

## Cumulative incidence and diagnosis of SARS-CoV-2 infection in New York

Eli S. Rosenberg PhD<sup>1</sup>, James M. Tesoriero PhD<sup>2</sup>, Elizabeth M. Rosenthal MPH<sup>1</sup>, Rakkoo Chung PhD<sup>2</sup>, Meredith A. Barranco MPH<sup>1</sup>, Linda M. Styer PhD<sup>3</sup>, Monica M. Parker PhD<sup>3</sup>, Shu-Yin John Leung MA<sup>2</sup>, Johanne E. Morne MS<sup>2</sup>, Danielle Greene DrPH<sup>2</sup>, David R. Holtgrave PhD<sup>1</sup>, Dina Hoefler PhD<sup>2</sup>, Jessica Kumar DO<sup>2</sup>, Tomoko Udo PhD<sup>1</sup>, Brad Hutton MPH<sup>2</sup>, Howard A. Zucker, MD<sup>2</sup>

1. University at Albany School of Public Health, State University of New York, Rensselaer NY
2. New York State Department of Health, Albany NY
3. Wadsworth Center, New York State Department of Health, Albany NY

### Corresponding Author:

Eli Rosenberg

Department of Epidemiology and Biostatistics,  
University at Albany School of Public Health  
1 University Place, Room 123,  
Rensselaer, NY 12144  
erosenberg2@albany.edu

### ABSTRACT

**Importance:** New York State (NYS) is an epicenter of the United States' COVID-19 epidemic. Reliable estimates of cumulative incidence of SARS-CoV-2 infection in the population are critical to tracking the extent of transmission and informing policies, but US data are lacking, in part because societal closure complicates study conduct.

**Objective:** To estimate the cumulative incidence of SARS-CoV-2 infection and percent of infections diagnosed in New York State, overall and by region, age, sex, and race and ethnicity.

**Design:** Statewide cross-sectional seroprevalence study, conducted April 19-28, 2020.

**Setting:** Grocery stores (n=99) located in 26 counties throughout NYS, which were essential businesses that remained open during a period of societal closure and attract a heterogeneous clientele.

**Participants:** Convenience sample of patrons ≥18 years and residing in New York State, recruited consecutively upon entering stores and via an in-store flyer.

**Exposures:** Region (New York City, Westchester/Rockland, Long Island, Rest of New York State), age, sex, race and ethnicity.

**Main Outcomes:** Primary outcome: cumulative incidence of SARS-CoV-2 infection, based on dry-blood spot (DBS) SARS-CoV-2 antibody reactivity; secondary outcome: percent of infections diagnosed.

**Results:** Among 15,101 adults with suitable DBS specimens, 1,887 (12.5%) were reactive using a validated SARS-CoV-2 IgG microsphere immunoassay (sensitivity 87.9%, specificity 99.75%). Following post-stratification weighting on region, sex, age, and race and ethnicity and adjustment for assay characteristics, estimated cumulative incidence through March 29 was 14.0% (95% CI: 13.3-14.7%), corresponding to 2,139,300 (95% CI: 2,035,800-2,242,800) infection-experienced adults. Cumulative incidence was higher among Hispanic/Latino (29.2%, 95% CI: 27.2-31.2%), non-Hispanic black/African American (20.2% 95% CI, 18.1-22.3%), and non-Hispanic Asian (12.4%, 95% CI: 9.4-15.4%) adults than non-Hispanic white adults (8.1%, 95% CI: 7.4-8.7%, p<.0001). Cumulative incidence was highest in New York City (NYC) 22.7% (95% CI: 21.5%-24.0). Dividing diagnoses reported to NYS by estimated infection-experienced adults, an estimated 8.9% (95% CI: 8.4-9.3%) of infections were diagnosed, with those ≥55 years most likely to be diagnosed (11.3%, 95% CI: 10.4-12.2%).

**Conclusions and Relevance:** Over 2 million adults were infected through late March 2020, with substantial variations by subpopulations. As this remains below herd immunity thresholds, monitoring, testing, and contact tracing remain essential public health strategies.

## Introduction

The first cases of COVID-19 were identified in New York State (NYS) in early March, 2020 and since then NYS, particularly the metropolitan New York City (NYC) area, has become one of the most-impacted communities in the United States.<sup>1,2</sup> As of May 9, 2020, over 335,000 laboratory-confirmed diagnoses have been made, accounting for approximately 25% of US diagnoses.<sup>2,3</sup> As with most infections, lab-confirmed diagnoses undercount the true population-level burden of infections; with SARS-CoV-2, the virus that causes COVID-19, key factors that contribute to underdiagnosis include absent or mild symptoms and access to testing.<sup>4</sup> Thus although NYS has tested more residents for COVID-19 than any other state (over 1,180,000 persons tested through May 9, 2020), it is likely that laboratory-confirmed cases represent a relatively small portion of the total number of persons with a history of infection in NYS.<sup>3</sup>

Estimates of COVID-19 cumulative incidence (i.e. prevalence of previous or current infection) can inform the extent of epidemic spread as well as the number of persons still susceptible and progress towards herd immunity, which are critical for parameterizing simulation models and informing policies, including those for altering societal restrictions.<sup>5</sup> Furthermore, such data provide needed denominators for understanding the extent of diagnosis and rates of hospitalization, morbidity, and mortality, and geographic differences.

Antibody testing for SARS-CoV-2 has emerged as an important tool for understanding infection history. Although a several-week window period for development of IgG antibodies and evidence that not all persons with infection develop an antibody response limits their utility for diagnostics, and their interpretation for short- and long-term immunity remain uncertain, as with other infections, antibody prevalence serostudies with validated assays can assess population-level cumulative incidence in the recent past.<sup>6-10</sup>

Antibody serostudies for SARS-CoV-2 are being conducted in other countries and in the US are occurring on the national and county levels, but none have been conducted at the state level and no findings have been peer-reviewed.<sup>11,12</sup> The current array of recommendations against individual movement and business operation during the pandemic complicates study specimen collection. A recent RNA survey in Iceland and serosurvey in Santa Clara County, California, conducted sampling at centralized testing sites, which offer ease of execution particularly in small geographies, with potentially large self-selection biases.<sup>11,13</sup> Alternative approaches include random at-home mail-in testing and community-intercept studies in high-traffic locations that remain open.<sup>12</sup>

To provide a statewide picture of COVID-19 infection through late-March and diagnoses by early-April 2020, during April 19-28 2020, the NYS Department of Health (NYSDOH) conducted a community-based serostudy throughout NYS. Cumulative incidence by geographic and demographic features was estimated from weighted reactivity rates that were adjusted for validated test characteristics. Combining these findings with cumulative diagnoses enabled estimation of the percent of infections diagnosed.

## **Methods**

### Field study

The NYS DOH conducted a convenience sample of over 15,000 New Yorkers attending 99 grocery stores across 26 counties, which contain 87.3% of the state's population, located in all regions of NYS (Figure). Grocery stores were chosen as the testing venue because they were classified an essential business to remain open and, due to the necessity of grocery shopping, they attract a heterogeneous clientele.<sup>14</sup> Store locations were chosen to increase sample coverage of the racial and ethnic diversity of the statewide population.

Testing occurred over 6 distinct days from 4/19/2020 through 4/28/2020. Each store had a team of 6-8 staff responsible for recruiting participants, collecting specimens, recording data, and managing specimen transport to Wadsworth Center Laboratory (Albany, NY) for analysis. Eligible subjects were adults  $\geq 18$  years, New York residents irrespective of county, recruited through a recruitment flyer posted at stores and by systematically approaching each patron as they entered the store. Patrons were given information about the testing and if interested, completed written informed consent. Procedures included a brief demographic questionnaire and dried blood spot (DBS) collection by trained personnel. Approximately 13% of participants initially had missing demographic data. Staff attempted to capture these data through >2,500 follow-up phone calls, reaching all but approximately 75 participants, who were subsequently excluded from analyses. Test results were delivered to participants by text-message if non-reactive and by phone if indeterminate or reactive.

### Testing approach

Blood was collected by fingerstick onto custom 903 filter paper cards labeled with a specimen ID. Cards were dried for 3-4 hours at ambient temperature and transported to the Wadsworth Center. A fully saturated  $\geq 3$ -mm diameter DBS was required. A total of 525 DBS cards from eligible individuals were rejected; 433 with insufficient or improperly collected blood, 92 with no specimen ID. Acceptable DBS cards were processed for testing.

SARS-CoV IgG testing was conducted using a microsphere immunoassay (MIA) developed and validated for DBS by the NYSDOH Wadsworth Center. Briefly, nucleocapsid (N) antigen-coupled magnetic beads were incubated with blood eluted from a 3-mm DBS punch. Phycoerythrin-labeled goat anti-human IgG secondary antibody was used to detect microsphere-bound IgG antibodies and median fluorescence intensity (MFI) was determined using a FlexMap 3D (Luminex Corp., Austin, TX). The mean MFI of 90-100 negative DBS was used to set cut-offs; results greater than the mean MFI plus 6 standard

deviations (SD) were reported as reactive; results less than the mean MFI + 3 SD were nonreactive and results between mean MFI +3 to +6 were indeterminate. Serosurvey testing was initiated with SARS-CoV IgG v1 which used SARS-CoV-1 N antigen (Wadsworth Center, Albany, NY), and was completed using SARS-CoV IgG v2 which used SARS CoV-2 N protein (Sino Biological, Wayne, PA) after validation studies confirmed comparable performance.

Assay validation studies are described in eTables 1-5 and eFigure in the Supplement. Based on testing DBS collected prior to December 2019, specificity was estimated between 99.5% (95% CI: 98.5-100%) to 100% (95% CI: 96.1-100%). Serum collected from individuals diagnosed with non-COVID-19 respiratory and non-respiratory agents were tested to assess cross-reactivity; only 1 of 85 samples was reactive. Of 232 SARS CoV-2 PCR-positive DBS collected a median 35 days post-symptom onset, 204 (87.9%, 95% CI: [83.7-92.1%]) were reactive, informing sensitivity and thus incorporating both test performance and the proportion of infected persons who never develop IgG.<sup>6,7</sup>

## Analysis

We estimated SARS-Cov-2 cumulative incidence from observed antibody reactivity using two sequential steps: 1) post-stratification weighting to standardize to the New York State population and 2) adjustment by estimated antibody test characteristics.

Using the National Center for Health Statistics bridged-race file, weights were assigned to each participant based on their membership in each of 160 strata of sex, race and ethnicity (Hispanic, non-Hispanic white, non-Hispanic black, non-Hispanic Asian, and non-Hispanic other), age (18-34, 35-44, 45-54, ≥55 years), and residential region (New York City, Westchester/Rockland, Long Island, Rest of State [ROS]).<sup>15</sup> Post-stratification weights were defined as the proportion each stratum is represented in the state's population divided by the analogous proportion in the sample.<sup>16,17</sup> Next, we computed weighted frequencies for the percent reactive statewide, with one-way stratifications by sex, race and ethnicity,

age group, and region, and two-way stratifications within levels of region, including 95% confidence intervals (CI), with differences assessed using Rao-Scott  $\chi^2$  tests.<sup>18</sup> Indeterminate results were assumed non-reactive and statistical procedures were two-sided at  $\alpha=0.05$ .

In the second step, weighted reactivity estimates ( $p_{reactive}$ ) and their 95% CI bounds were corrected for test sensitivity and specificity, based on validation data, to yield cumulative incidence, per Bayes' Rule as applied to the diagnostic 2x2 table:  $cumulative\ incidence = \frac{p_{reactive} + specificity - 1}{sensitivity + specificity - 1}$ .<sup>11,19</sup> Primary analyses used the sensitivity and specificity point-estimates from the validation studies, with sensitivity analyses at the extremes of test characteristics' 95% CI ([96.1% specificity, 92.1% sensitivity], [100% specificity, 83.7% sensitivity]). Test-characteristic adjusted cumulative incidence values were multiplied by the one- and two-way non-institutionalized adult populations from the American Community Survey 2014-2018 Public Use Microdata Sample file.<sup>20</sup> This yielded the estimated total 'infection-experienced' adults with SARS-CoV-2 within each stratum, accounting for excluded settings such as prisons and nursing homes. With a study mid-point of April 23, and literature estimates of mean 4 days from infection to symptom onset and mean 21 days from onset to IgG detection, results represent cumulative incidence through approximately March 29.<sup>6,8,21</sup>

In NYS, diagnostic testing for SARS-CoV-2 is mandatorily reported electronically to NYSDOH. Using cumulative diagnoses reported and total numbers of infection-experienced adults, we estimated the percent of infections diagnosed overall and by region, sex, and age. For primary analyses, we accumulated diagnoses through April 9, based on the March 29 final infection date, 4 days to symptom onset, and mean 7 days from onset to diagnosis. Supplemental upper-bound estimates used the last plausible diagnosis date of May 8<sup>th</sup>, based on the April 28 final study day, 4 days being earliest time from onset to IgG detection and allowing PCR detection up to 14 days post-onset.<sup>8</sup>

## Results

Across NYS, a total of 15,626 adult residents with complete data were tested, of whom 15,101 (96.6%) had suitable specimens, of which 1,887 (12.5%) were reactive and 340 (2.3%) indeterminate. Following weighting, 12.5% were estimated reactive and following further adjustment for test characteristics, estimated cumulative incidence was 14.0% (95% CI: 13.3-14.7%), corresponding to 2,139,300 (95% CI: 2,035,800-2,242,800) infection-experienced adults in NYS through approximately March 29 (Table 1). In sensitivity analyses at the extremes of test characteristics, cumulative incidence ranged from 9.8% (95% CI: [9.1-10.5%]) to 15.0% (95% CI: [14.3-15.7%]), representing a total of 1,494,700 (95% CI: [1,391,800-1,597,600]) to 2,286,600 (95% CI: [2,178,200-2,395,100]) adults in NYS (eTables 6-7).

Cumulative incidence was higher among males (14.8%, 95% CI: [13.8-15.8%]) than females (13.3%, 95% CI: [12.4-14.2%],  $p=0.03$ ), with males comprising 50.3% of adult infections. This differed significantly by race and ethnicity, with Hispanic/Latino (29.2%, 95% CI: 27.2-31.2%), non-Hispanic black/African American (20.2% [95% CI, 18.1-22.3%]), and non-Hispanic Asian (12.4%, 95% CI: [9.4-15.4%]) adults having higher cumulative incidence than non-Hispanic white adults (8.1%, 95% CI: [7.4-8.7%],  $p<.0001$ ). Given these differences, Hispanics comprised the plurality (36.6%) of infection-experienced adults. Significant differences were also observed by age ( $p=0.0002$ ), ranging from highest levels among persons 45-54 years old (16.0%, 95% CI: [14.6-17.5%]) to lowest among persons  $\geq 55$  years (12.1% [95% CI: 11.2-13.1%]).

We observed regional heterogeneity in cumulative incidence, ranging from 22.7% (95% CI: 21.5%-24.0%) in NYC residents, to 16.1% (95% CI: 13.2-19.0%) and 13.2% (11.4-15.1%) in the respective metropolitan areas of Westchester/Rockland Counties and Long Island, to 3.6% (95% CI: [3.0-4.1]) in ROS ( $p<0.0001$ ). Demographic patterns were heterogeneous by region (Table 2). Males had significantly higher cumulative incidence in all regions outside of, but not within NYC. The patterns of racial disparity

observed statewide were similar and statistically significant within NYC, Westchester/Rockland, and Long Island, but not in ROS. In each of the former 3 regions, Hispanic/Latino persons represented >37% of infection-experienced adults, whereas in the latter non-Hispanic whites comprised a majority of infection-experienced adults (79.4%).

An estimated 8.9% (95% CI: 8.4-9.3%) of infections in NYS were diagnosed as of April 9<sup>th</sup> 2020 (Table 3). Males (9.4%, 95% CI: 8.8-10.1%) had higher diagnosis levels than females (8.2%, 95% CI: 7.7-8.8%). Those  $\geq 55$  years were most likely to be diagnosed (11.3%, 95% CI: 10.4-12.2%). Diagnosis rates in NYC (7.1%, 95% CI: 6.7-7.5%) and ROS (7.5%, 95% CI: 6.4-8.9%) were about half those observed in the other regions. Considering the May 8 upper-bound for diagnoses, a maximum of 15.7% (95% CI: 15.0-16.5%) of overall infections could have been diagnosed, with similar patterns observed across levels of each factor (eTable 8).

## Discussion

From the largest US SARS-CoV-2 serosurvey to-date, we estimated that over 2 million adult NYS residents were infected through the end of March. Our findings estimate the extent of transmission of and community experience with SARS-CoV-2, particularly in the NYC metropolitan region. Despite large numbers of persons acquiring SARS-CoV-2, this represents only 14.0% of adult residents, suggesting that, even in this COVID-19 epicenter, the epidemic is substantially below the ~70% US herd immunity threshold.<sup>22</sup> Against this remaining epidemic potential, ongoing vigilance through rigorous and extensive epidemic monitoring, testing, and contact tracing are necessary components for predicting, preventing, and/or mitigating a second epidemic wave, consistent with state and federal guidance for reopening.<sup>5,23</sup> This vigilance is needed even in the rest of NYS outside the metropolitan region, portions of which are



beginning the first phases of reopening in NYS, and where lowest cumulative incidence suggests the highest proportion susceptible.

Our finding of higher cumulative incidence in the regions of the NYC metropolitan area, particularly NYC, is consistent with the known distribution of diagnoses. Further, in these regions of high urbanicity, significant racial/ethnic disparities in infection history were found, with minority communities experiencing disproportionate risk. The drivers of greater COVID-19 risk and disparities in urban areas continue to be studied, but may relate to population density and the mechanisms by which transportation, employment, housing, and other socioeconomic or environmental factors shape opportunities for transmission.<sup>24-26</sup> A recent NYS study on a random sample of COVID-19 hospitalizations showed limited racial/ethnic differences in clinical outcomes, suggesting that observed differences in mortality by race and ethnicity are likely driven by the different infection histories reported here.<sup>3,27</sup> Research is needed to understand the drivers of increased COVID-19 risk experienced by minority communities, followed by actions to improve health equity.

The finding that over 8.9% of adults were diagnosed reveals both the opportunities for further expansion of diagnostic testing in NYS, yet in context of far higher diagnosis and testing levels than other US settings suggests substantial progress to-date.<sup>1,11</sup> Compared to all persons with infection history, there was a higher representation of males and those over age 55 among diagnosed persons. Given the lower reactivity rates observed among this age group, our results expand observations from previous studies that older adults may be more likely to exhibit symptoms or illness or be more likely to seek care.<sup>27-30</sup>

Strengths of our study include a large sample, which contained 0.1% of the adult NYS population, and a systematic sampling approach in one of the only open public venues in the state, where a necessary commodity is purchased. Although a convenience sample, survey weights adjusted for biased demographic/geographic representation, and we further adjusted results for assay

performance, under varied scenarios. Our study may nevertheless be limited by residual non-representativeness of the underlying population. This includes potential undersampling of persons from vulnerable groups who might be less likely to go grocery shopping. For this to impact our findings, those remaining home would need to have differential antibody prevalence compared to their age/sex/racial-ethnic/regional group peers. If persons staying at home had lower prevalence due to self-isolation, our study's cumulative incidence would be a slight overestimate. Further, our sample did not include those who have died from COVID-19 or those who reside in long term care facilities which have been differentially impacted, causing a slight underestimate, nor those in the hospital or at-home due to COVID-19 illness, some of whom would be expected to have detectable antibodies.<sup>31,32</sup> Such actively symptomatic persons would be expected to be a small portion of the cumulative infection burden since the outbreak's commencement, and given most would have been infected after March 29, their exclusion also likely causes observed values to be overestimated.

Although data are limited on the potential for self-selection to alter our results, a recent Icelandic study found comparable prevalence when participants were tested following online self-registration vs. random invitation.<sup>13</sup> This finding, in conjunction with our systematic community intercept approach, suggest that this bias may be small, outside of outright non-response. Although data were not captured, field workers reported very-high willingness to participate due to the public concern over COVID-19, supporting low non-response. Results presented may differ from publicly discussed preliminary estimates, given both our inclusion of more participants and analytic adjustments for test characteristics. Timeframes utilized for cumulative infections and diagnoses are approximate, being based on the evolving SARS-CoV-2 immunological and testing literature, with the 10-day sampling period during a linear-growth phase of the epidemic.

The findings of this study suggest extensive SARS-CoV-2 transmission in NYS and highlight the remaining opportunities for prevention and diagnosis. As the epidemic grows in other regions of the

country, this study offers a potential model for other jurisdictions to monitor their epidemic. Estimates of cumulative incidence can be combined with diagnostic totals, or other epidemic markers such as mortality, to provide a holistic epidemic view during a time of unprecedented pandemic and to best craft high-impact approaches to prevention, containment, treatment and mitigation.

**Acknowledgements:**

The co-authors wish to acknowledge the following contributors to this work. Data preparation team members Janson Ganns, Peter Cichetti, Melissa Kamal, Alison Pingelski and Mary McCormick. Michelle Cummings for data management. Amy Kelly for literature review contributions. The New York State Antibody Sampling and Testing Team, including the Call Center team who placed thousands of phone calls to deliver test results and to collect missing demographic data. Office of Quality and Patient Safety team members James Kirkwood and Meng Wu. Eric Hall at Emory University for map assistance. We thank Dr. William Lee for assistance with assay development and validation, Jean Rock and the Wadsworth Center COVID-19 serology team especially Rachel Bievenue, Seth Blumerman, Theresa Hattenrath, Jim Long, Kate Mastraccio, Erica Miller, Katie Nemeth, and Alyssa Sossei, and numerous members of Wadsworth Center's Newborn Screening Program especially Beth Vogel, Michele Caggana and Rhonda Hamel.

**Table 1. Reactivity and test-characteristic adjusted cumulative incidence of COVID-19, overall and by demographic factors and region**

	Reactivity		Test-characteristic adjusted estimated cumulative incidence					p-value	
	Unweighted #				Infection-experienced		% of Infection-experienced		
	Reactive	Total	Weighted	%	(95% CI)	adults <sup>a</sup>	(95% CI)		adults
	Sample	percent							
<b>Overall</b>	1,887/15,101	12.5		14.0 (13.3-14.7)		2,139,300 (2,035,800-2,242,800)		100.0	
<b>Sex</b>									0.03
Male	918/6,635	13.2		14.8 (13.8-15.8)		1,076,500 (1,001,900-1,151,100)		50.3	
Female	969/8,466	11.9		13.3 (12.4-14.2)		1,062,200 (990,500-1,133,800)		49.7	
<b>Race and Ethnicity</b>									<.0001
Hispanic or Latino	757/2,735	25.8		29.2 (27.2-31.2)		775,800 (722,700-829,000)		36.6	
NH-White	623/9,545	7.3		8.1 (7.4-8.7)		715,400 (657,100-773,700)		33.7	
NH-Black/African American	388/1,913	18.0		20.2 (18.1-22.3)		428,000 (382,700-473,400)		20.2	
NH-Asian	75/629	11.1		12.4 (9.4-15.4)		161,700 (122,600-200,800)		7.6	
Multiracial/Other	44/279	10.7		11.9 (6.4-17.5)		38,800 (20,800-56,800)		1.8	
<b>Age group</b>									0.0002
18-34	377/3,151	13.0		14.6 (13.1-16.1)		682,600 (612,000-753,200)		31.8	
35-44	334/2,628	13.7		15.3 (13.7-17.0)		371,800 (331,700-411,900)		17.3	
45-54	479/3,345	14.3		16.0 (14.6-17.5)		424,700 (386,400-463,100)		19.8	
55+	697/5,977	10.9		12.1 (11.2-13.1)		667,800 (615,600-719,900)		31.1	
<b>Region</b>									<.0001
New York City <sup>b</sup>	1,319/5,946	20.2		22.7 (21.5-24.0)		1,504,400 (1,421,300-1,587,500)		70.1	
Westchester/ Rockland Counties	134/980	14.4		16.1 (13.2-19.0)		156,500 (128,400-184,600)		7.3	
Long Island <sup>c</sup>	241/2,074	11.9		13.2 (11.4-15.1)		291,800 (250,600-332,900)		13.6	
Rest of NYS <sup>d</sup>	193/6,101	3.4		3.6 (3.0-4.1)		194,600 (162,600-226,600)		9.1	

- Stratified estimates may not exactly sum to total due to rounding and differences between weighting scheme and non-institutionalized population totals
- Boroughs of Bronx, Brooklyn, Manhattan, Queens, Staten Island
- Nassau and Suffolk Counties
- Albany, Allegany, Broome, Cattaraugus, Cayuga, Chautauqua, Chemung, Chenango, Clinton, Columbia, Cortland, Delaware, Dutchess, Erie, Essex, Franklin, Fulton, Genessee, Greene, Hamilton, Herkimer, Jefferson, Lewis, Livingston, Madison, Monroe, Montgomery, Niagara, Oneida, Onondaga, Ontario, Orange, Orleans, Oswego, Otsego, Putnam, Rensselaer, St. Lawrence, Saratoga, Schenectady, Schoharie, Schuyler, Seneca, Steuben, Sullivan, Tioga, Tompkins, Ulster, Warren, Washington, Wayne, Wyoming, Yates counties

**Table 2. Reactivity and test-characteristic adjusted cumulative incidence of COVID-19, demographic factors within region**

Reactivity			Test-characteristic adjusted estimated cumulative incidence					p-value
Unweighted # Reactive / Total Sample <sup>a</sup>	Weighted percent		Infection- experienced adults <sup>b</sup>		% of Infection- experienced adults			
			%(95% CI)	(95% CI)				
<b>New York City <sup>c</sup></b>								
<b>Sex</b>								0.26
Male	629/2,727	20.9	23.5	(21.6-25.4)	726,300	(668,500-784,100)	48.3	
Female	690/3,219	19.6	22.1	(20.4-23.7)	778,000	(718,300-837,600)	51.7	
<b>Race and Ethnicity</b>								<.0001
Hispanic or Latino	624/2,103	29.2	33.0	(30.6-35.4)	599,900	(556,800-643,000)	39.8	
NH-White	264/1,758	14.8	16.6	(14.6-18.5)	371,300	(327,800-414,800)	24.6	
NH-Black/African American	329/1,392	22.4	25.2	(22.5-27.9)	361,700	(322,900-400,500)	24.0	
NH-Asian	68/509	13.0	14.5	(11.0-18.0)	139,000	(105,400-172,700)	9.2	
Multiracial/Other	34/184	18.2	20.4	(13.7-27.2)	34,400	(23,100-45,700)	2.3	
<b>Age group</b>								0.04
18-34	252/1,257	19.3	21.8	(19.2-24.4)	490,200	(432,200-548,300)	32.5	
35-44	243/1,144	20.8	23.4	(20.6-26.2)	270,400	(238,100-302,700)	18.0	
45-54	334/1,328	23.5	26.5	(23.8-29.2)	286,700	(257,700-315,700)	19.0	
55+	490/2,217	19.1	21.5	(19.6-23.5)	459,000	(417,700-500,300)	30.5	
<b>Westchester/Rockland</b>								
<b>Sex</b>								0.049
Male	72/450	17.1	19.2	(14.6-23.8)	88,700	(67,400-109,900)	56.7	
Female	62/530	11.9	13.3	(9.8-16.9)	67,800	(49,600-85,900)	43.3	
<b>Race and Ethnicity</b>								0.0008
Hispanic or Latino	37/141	25.3	28.6	(20.2-37.0)	58,300	(41,200-75,400)	37.8	
NH-White	62/654	9.7	10.8	(8.0-13.5)	60,400	(45,200-75,600)	39.1	
NH-Black/African American	32/152	20.1	22.7	(15.0-30.3)	29,200	(19,400-39,000)	18.9	
NH-Asian	**	7.7	8.5	(0-25.1)	5,200	(0-15,300)	3.4	
Multiracial/Other	**	7.1	7.8	(0-19.4)	1,200	(0-2,900)	0.8	
<b>Age group</b>								0.78
18-34	29/184	16.1	18	(11.8-24.3)	46,900	(30,600-63,200)	29.9	
35-44	22/156	15.2	17	(9.3-24.8)	27,000	(14,700-39,200)	17.2	
45-54	33/247	14.5	16.2	(10.7-21.8)	29,800	(19,600-40000)	19.0	
55+	50/393	12.9	14.4	(10.0-18.9)	53,100	(36,700-69,400)	33.9	
<b>Long Island <sup>d</sup></b>								
<b>Sex</b>								0.15
Male	122/913	13.1	14.7	(11.7-17.6)	156,100	(125,100-187,100)	53.5	
Female	119/1,161	10.7	11.9	(9.5-14.3)	135,700	(108,700-162,600)	46.5	
<b>Race and Ethnicity</b>								<.0001
Hispanic or Latino	89/301	28.3	32.0	(26.1-37.9)	112,100	(91,400-132,900)	38.4	
NH-White	126/1,599	7.9	8.7	(7.2-10.3)	129,500	(106,800-152,200)	44.4	
NH-Black/African American	16/111	14.1	15.8	(6.6-25.0)	30,800	(12,900-48,700)	10.6	
NH-Asian	**	7.7	8.4	(0-18.7)	11,800	(0-26,200)	4.0	
Multiracial/Other	**	18.4	20.7	(3.9-37.5)	7,600	(1,400-13,800)	2.6	
<b>Age group</b>								0.73
18-34	45/429	12.6	14.1	(9.2-19.0)	81,800	(53,500-110,000)	28.0	
35-44	40/317	12.4	13.8	(9.6-18.1)	47,300	(32,800-61,800)	16.2	

45-54	61/468	12.8	14.3	(10.5-18.0)	61,200	(45,100-77,300)	20.9	
55+	95/860	10.7	12.0	(9.5-14.4)	102,200	(81,500-122,900)	34.9	
<b>Rest of New York State<sup>e</sup></b>								
<b>Sex</b>								
Male	95/2,545	3.9	4.2	(3.2-5.1)	111,200	(85,800-136,600)	57.3	0.04
Female	98/3,556	2.8	2.9	(2.3-3.6)	83,100	(63,800-102,300)	42.7	
<b>Race and Ethnicity</b>								
Hispanic or Latino	**	4.4	4.7	(0.8-8.7)	13,600	(2,300-24,900)	7.0	0.90
NH-White	171/5,534	3.2	3.4	(2.8-3.9)	154,300	(128,800-179,800)	79.4	
NH-Black/African American	11/258	4.0	4.3	(1.4-7.1)	15,500	(5,100-25,800)	8.0	
NH-Asian	**	4.0	4.3	(0-9.4)	6,300	(0-13,900)	3.2	
Multiracial/Other	**	4.2	4.5	(0-13.6)	4,700	(0-14,300)	2.4	
<b>Age group</b>								
18-34	51/1,281	4.1	4.3	(3.0-5.7)	69,100	(47,500-90,700)	35.2	0.04
35-44	29/1,011	3.5	3.8	(2.1-5.4)	29,100	(16,300-41,900)	14.8	
45-54	51/1,302	4.3	4.6	(3.2-6.0)	43,900	(30,500-57,400)	22.4	
55+	62/2,507	2.5	2.5	(1.8-3.2)	54,000	(39,100-68,900)	27.5	

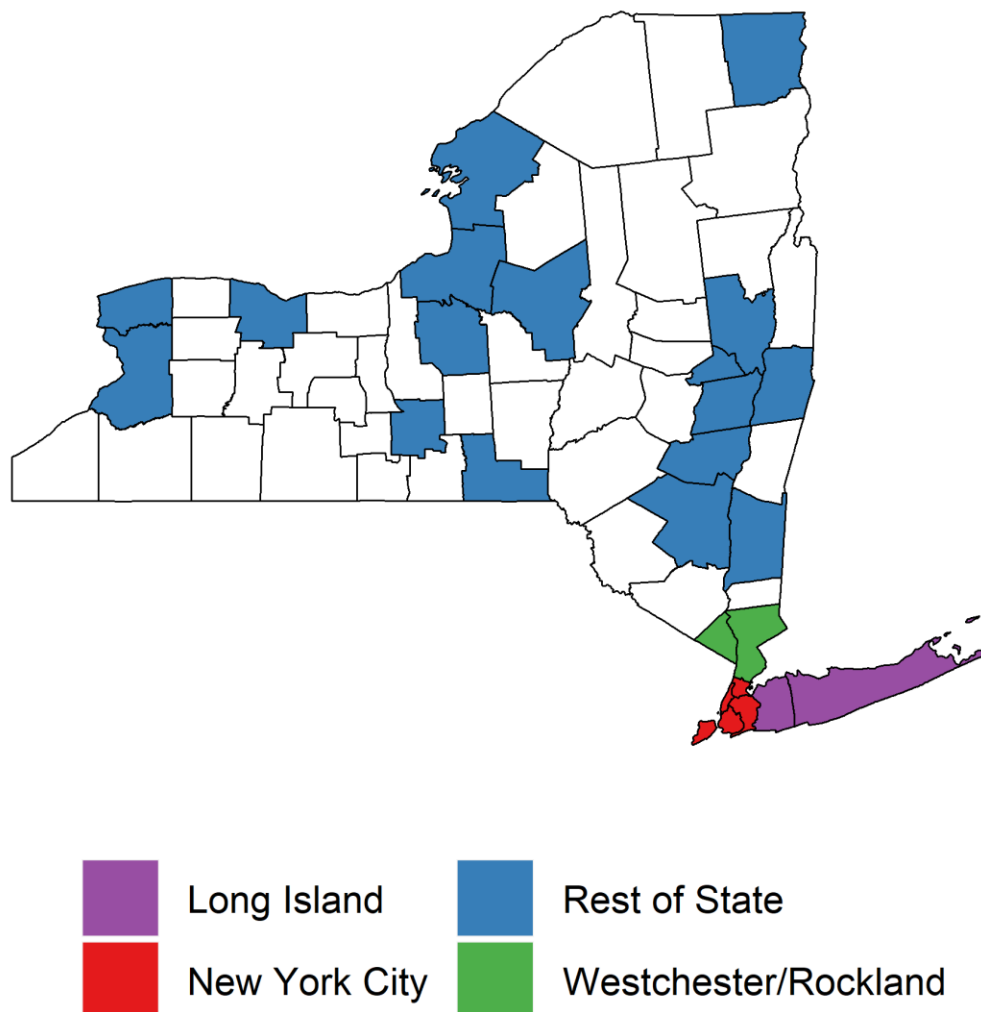
- Unweighted results with numerator < 10 are suppressed to protect participant confidentiality, indicated by \*\*
- Stratified estimates may not exactly sum to total due to rounding and differences between weighting scheme and non-institutionalized population totals
- Boroughs of Bronx, Brooklyn, Manhattan, Queens, Staten Island
- Nassau and Suffolk Counties
- Albany, Allegany, Broome, Cattaraugus, Cayuga, Chautauqua, Chemung, Chenango, Clinton, Columbia, Cortland, Delaware, Dutchess, Erie, Essex, Franklin, Fulton, Genessee, Greene, Hamilton, Herkimer, Jefferson, Lewis, Livingston, Madison, Monroe, Montgomery, Niagara, Oneida, Onondaga, Ontario, Orange, Orleans, Oswego, Otsego, Putnam, Rensselaer, St. Lawrence, Saratoga, Schenectady, Schoharie, Schuyler, Seneca, Steuben, Sullivan, Tioga, Tompkins, Ulster, Warren, Washington, Wayne, Wyoming, Yates counties

**Table 3. Estimated percentage of SARS-CoV-2 infections diagnosed<sup>a</sup>**

	Estimated infection-experienced adults		Diagnosed adults through April 9, 2020			
	Adults	(95% CI)	Diagnoses	% diagnosed	(95% CI)	% of diagnosed adults
<b>Overall</b>	2,139,300	(203,5800-2,242,800)	189,383	8.9	(8.4-9.3)	100.0
<b>Sex</b>						
Male	1,076,500	(1,001,900-1,151,100)	101,030	9.4	(8.8-10.1)	53.7
Female	1,062,200	(990,500-1,133,800)	87,196	8.2	(7.7-8.8)	46.3
Unknown			1,157			
<b>Age</b>						
18-34	682,600	(612,000-753,200)	41,335	6.1	(5.5-6.8)	22.4
35-44	371,800	(331,700-411,900)	32,845	8.8	(8.0-9.9)	17.8
45-54	424,700	(386,400-463,100)	35,307	8.3	(7.6-9.1)	19.1
55+	667,800	(615,600-719,900)	75,124	11.3	(10.4-12.2)	40.7
Missing/Invalid			491			
<b>Region</b>						
New York City <sup>b</sup>	1,504,400	(1,421,300-1,587,500)	106,401	7.1	(6.7-7.5)	56.2
Westchester/ Rockland Counties	156,500	(128,400-184,600)	23,557	15.1	(12.8-18.3)	12.4
Long Island <sup>c</sup>	291,800	(250,600-332,900)	44,907	15.4	(13.5-17.9)	23.7
Rest of State <sup>d</sup>	194,600	(162,600-226,600)	14,518	7.5	(6.4-8.9)	7.7

- a. Complete statewide data on case race and ethnicity not currently available
- b. Boroughs of Bronx, Brooklyn, Manhattan, Queens, Staten Island
- c. Nassau and Suffolk Counties
- d. Albany, Allegany, Broome, Cattaraugus, Cayuga, Chautauqua, Chemung, Chenango, Clinton, Columbia, Cortland, Delaware, Dutchess, Erie, Essex, Franklin, Fulton, Genessee, Greene, Hamilton, Herkimer, Jefferson, Lewis, Livingston, Madison, Monroe, Montgomery, Niagara, Oneida, Onondaga, Ontario, Orange, Orleans, Oswego, Otsego, Putnam, Rensselaer, St. Lawrence, Saratoga, Schenectady, Schoharie, Schuyler, Seneca, Steuben, Sullivan, Tioga, Tompkins, Ulster, Warren, Washington, Wayne, Wyoming, Yates counties

**Figure. New York State counties included in the New York State Department of Health Serological Testing Survey <sup>1</sup>**



1. Sampled counties - Long Island: Nassau, Suffolk; New York City: Boroughs of Bronx, Brooklyn, Manhattan, Queens, Staten Island; Westchester, Rockland Counties; Rest of State: Albany, Broome, Clinton, Dutchess, Erie, Greene, Jefferson, Monroe, Niagara, Oneida, Onondaga, Oswego, Rensselaer, Saratoga, Schenectady, Tompkins, Ulster



## References

1. Rosenberg ES, Dufort EM, Blog DS, et al. COVID-19 Testing, Epidemic Features, Hospital Outcomes, and Household Prevalence, New York State-March 2020. *Clin Infect Dis* 2020.
2. COVID-19 United States Cases 2020. (Accessed 5/13/2020, at <https://coronavirus.jhu.edu/us-map>.)
3. NYSDOH COVID-19 Tracker. 2020. at <https://covid19tracker.health.ny.gov/views/NYS-COVID19-Tracker/NYSDOHCOVID-19Tracker-DailyTracker?%3Aembed=yes&%3Atoolbar=no&%3Atabs=n>.)
4. Li R, Pei S, Chen B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* 2020;368:489-93.
5. NYS. Amid Ongoing COVID-19 Pandemic, Governor Cuomo Outlines Blueprint to Un-PAUSE New York. 2020.
6. Xiang F, Wang X, He X, et al. Antibody Detection and Dynamic Characteristics in Patients with COVID-19. *Clin Infect Dis* 2020.
7. Zhao J, Yuan Q, Wang H, et al. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. *Clin Infect Dis* 2020.
8. Sethuraman N, Jeremiah SS, Ryo A. Interpreting Diagnostic Tests for SARS-CoV-2. *JAMA* 2020.
9. Kirkcaldy RD, King BA, Brooks JT. COVID-19 and Postinfection Immunity: Limited Evidence, Many Remaining Questions. *JAMA* 2020.
10. Hofmeister MG, Rosenthal EM, Barker LK, et al. Estimating Prevalence of Hepatitis C Virus Infection in the United States, 2013-2016. *Hepatology* 2019;69:1020-31.
11. Bendavid E, Mulaney B, Sood N, et al. COVID-19 Antibody Seroprevalence in Santa Clara County, California. *medRxiv* 2020:2020.04.14.20062463.
12. NIH begins study to quantify undetected cases of coronavirus infection. 2020. (Accessed 5/13/2020, at <https://www.nih.gov/news-events/news-releases/nih-begins-study-quantify-undetected-cases-coronavirus-infection>.)
13. Gudbjartsson DF, Helgason A, Jonsson H, et al. Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J Med* 2020.
14. Executive Order [Cuomo] No. 202.6 [9 NYCRR 8.202.6]: New York State on PAUSE. 2020. at <https://www.governor.ny.gov/news/governor-cuomo-issues-guidance-essential-services-under-new-york-state-pause-executive-order>.)
15. U.S. Census Populations With Bridged Race Categories. 2020. (Accessed 5/13/2020, at [https://www.cdc.gov/nchs/nvss/bridged\\_race.htm](https://www.cdc.gov/nchs/nvss/bridged_race.htm).)
16. Holt D, Smith T. Post Stratification. *Journal of the Royal Statistical Society* 1979;142:33-46.

17. Korn EL, Graubard BI. Analysis of Health Surveys. New York: John Wiley & Sons; 1999.
18. Rao JNK, Scott AJ. On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data. *Ann Statist* 1984;12:46-60.
19. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
20. Public Use Microdata Sample (PUMS) Documentation. 2020. (Accessed 5/13/2020, at <https://www.census.gov/programs-surveys/acs/technical-documentation/pums.html>.)
21. Lee YL, Liao CH, Liu PY, et al. Dynamics of anti-SARS-Cov-2 IgM and IgG antibodies among COVID-19 patients. *J Infect* 2020.
22. Kwok KO, Lai F, Wei WI, Wong SYS, Tang JWT. Herd immunity - estimating the level required to halt the COVID-19 epidemics in affected countries. *J Infect* 2020.
23. Opening Up America Again. 2020. (Accessed 5/13/2020, at <https://www.whitehouse.gov/openingamerica/>.)
24. Rocklöv J, Sjödin H. High population densities catalyse the spread of COVID-19. *Journal of Travel Medicine* 2020.
25. amfAR. COVID-19 Racial Disparities in U.S. Counties. 2020.
26. Chowkwanyun M, Reed AL. Racial Health Disparities and Covid-19 — Caution and Context. *New England Journal of Medicine* 2020.
27. Rosenberg ES, Dufort EM, Udo T, et al. Association of Treatment With Hydroxychloroquine or Azithromycin With In-Hospital Mortality in Patients With COVID-19 in New York State. *JAMA* 2020.
28. Coronavirus Disease 2019 (COVID-19): Cases in U.S. 2020. at <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html>.)
29. Livingston E, Bucher K. Coronavirus Disease 2019 (COVID-19) in Italy. *JAMA* 2020.
30. Goyal P, Choi JJ, Pinheiro LC, et al. Clinical Characteristics of Covid-19 in New York City. *New England Journal of Medicine* 2020.
31. McMichael TM, Clark S, Pogojans S, et al. COVID-19 in a Long-Term Care Facility - King County, Washington, February 27-March 9, 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:339-42.
32. Kimball A, Hatfield KM, Arons M, et al. Asymptomatic and Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled Nursing Facility - King County, Washington, March 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:377-81.