Supplementary materials for:

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

Supplementary methods

eAppendix 1. 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.

eAppendix 2. 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 in the main text), 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.

eAppendix 3. 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 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.[1] Among those who died on day t (d_t), we calculated the number of individuals who developed symptoms on day t-m (o_{t-m}) as follows:

$$d_t = \sum_{m=1}^{\infty} o_{t-m} \times (F_m - F_{m-1}), 1 \leq m < \infty$$

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.

eAppendix 4. 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.

We compared the estimated P_t (daily seroprevalence as defined in the "Model" subsection of the Methods in the main text) with the reported seroprevalence in each round of the CDC commercial laboratory serosurvey. We calculated the log-likelihood assuming the binomial distribution with size (i.e., number of total sera samples tested in each round of the serosurvey), number of observations (i.e., number of seropositive samples in each round of the serosurvey), and probability (i.e., estimated seroprevalence on the median date of each round of serosurvey).

References:

1. Delignette-Muller ML, Dutang C. fitdistrplus: Package for Fitting Distributions. *J Stat Soft*. 2015;64(4). doi:10.18637/jss.v064.i04
2. Wu JT, Leung K, Bushman M, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med*. 2020;26(4):506-510. doi:10.1038/s41591-020-0822-7

Supplementary tables

eTable 1. Date of the CDC longitudinal commercial laboratory seroprevalence surveys in New York City metro area and Connecticut in 2020.

|	Sampling period (number of samples collected)	
Rounds of the survey	**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.

eTable 2. Results of the main analysis and sensitivity analysis.

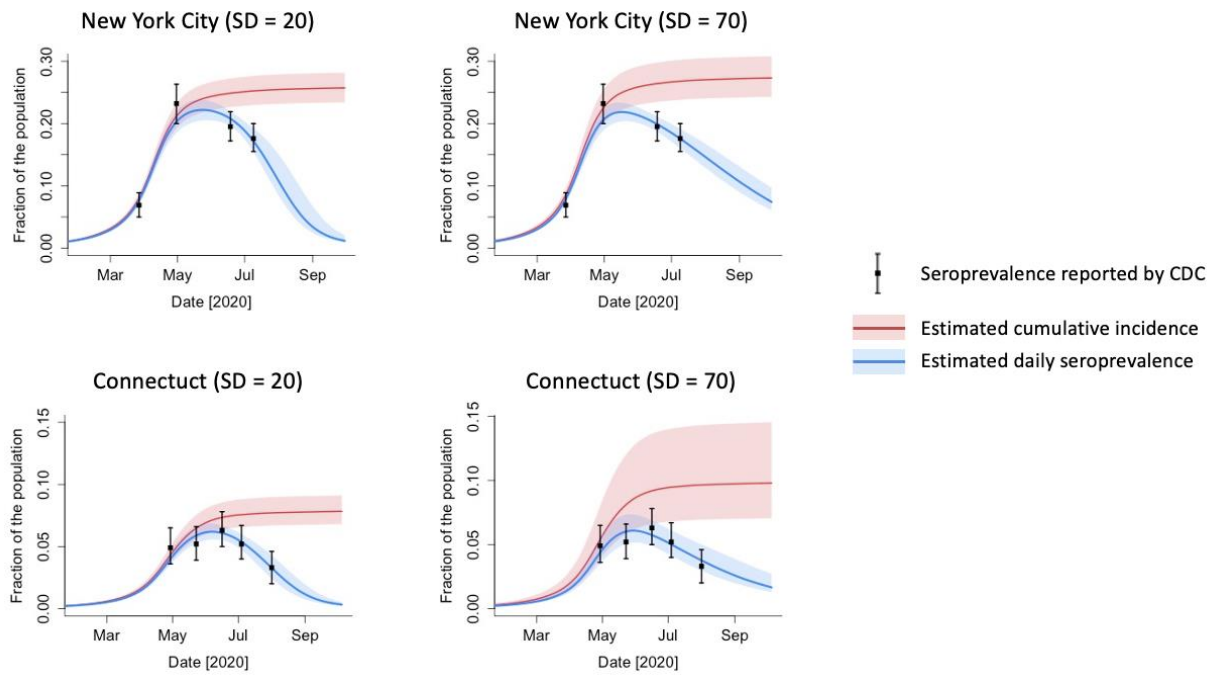| Site	Analysis	Standard deviation for time from seroconversion to seroreversion (Fixed)	Estimated average months from seroconversion to seroreversion (95% CrI)	Estimated IFR % (95% CrI)	Estimated highest daily seroprevalence % (95% CrI)	Estimated daily seroprevalence on September 30, 2020 % (95% CrI)	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
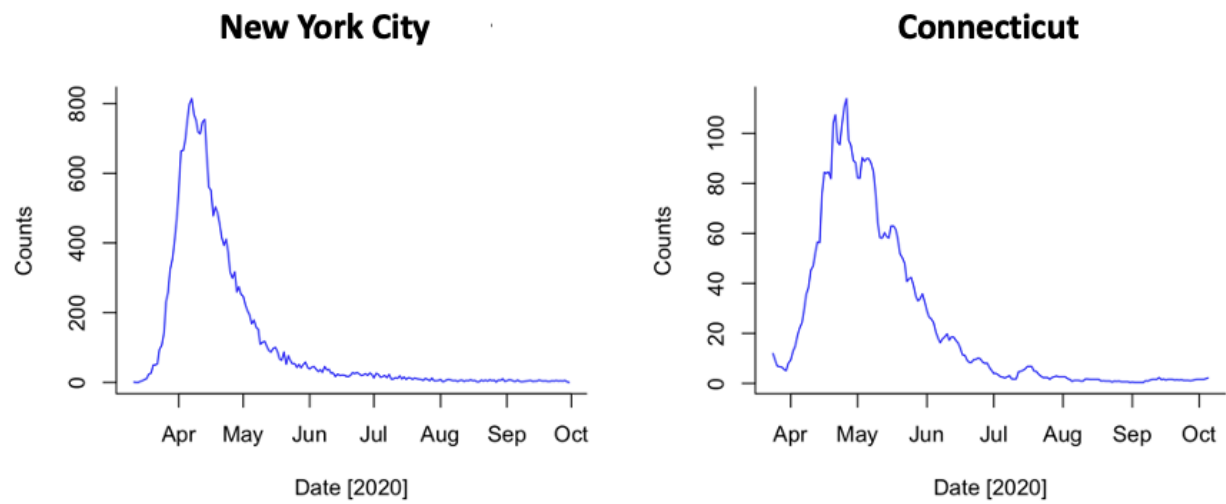
eFigure 1. The number of reported deaths in New York City and Connecticut in 2020.
A seven-day rolling average is shown for Connecticut.

eFigure 2. 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. Abbreviations: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SD, standard deviation; CDC, Centers for Disease Control and Prevention.
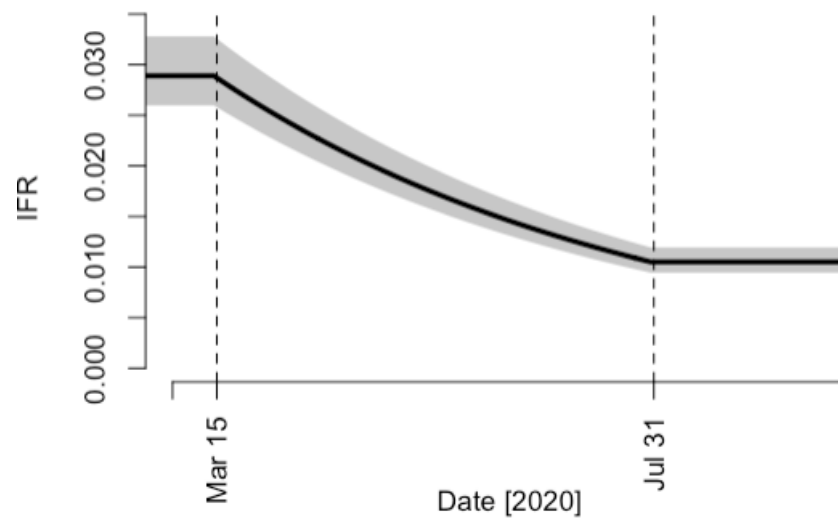
eFigure 3. Estimated infection fatality ratio in New York City under the assumption that it decreased by 5% per week from March 15 to July 31 in 2020.

A line represents the median of posterior samples and a shaded area represents the 95% credible interval.

Abbreviation: IFR, infection fatality ratio.
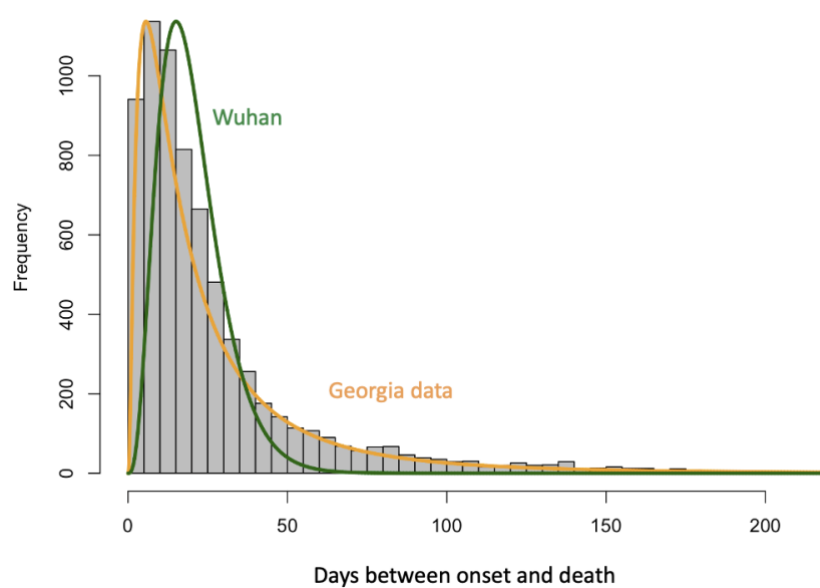
eFigure 4. Time from symptom onset to death for COVID-19 fatal cases in Georgia, USA and Wuhan, China.
The histogram represents the raw data on time from onset to death in Georgia, and an orange line is its best fitted distribution. A green line represents the best fitted distribution to the data from Wuhan, China.[2] The original data from Wuhan, China can be found in the previous study and is not reproduced here.[2] Abbreviations: COVID-19: coronavirus disease 2019.