-level data that likely have greater variation in demographic and clinical characteristics. Analyzing seroprevalence data in August and after could provide critical information on how population-level seroprevalence declines over time and would enable better estimates of the timeline of antibody waning.

The data used in this study have certain limitations. The samples collected for the CDC seroprevalence data may not be representative of the general population, as Havers *et al.* discussed.[6] The data on the date of symptom onset among cases who died in Georgia may be incorrectly recorded and subject to recall bias as they are largely self-reported. Changes in testing and reporting practice over time likely affected the number of reported deaths, although mortality data are less sensitive to these changes compared to case data. We conducted the sensitivity analysis with a different source of data on time from onset to death and considered time-varying IFR instead of constant IFR, and found that results were robust to these changes (Table S2).

Detection of antibodies by serological assays may not correlate to protection from reinfection or disease, and thus, time from seroconversion to seroreversion estimated in our study is not necessarily equivalent to the duration of protective immunity. Several studies have noted strong correlation between results from assays based on antibody binding (such as ELISA) and neutralization testing,[16,20,25] which may allow protective immunity to be inferred from simpler serologic tests once more comprehensive longitudinal data sets are available. Moreover, T cell immunity may make an important contribution to protective immunity but is not assessed by serologic methods.[30] Therefore, additional studies are needed to define immunologic determinants of SARS-CoV-2 protection against reinfection and severe disease.

Our findings suggest that the cumulative incidence estimated from serology data needs to be adjusted for seroreversion. We intend to apply this framework to other seroprevalence studies[31] and we suggest others conducting serological surveys consider doing so as well. Our framework could readily be applied to other data to estimate the duration of seroreversion, IFR, ascertainment ratios, daily seroprevalence, and cumulative incidence.

Acknowledgements

Conflicts of interest

BAL reports personal fees from Takeda Pharmaceutical, personal fees from CDC Foundation, and personal fees from Hall Booth Smith, P.C., outside the submitted work. Other authors do not have conflicts of interest.

References

1   Zhu N, Zhang D, Wang W, *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020; **382**: 727–33.

2   COVID-19 Map - Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/map.html (accessed Oct 8, 2020).

3   Wu SL, Mertens AN, Crider YS, *et al.* Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications* 2020; **11**: 4507.

4   Eckerle I, Meyer B. SARS-CoV-2 seroprevalence in COVID-19 hotspots. *The Lancet* 2020; **396**: 514–5.

5   Arora RK, Joseph A, Wyk JV, *et al.* SeroTracker: a global SARS-CoV-2 seroprevalence dashboard. *The Lancet Infectious Diseases* 2020; **0**. DOI:10.1016/S1473-3099(20)30631-9.

6   Havers FP, Reed C, Lim T, *et al.* Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May 12, 2020. *JAMA Intern Med* 2020; published online July 21. DOI:10.1001/jamainternmed.2020.4130.

7   Ibarrondo FJ, Fulcher JA, Goodman-Meza D, *et al.* Rapid Decay of Anti–SARS-CoV-2 Antibodies in Persons with Mild Covid-19. *New England Journal of Medicine* 2020; **383**: 1085–7.

8   Patel MM, Thornburg NJ, Stubblefield WB, *et al.* Change in Antibodies to SARS-CoV-2 Over 60 Days Among Health Care Personnel in Nashville, Tennessee. *JAMA* 2020; published online Sept 17. DOI:10.1001/jama.2020.18796.

9   Poland GA, Ovsyannikova IG, Kennedy RB. SARS-CoV-2 immunity: review and applications to phase 3 vaccine candidates. *The Lancet* 2020; **0**. DOI:10.1016/S0140-6736(20)32137-1.

10   Havers FP, Reed C, Lim TW, *et al.* Seroprevalence of Antibodies to SARS-CoV-2 in Six Sites in the United States, March 23-May 3, 2020. *medRxiv* 2020; : 2020.06.25.20140384.

11   Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19) Commercial Laboratory Seroprevalence Surveys. 2020; published online Feb 11. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/commercial-lab-surveys.html (accessed Oct 8, 2020).

12   Centers for Disease Control and Prevention. CDC COVID Data Tracker. 2020; published online March 28. https://covid.cdc.gov/covid-data-tracker (accessed Oct 8, 2020).

13   COVID-19 Tests, Cases, Hospitalizations, and Deaths (Statewide) | Connecticut Data. https://data.ct.gov/Health-and-Human-Services/COVID-19-Tests-Cases-Hospitalizations-and-Deaths-S/rf3k-f8fg (accessed Oct 15, 2020).

14   COVID-19: Data Details on Deaths - NYC Health. https://www1.nyc.gov/site/doh/covid/covid-19-data-deaths.page (accessed Oct 15, 2020).

15   Iyer AS, Jones FK, Nodoushania A, *et al.* Dynamics and significance of the antibody response to SARS-CoV-2 infection. *medRxiv* 2020; published online July 20. DOI:10.1101/2020.07.18.20155374.

16   Okba NMA, Müller MA, Li W, *et al.* Severe Acute Respiratory Syndrome Coronavirus 2−Specific Antibody Responses in Coronavirus Disease Patients - Volume 26, Number 7—July 2020 - Emerging Infectious Diseases journal - CDC. DOI:10.3201/eid2607.200841.

17   Zhao J, Yuan Q, Wang H, *et al.* Antibody Responses to SARS-CoV-2 in Patients With

Novel Coronavirus Disease 2019. *Clin Infect Dis* DOI:10.1093/cid/ciaa344.

18 Huang AT, Garcia-Carreras B, Hitchings MDT, *et al.* A systematic review of antibody mediated immunity to coronaviruses: kinetics, correlates of protection, and association with severity. *Nature Communications* 2020; **11**: 4704.

19 To KK-W, Tsang OT-Y, Leung W-S, *et al.* Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *The Lancet Infectious Diseases* 2020; **20**: 565–74.

20 Iyer AS, Jones FK, Nodoushani A, *et al.* Persistence and decay of human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in COVID-19 patients. *Science Immunology* 2020; **5**. DOI:10.1126/sciimmunol.abe0367.

21 Yang W, Kandula S, Huynh M, *et al.* Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: a model-based analysis. *The Lancet Infectious Diseases* 2020; : S1473309920307696.

22 Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China | Nature Medicine. https://www.nature.com/articles/s41591-020-0822-7 (accessed Nov 7, 2020).

23 O'Driscoll M, Dos Santos GR, Wang L, *et al.* Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* 2020; : 1–9.

24 Long Q-X, Tang X-J, Shi Q-L, *et al.* Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat Med* 2020; **26**: 1200–4.

25 Wajnberg A, Amanat F, Firpo A, *et al.* SARS-CoV-2 infection induces robust, neutralizing antibody responses that are stable for at least three months. *medRxiv* 2020; : 2020.07.14.20151126.

26 Gudbjartsson DF, Norddahl GL, Melsted P, *et al.* Humoral Immune Response to SARS-CoV-2 in Iceland. *N Engl J Med* 2020; **383**: 1724–34.

27 Takahashi S, Greenhouse B, Rodríguez-Barraquer I. Are SARS-CoV-2 seroprevalence estimates biased? *J Infect Dis* DOI:10.1093/infdis/jiaa523.

28 Viral Kinetics and Antibody Responses in Patients with COVID-19 | medRxiv. https://www.medrxiv.org/content/10.1101/2020.03.24.20042382v1 (accessed Oct 19, 2020).

29 Stadlbauer D, Tan J, Jiang K, *et al.* Repeated cross-sectional sero-monitoring of SARS-CoV-2 in New York City. *Nature* 2020; : 1–7.

30 Cox RJ, Brokstad KA. Not just antibodies: B cells and T cells mediate immunity to COVID-19. *Nature Reviews Immunology* 2020; **20**: 581–2.

31 Siegler AJ, Sullivan PS, Sanchez T, *et al.* Protocol for a national probability survey using home specimen collection methods to assess prevalence and incidence of SARS-CoV-2 infection and antibody response. *Annals of Epidemiology* 2020; **49**: 50–60.

32 Delignette-Muller ML, Dutang C. fitdistrplus: Package for Fitting Distributions. *J Stat Soft* 2015; **64**. DOI:10.18637/jss.v064.i04.

Table 1. Percentage of infected individuals whose antibody level has declined and become undetectable by serological assays within *t* months since seroconversion.

Months since seroconversion	New York City (95% CrI)	Connecticut (95% CrI)
1 month	1.8% (0.6-3.9)	8.7% (1.7-22.2)
2 months	10.5% (5.2-17.7)	29.2% (10-49.4)
3 months	28.4% (17.3-39.8)	53.7% (27.5-71.2)
4 months	50.8% (36.5-62.3)	73.6% (49.7-84.7)
5 months	72.5% (59.8-80.8)	87.3% (71.7-92.6)
6 months	87.5% (79.4-91.7)	94.6% (87-96.6)
9 months	99.8% (99.6-99.8)	99.8% (99.8-99.8)
12 months	100.0% (100.0-100.0)	100% (100-100)

The mean of the Weibull distribution for time from seroconversion to seroreversion was estimated while fixing the standard deviation at 50 days. Abbreviation: CrI, credible interval.
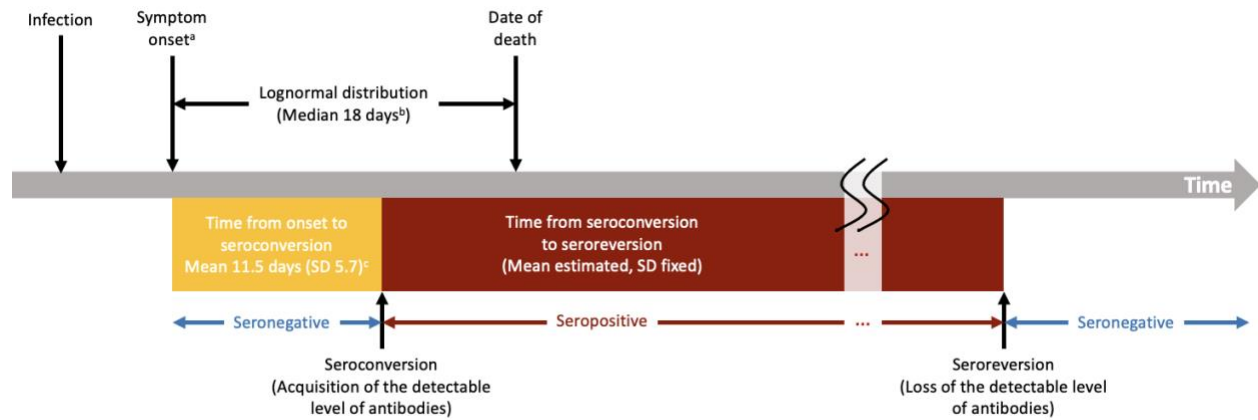
Figure 1. Structure of the analytic framework.

[a]Onset of infectiousness for asymptomatic cases; [b]Data from Georgia Department of Public Health; [c]Iyer *et al.* medRxiv 2020. Abbreviations: SD, standard deviation.
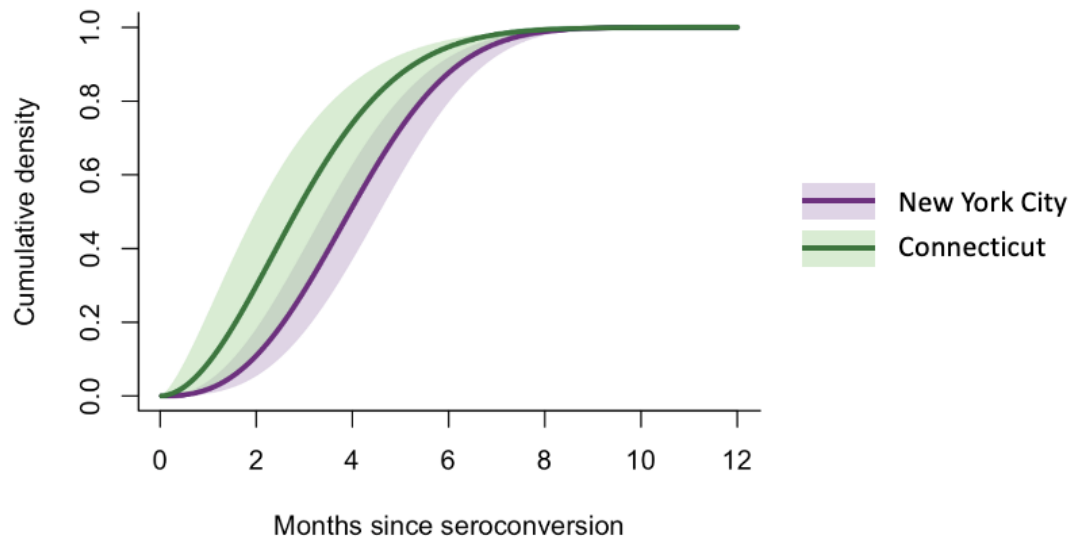
Figure 2. Cumulative density function for the estimated Weibull distribution for time from seroconversion to seroreversion (New York City and Connecticut).
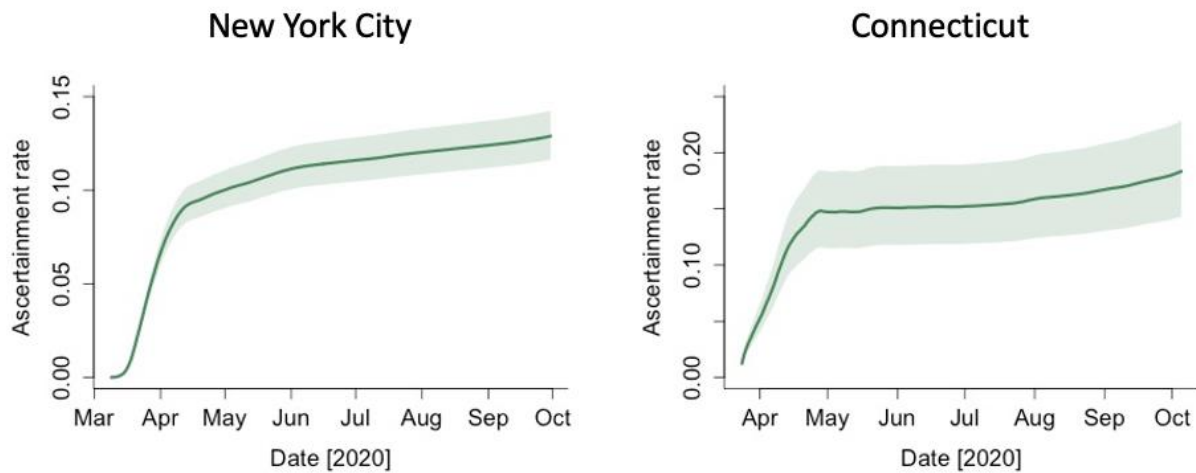Lines and shaded areas represent the posterior median and 95% credible intervals, respectively.

Figure 3. Estimated case ascertainment ratio in New York City and Connecticut in 2020.
The case ascertainment ratio was calculated in Equation 3. Lines represent the 50th percentile
of the posterior distributions, and shaded areas represent 95% credible intervals.
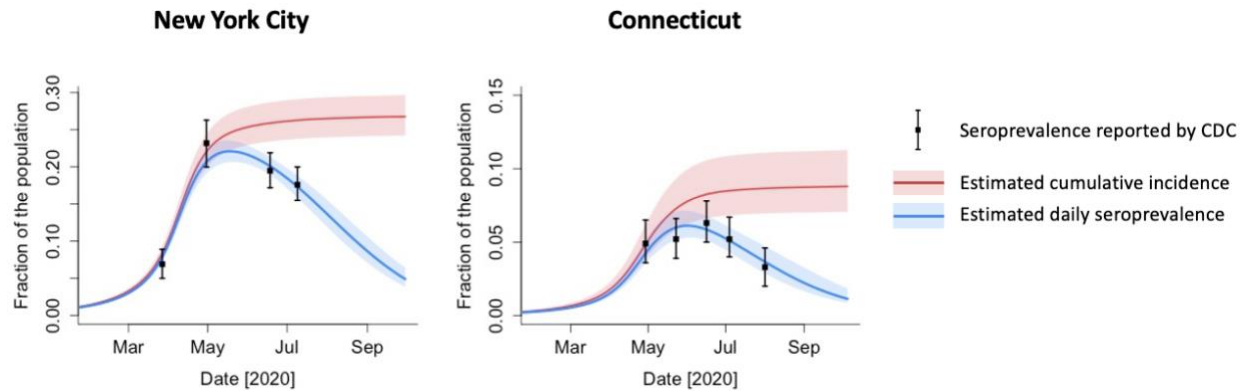
Figure 4. Estimated daily seroprevalence and cumulative incidence of SARS-CoV-2 infection in New York City and Connecticut in 2020.

Black dots and bars represent the point estimates and 95% confidence intervals for the seroprevalence reported by CDC.[12] Abbreviations: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; CDC, Centers for Disease Control and Prevention.

Supplementary materials

Supplementary methods

More details on the population-level cross-sectional seroprevalence data

For New York City, the catchment area covered Manhattan, Bronx, Queens, Kings, and Nassau Counties in the first round (catchment population: 9.26 million), with Suffolk, Westchester, and Richmond Counties added in the second, third, and fourth round (total catchment population: 12.2 million). Because the county-specific serology data were not available, we assumed that the seroprevalence was similar across the different catchment areas in the New York City metro area. For Connecticut, the catchment area included all counties in the state.

More details on the mortality data and case data

The mortality data for New York City covered Bronx, Kings, Manhattan. Queens, Richmond Counties. The mortality data covered all counties in the state. For Connecticut, we calculated a 7-day rolling average, because there were no data (i.e., zero deaths reported) on weekends. For New York City, we used the reported daily counts.

For the number of documented cases (T_t in Equation 3), we used the number of individuals with positive PCR and/or antigen tests in New York City. In Connecticut, T_t denotes the number of total cases (sum of confirmed and probable cases), because confirmed cases and probable cases were not reported separately before May 31, 2020.

Total number of SARS-CoV-2 infections

To estimate the total number of SARS-CoV-2 infections (I_t) and the IFR, we first estimated the date of symptom onset for fatal cases, by fitting gamma, Weibull, and lognormal distributions to the observed data on time from symptom onset to death for 6,999 COVID-19

21

associated deaths in Georgia. The best fitted distribution was selected based on goodness-of-fit tests (Cramer-von Mises, Kolmogorov-Smirnov and Anderson-Darling statistics), using the fitdistrplus R package.[32] Among those who died on day t (d_t), we calculated the number of individuals who developed symptoms on day $t\text{-}k$ (o_{t-k}) as follows:

$$d_t = \sum_{k=1}^{\infty} o_{t-k} \times (F_k - F_{k-1}), 1 \le k < \infty$$

where F is the cumulative density function of the best fitted distribution for time from onset to death, conditional upon fatality. We then calculated the total number of infections on day t (I_t) as follows:

$$I_t = \frac{1}{IFR} \times o_t$$

This framework assumes that the duration of incubation period is approximately equal to the duration of latent period, and therefore, day t represents the onset of infectiousness for asymptomatic individuals.

Markov chain Monte Carlo (MCMC) analysis

In the MCMC analysis, we obtained at least 50,000 posterior samples after a burn-in period of 20,000 iterations. Model convergence was assessed based on trace plots of posterior samples and the Geweke diagnostic test. We fixed the standard deviation for the Weibull distribution for time from seroconversion to seroreversion in the analysis, as parameters were not identifiable otherwise.

Supplementary tables

Table S1. Date of the CDC longitudinal commercial laboratory seroprevalence surveys in New York City metro area and Connecticut in 2020.

Rounds of the survey	Sampling period (number of samples collected)	
	New York City metro area	Connecticut
Round 1	March 23 - April 1 (n=2482)	April 26 - May 3 (n=1431)
Round 2	April 25 - May 6 (n=1116)	May 21- May 26 (n=1800)
Round 3	June 15 - June 21 (n=1581)	June 15 - June 17 (n=1798)
Round 4	July 7 - July 11 (n=1602)	July 3 - July 6 (n=1802)
Round 5	NA*	July 30 - August 3 (n=992)

*From August, the CDC commercial laboratory survey was expanded to cover other parts of New York State. Results were reported at state level, and data specific for the New York City metro area was not available. We therefore did not include the Round 5 data in this study. Abbreviation: CDC, Centers for Disease Control and Prevention.

Table S2. Results of the main analysis and sensitivity analysis.

Site	Analysis	Standard deviation for time from seroconversion to seroreversion (Fixed)	Estimated average months from seroconversion to seroreversion (95% CrI)	Estimated IFR % (95% CrI)	Estimated highest daily seroprevalence % (95% CrI)	Estimated daily seroprevalence on September 30, 2020 % (95% CrI)	Estimated cumulative incidence on September 30, 2020 % (95% CrI)
Connecticut	Sensitivity analysis with a smaller SD	20 days	3.0 (2.6-3.5)	1.5 (1.3-1.8)	6.2 (5.6-6.9)	0.4 (0.3-0.6)	7.8 (6.8-9.1)
Connecticut	Main analysis	50 days	3.0 (2.3-4.1)	1.4 (1.1-1.7)	6.1 (5.3-7.1)	1.3 (0.9-2.0)	8.8 (7.1-11.3)
Connecticut	Sensitivity analysis with a larger SD	70 days	3.0 (2.0-4.7)	1.2 (0.8-1.7)	6.1 (5.2-7.4)	1.8 (1.4-2.9)	9.8 (7.0-14.5)
New York City	Sensitivity analysis with a smaller SD	20 days	3.6 (3.3-4.1)	1.1 (1.0-1.2)	22.2 (20.5-23.8)	1.2 (0.9-2.1)	25.7 (23.4-28.2)
New York City	Main analysis	50 days	4.0 (3.6-4.6)	1.1 (1.0-1.2)	22.1 (20.5-23.6)	4.9 (3.9-6.9)	26.8 (24.2-29.7)
New York City	Sensitivity analysis with a larger SD	70 days	4.4 (3.8-5.2)	1.0 (0.9-1.2)	21.9 (20.4-23.3)	7.4 (6.1-9.7)	27.3 (24.3-30.8)
New York City	Sensitivity analysis with a decreasing IFR	50 days	4.0 (3.5-4.8)	1.4 (1.2-1.6)*	22.1 (20.3-23.6)	4.9 (3.8-7.7)	26.9 (23.7-29.9)
New York City	Sensitivity analysis with the Wuhan data for delay between onset and death	50 days	3.7 (3.3-4.1)	1.0 (0.9-1.1)	23.3 (21.8-24.9)	4.2 (3.5-5.4)	27.8 (25.2-30.6)

*Average IFR between March and September 2020. Abbreviations: IFR, infection fatality ratio; SD, standard deviation.
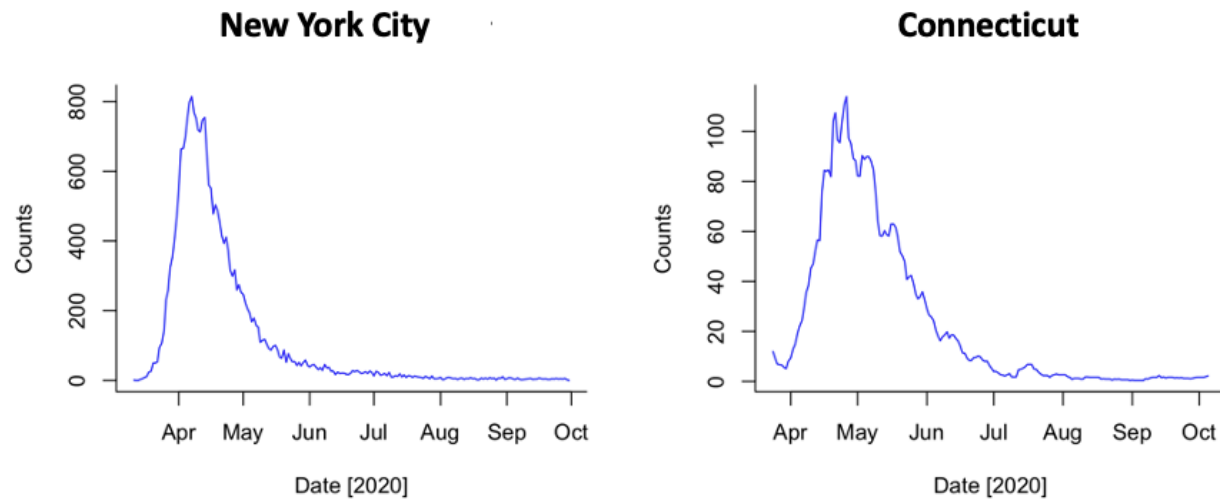
Supplementary figures

Figure S1. The number of reported deaths in New York City and Connecticut in 2020.
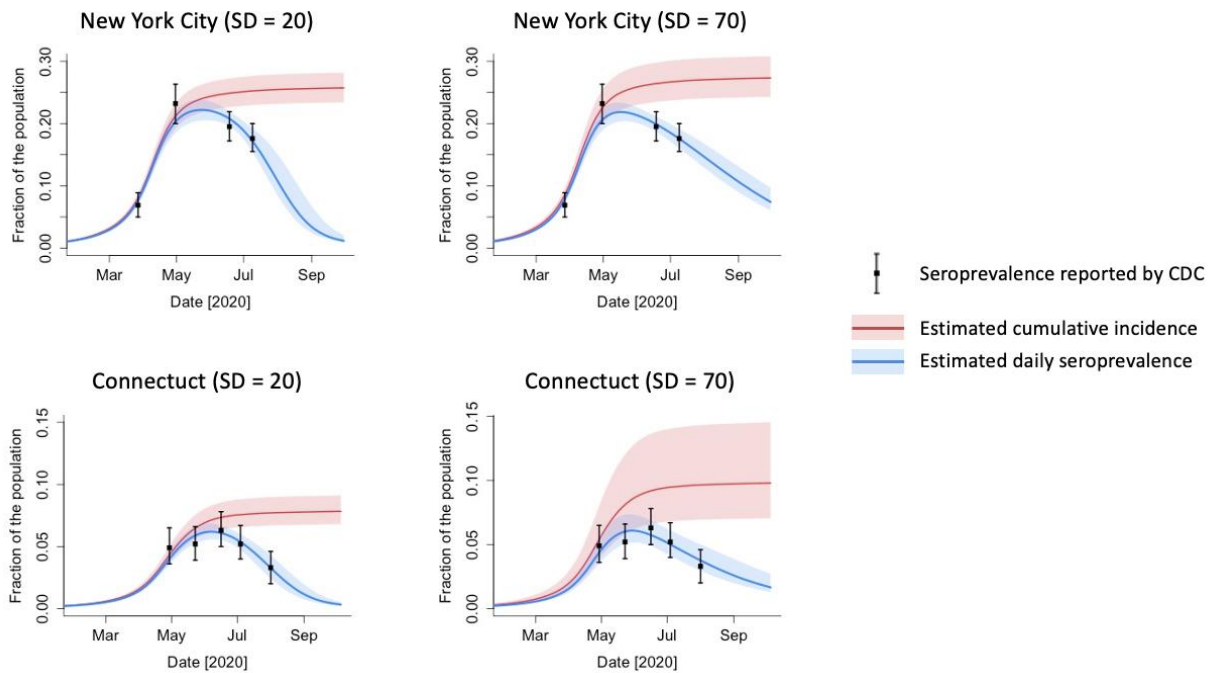A seven-day rolling average is shown for Connecticut.

Figure S2. Estimated daily seroprevalence and cumulative incidence of SARS-CoV-2 infection in New York City and Connecticut in 2020 with different values of the standard deviation for time from seroconversion to seroreversion.
Black dots and bars represent are the point estimates and 95% confidence intervals for the seroprevalence reported by CDC.[12] Abbreviations: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SD, standard deviation; CDC, Centers for Disease Control and Prevention.
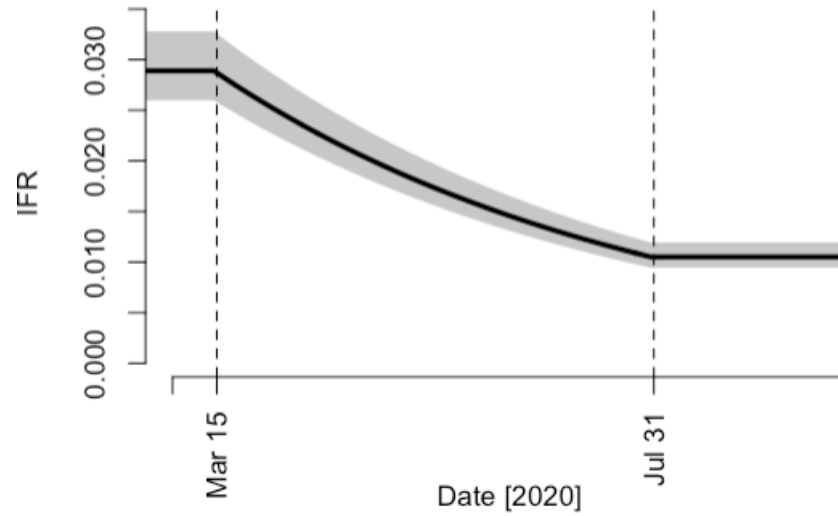
Figure S3. Estimated infection fatality ratio in New York City under the assumption that it decreased by 5% per week from March 15 to July 31 in 2020.
A line represents the median of posterior samples and a shaded area represents the 95% credible interval. Abbreviation: IFR, infection fatality ratio.
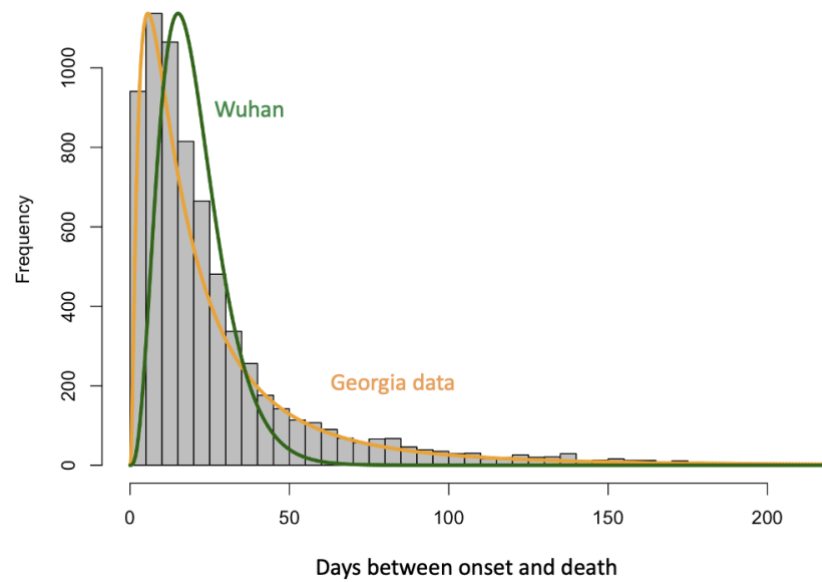
Figure S4. Time from symptom onset to death for COVID-19 fatal cases in Georgia, USA and Wuhan, China.
The histogram represents the raw data on time from onset to death in Georgia, and an orange line is its best fitted distribution. A green line represents the best fitted distribution to the data from Wuhan, China.[22] The original data from Wuhan, China can be found in the previous study and is not reproduced here.[22] Abbreviations: COVID-19: coronavirus disease 2019.