MAJOR ARTICLE

Severe Acute Respiratory Syndrome Coronavirus 2 Cumulative Incidence, United States, August 2020–December 2020

Patrick Sean Sullivan,[1,a] Aaron J. Siegler,[1,a] Kayoko Shioda,[2] Eric W. Hall,[1] Heather Bradley,[3] Travis Sanchez,[1] Nicole Luisi,[1] Mariah Valentine-Graves,[1] Kristin N. Nelson,[1] Mansour Fahimi,[4] Amanda Kamali,[5] Charles Sailey,[6] and Benjamin A. Lopman[1]

[1]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA; [2]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA; [3]Department of Population Health Sciences, Georgia State University School of Public Health, Atlanta, Georgia, USA; [4]Marketing Systems Group, Horsham, Pennsylvania, USA; [5]California Department of Public Health, Sacramento, California, USA; and [6]Molecular Testing Labs, Vancouver, Washington, USA

Background. Reported coronavirus disease 2019 (COVID-19) cases underestimate severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections. We conducted a national probability survey of US households to estimate cumulative incidence adjusted for antibody waning.

Methods. From August–December 2020 a random sample of US addresses were mailed a survey and self-collected nasal swabs and dried blood spot cards. One adult household member completed the survey and mail specimens for viral detection and total (immunoglobulin [Ig] A, IgM, IgG) nucleocapsid antibody by a commercial, emergency use authorization–approved antigen capture assay. We estimated cumulative incidence of SARS-CoV-2 adjusted for waning antibodies and calculated reported fraction (RF) and infection fatality ratio (IFR). Differences in seropositivity among demographic, geographic, and clinical subgroups were explored.

Results. Among 39 500 sampled households, 4654 respondents provided responses. Cumulative incidence adjusted for waning was 11.9% (95% credible interval [CrI], 10.5%–13.5%) as of 30 October 2020. We estimated 30 332 842 (CrI, 26 703 753–34 335 338) total infections in the US adult population by 30 October 2020. RF was 22.3% and IFR was 0.85% among adults. Black non-Hispanics (Prevalence ratio (PR) 2.2) and Hispanics (PR, 3.1) were more likely than White non-Hispanics to be seropositive.

Conclusions. One in 8 US adults had been infected with SARS-CoV-2 by October 2020; however, few had been accounted for in public health reporting. The COVID-19 pandemic is likely substantially underestimated by reported cases. Disparities in COVID-19 by race observed among reported cases cannot be attributed to differential diagnosis or reporting of infections in population subgroups.

Keywords. SARS-CoV-2; serology; probability survey; incidence; viral detection.

A complete understanding of the US coronavirus disease 2019 (COVID-19) epidemic requires measuring unreported (ie, not diagnosed or diagnosed but not reported to public health surveillance systems) severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections. Cumulative SARS-CoV-2 incidence must account for unreported cases and systematic differences between documented and undocumented cases related to healthcare access or health-seeking behaviors (eg, people experiencing symptoms are more likely to test). Serosurveys identify people who have developed an immune response to SARS-CoV-2, regardless of symptoms, seeking medical care or being diagnosed or reported to public health surveillance

systems. However, most serosurveys to date are subject to selection biases by overrepresenting people concerned about symptoms or exposures, people seeking medical evaluation, or high-risk subpopulations (eg, healthcare workers). Accurate US national estimates of the cumulative incidence of SARS-CoV-2 infection require minimally biased, population-based surveys and screening with viral and antibody detection assays.

The natural history of SARS-CoV-2 infection and immunity informs this effort. Relying solely on detectable levels of SARS-CoV-2 antibodies to estimate cumulative incidence is inadequate because antibodies wane in the months following primary infection [1, 2]. Because of antibody waning, population anti–SARS-CoV-2 antibody prevalence in New York City and the United Kingdom decreased during a time of increasing total reported cases [2-4]. Further, antibodies against the nucleocapsid (N) protein likely wane faster than antibodies against the spike (S) protein [5]. Thus, cross-sectional prevalence estimates that rely on antibody testing, especially studies conducted after spring 2020, likely substantially underestimate cumulative

incidence. Specimens collected later in epidemic are increasingly subject to false-negative antibody results, that is, failing to identify antibodies in previously infected persons.

To develop a nationally representative estimate of the cumulative incidence of SARS-CoV-2, we conducted a national probability survey of US households with mailed at-home specimen collection and polymerase chain reaction (PCR) and serology testing [6]. We calculated adjusted seroprevalence and used a Bayesian model to account for waning antibodies to estimate the overall cumulative incidence in the United States as of 30 October 2020 [7].

METHODS

Sampling

As previously described [6], we used a national address-based household sample of all residential delivery points in the United States (about 130 million addresses) that has been used in numerous health research studies [8-10]. To recruit ≥4000 responding households, 39 500 addresses were sampled. Due to state-level interest in estimates of key parameters, households were oversampled in California (6500 oversampled) and Georgia (12 000 oversampled). In response to differentially low return rates by Black and Hispanic respondents, households in census tracts with >50% Black residents and households with surnames likely to represent Hispanic ethnicity [8] were also oversampled.

Survey and Laboratory Procedures

One person per selected household was asked to enumerate household members and each person's age; 1 household member aged ≥18 years was randomly selected to participate in the COVIDVu study. Consenting participants completed an online survey and provided a self-collected anterior nares (AN) swab and a self-collected dried blood spot (DBS) card as previously described [11] and returned specimens to a central laboratory by mail [12]. AN swabs were tested by PCR using the Thermo EUA (emergency use authorization) Version 2 kit (Thermo Fisher, Waltham, MA). DBS specimens were tested using the BioRad Platelia Total Antibody test (BioRad, Hercules, CA) that targets the NC protein as a laboratory-developed test under Clinical Laboratory Improvements Act/College of American Pathologists (CLIA/CAP) protocols. The Platelia assay has advantages for the purpose of a serosurvey: it detects multiple antibody isotypes; targets the NC protein, which indicates natural infection but not vaccination; and has robust sensitivity (98.0%) and specificity (99.3%) [13]. To characterize potential misclassification biases associated with test performance, we adjusted prevalence estimates for test performance per Sempos and Tian. [14]. We resampled each adjusted prevalence estimate and test performance parameter

estimate (ie, sensitivity and specificity) to estimate confidence intervals (CIs; k = 100 000 iterations) [15].

Antibodies to NC wane more quickly than antibodies to S [5]. Therefore, we quantified the magnitude of potential bias of lower sensitivity of the BioRad test by retesting a subset of BioRad antibody-negative specimens with the EUROIMMUN immunoglobulin (Ig) G assay (Lübeck, Germany) that targets the S protein. The specimen subset comprised participants with negative total Ig results and a high pretest probability of prior infection (n = 122; eg, participants reporting previous diagnosis, hospitalization for COVID-19, or reported loss of smell or taste since 1 January 2020) and a group of randomly selected total Ig-negative participants (n = 275).

The Emory University Institutional Review Board approved the COVIDVu study.

Computation of Sample Weights

Sample weights were developed to facilitate unbiased estimation of key parameters that represent the noninstitutionalized, housed adults (ie, aged ≥18 years US population). Hierarchical hot deck imputation [16] was performed to ensure no participants were missing data for key variables (gender, 0.1% missing; education, 1.2% missing; race, 3.2% missing; ethnicity, 1.6% missing; marital status, 2.2% missing; income, 13.8% missing) needed for weighting. These imputation steps were carried out sequentially within homogeneous imputation cells, each time using the variables previously imputed for the construction of cells for the next variable to be imputed. Next, design weights were computed to reflect the selection probabilities for household addresses and the selection of 1 adult per household and adjusted to account for differential nonresponse. For this purpose, Classification and Regression Tree analysis was used to identify characteristics that were differentially distributed among responding vs nonresponding households. Variables identified as key predictors of nonresponse were homeownership status (rent vs own), residing in a household located in a census tract with >50% Black residents, presence of Hispanic surname, and presence of household information about income or number of adults on the address-based sampling frame.

In the next step, nonresponse-adjusted design weights were post-stratified to distributions of demographic characteristics among US adults. Specifically, an iterative proportional fitting (raking) procedure was used to align weighted distributions of respondents with respect to gender, age, race/ethnicity, education, income, marital status, and census division [17]. Weights were examined to detect extreme outliers and trimmed at the 99th percentile on both ends of the distribution.

Seroprevalence analyses were conducted in SAS v9.4 and SUDAAN. Using the sampling weights, we estimated the weighted seroprevalence and 95% modified Wilson score

confidence limits of total Ig for the entire sample and for demographic and clinical factors of interest. To identify significant differences, prevalence ratios (PRs) and corresponding 95% CIs were estimated using weighted logistic regression procedures in SUDAAN. A χ^2 test for linear trend in proportions was performed for seroprevalence across levels of education.

Estimation of SARS-CoV-2 Cumulative Incidence and Infection Fatality Ratio Accounting for Waning Antibodies

To adjust for SARS-CoV-2 antibodies waning below the detectable levels [18, 19], we used a Bayesian model to estimate the cumulative incidence of SARS-CoV-2 at the median date of our sample (30 October 2020). The model uses population-level cross-sectional data from the present study and accounts for both the expected timeline of seroconversion and the timeline for seroreversion. Details of this model have been described [7]. Briefly, the model estimates the timing of infection based on empirical data on the distribution of time from symptom onset to death and is calibrated with the national weighted seroprevalence estimate from the present study by applying cumulative density functions for the time from seroconversion to seroreversion. The model generates a daily estimate of new infections and derives a cumulative incidence estimate by summing the total number of modeled infections since the beginning of the epidemic. The model directly estimates the infection fatality ratio (IFR) [7]. We also estimated the IFR for 2 age strata (55–64 and ≥65 years) where adequate age-specific time-series data were available in Centers for Disease Control and Prevention (CDC) public use datasets. An exploratory analysis of cumulative incidence was conducted for CI through 31 December 2020 using updated mortality data reported through 15 April 2021.

Calculation of Reported Fraction

We defined reported fraction as the ratio of reported cases in the United States as of 30 October 2020 (using data from the CDC's public use dataset [20] and assuming that those aged 18–19 years represented 21% of the 10- to 19-year age group) and the cumulative incidence as of the same date. Credible intervals (CrIs) were constructed using the 95% CrIs for the cumulative incidence of the denominator [21].

RESULTS

Sampling, Participation Rates, and Representation of Racial/Ethnic Minorities

A total of 39 500 registration packages were mailed to sampled US households from July 2020 through October 2020 (Figure 1). There were 2444 addresses (6.2%) that were unable to receive mail and excluded from the sample. A total of 5666 surveys (15.3%) were completed. Of those completing surveys, 4654 (12.6% of sampled households) also returned a DBS specimen collected during the period 9 August 2020–8 December 2020 with a valid antibody result. There were 450 other participants (7.9%) who did not have a total Ig result but had a valid PCR test. The overall participation rate was 15.3% for the survey only and 12.6% for the survey and a valid antibody test result.

Antibody and PCR RNA Positivity

Overall, 229 of 4654 (4.92%) DBS specimens were reactive for total Ig (ie, unadjusted seroprevalence); these made inference to the seroprevalence among 242 875 582 US adults (Table 1). The weighted seroprevalence was 5.24% (CI, 4.14%–6.60%); seroprevalence results suggested that the number of US adults with prevalent anti–SARS-CoV-2 Ig not adjusted for waning antibodies for the period 9 August 2020–8 December 2020 was 12 722 882. In a sensitivity analysis adjusting for test performance [13], the overall prevalence of antibodies was lower (4.71%; CI, 3.3–6.11; Supplementary Table 1). There were 36 of 4984 (0.72%) AN specimens that were positive by PCR testing, of which 10 (29%) were also reactive for total Ig.

### Characterizing Potential Bias From Lower Sensitivity for Detection of Antibodies to NC Protein

Among 122 samples with a negative NC Ig assay and a clinical history compatible with COVID-19 disease, 1 of 122 (0.8%) had a reactive result on the IgG assay for the S protein. No specimen from the 275 randomly selected NC Ig-nonreactive specimens was reactive on the IgG assay for the S protein. Therefore, we believed that the choice of the NC target did not result in misclassification bias and used the results of the BioRad assay for all analyses reported here.

Associations of Antibody Positivity

Weighted seroprevalence was 3-fold higher among Hispanic and 2-fold higher among Black, non-Hispanic participants compared with White, non-Hispanic participants (Table 2). Compared with persons aged ≥65 years, weighted seroprevalence was 3 times higher in those aged 18–34 or 35–44 years. Weighted seroprevalence was nearly double among persons living in the South compared with the West, and results showed an inverse relationship between educational attainment and seroprevalence (trend in proportions, $P = .008$). Seroprevalence was higher among participants residing in metropolitan areas and who reported cold/flu symptoms or loss of taste or smell since 1 January 2020. Overall, nearly 9 in 10 Ig-seropositive participants reported at least 1 symptom (loss of taste/smell, flu, or any of the other potential symptoms listed in the Table 2 footnote), and 8 in 10 of those who were SARS-CoV-2–seronegative reported ≥1 symptom since 1 January 2020. There was no difference in seropositivity by comorbidities.

Estimated Cumulative Incidence of SARS-CoV-2 Infections and IFR Adjusted for Waning Antibodies

Estimated cumulative incidence adjusted for waning antibodies was 11.9% (CrI, 10.5%–13.5%) on 30 October 2020
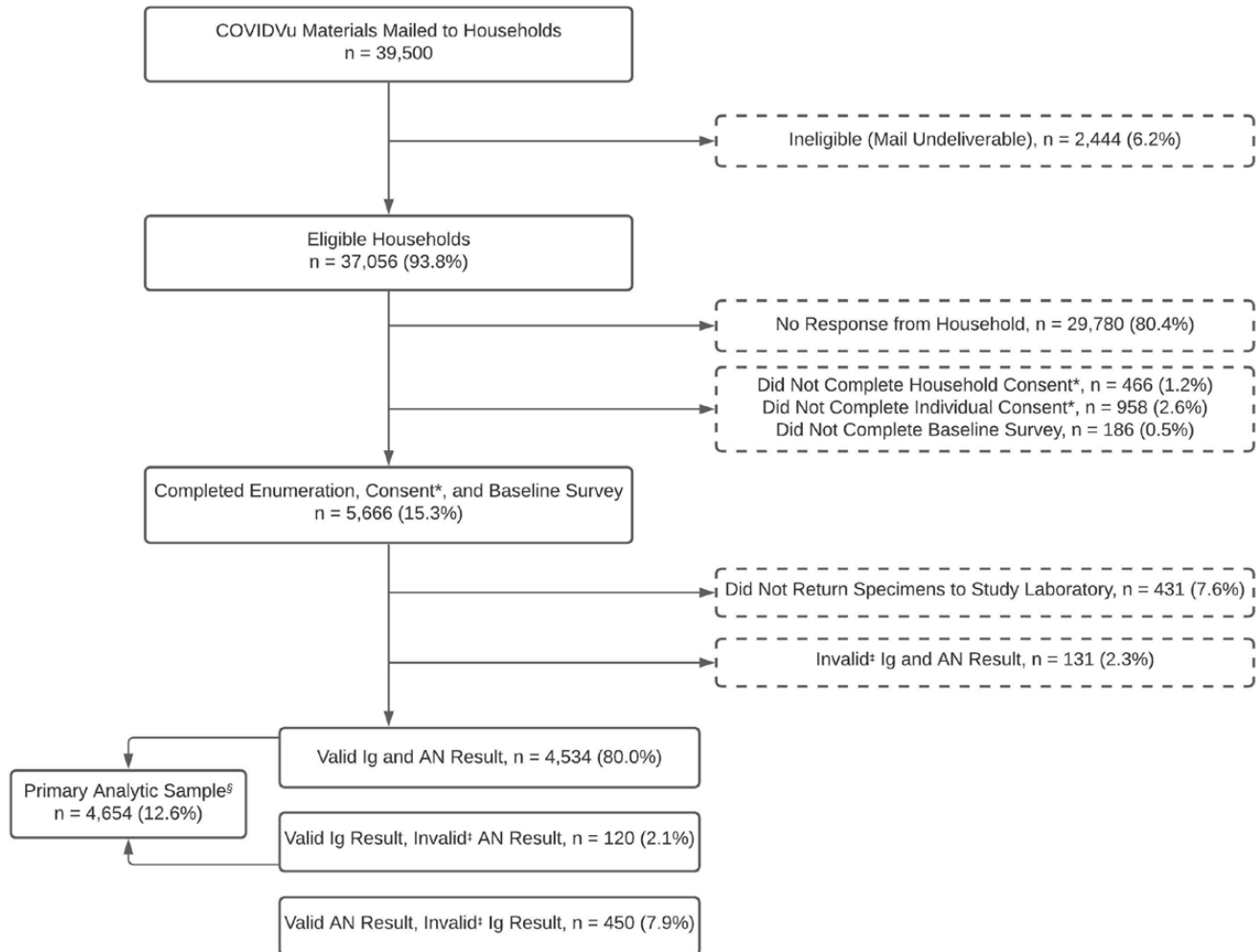
Figure 1. Consort diagram for a national household probability sample of US households to estimate the cumulative incidence of severe acute respiratory syndrome coronavirus 2 infection in the United States, 2020. Abbreviations: AN, anterior nares; COVIDVu, coronavirus disease 2019 study; Ig, immunoglobulin.

(Figure 2). The estimated IFR was 0.85% (CrI, 0.76%–0.97%) for adults aged ≥18 years, 0.59% (0.45%–0.83%) for those aged 55–64 years, and 7.1% (5.04%–10.38%) among those aged ≥65 years. We estimated 30 332 842 (CrI, 26 703 753– 34 335 338) infections among adults aged ≥18 years by 30 October 2020. There were 6 769 219 cumulative reported COVID-19 cases in adults through 30 October 2020, suggesting that about 1 in 5 (22.3%; CrI, 19.7%–25.3%) of adult SARS-CoV-2 infections had been reported as a COVID-19 case by 30 October 2020. The exploratory estimate for adult cumulative incidence through 31 December 2020 was 18.2% (CrI, 16.1%–20.4%). Estimated daily seroprevalence is also presented in Figure 2. Estimated daily seroprevalence tracked in parallel to cumulative incidence through summer 2020 but then began increasing more slowly than cumulative incidence.

DISCUSSION

By accounting for data on the distribution of time from exposure to seroconversion, seroreversion, and time to death, we report that although the daily seroprevalence of antibodies to SARS-CoV-2 remained relatively stable at between 4% and 5% from August 2020 to October 2020, cumulative incidence continued to climb. The cumulative incidence rose to more than 30 million US adults, and nearly 1 in 8 had been infected with the virus by the end of October 2020.

Understanding the extent of the SARS-CoV-2 epidemic in the United States has been challenging since the beginning of the epidemic for multiple reasons. First, deficits in testing capacity were acute in the early months of the epidemic, resulting in substantial underdiagnosis of COVID-19 cases, especially mildly symptomatic cases [22]. Second, early serosurveys were frequently based on convenience samples and subject to

Table 1. Severe Acute Respiratory Syndrome Coronavirus 2 Serology and Viral Detection Results for a Probability Sample of 4654 US Households and Weighted Results Compared With the US Population Aged ≥ 18 Years, United States, 2020

Characteristic	Ig Only				Ig or AN				US Population Aged ≥18 Years[a]	
	Sample		Weighted Sample		Sample		Weighted Sample			
	N	%	Weighted N	Column %	N	%	Weighted N	Column %	N	%
Overall	4654	100	242 875 582	100	5104	100	242 972 595	100	255 200 373	100
Sex										
Male	1927	41.4	115 613 214	47.6	2129	45.7	115 725 392	47.6	124 348 656	48.7
Female	2727	58.6	127 262 368	52.4	2975	63.9	127 247 203	52.4	130 851 717	51.3
Race/Ethnicity										
Hispanic	607	13	40 277 007	16.6	668	14.4	40 389 513	16.6	41 884 672	16.4
Non-Hispanic Black	683	14.7	27 643 982	11.4	797	17.1	28 062 416	11.6	32 169 434	12.6
Non-Hispanic White	3063	65.8	153 881 404	63.4	3316	71.3	153 414 972	63.2	162 644 095	63.7
Other	301	6.5	21 073 189	8.7	323	6.9	21 105 695	8.7	18 502 172	7.3
Age, years										
18–34	1013	21.8	67 946 989	28	1103	23.7	68 229 816	28.1	76 159 527	29.8
35–44	777	16.7	40 347 844	16.6	850	18.3	40 347 557	16.6	41 659 144	16.3
45–54	765	16.4	39 524 761	16.3	833	17.9	39 481 380	16.3	40 874 902	16
55–64	926	19.9	41 638 646	17.1	1012	21.7	41 389 099	17	42 448 537	16.6
65+	1173	25.2	53 417 341	22	1306	28.1	53 524 744	22	54 058 263	21.2
US census region										
Northeast	476	10.2	42 937 799	17.7	519	11.2	43 151 385	17.8	44 478 478	17.4
Midwest	591	12.7	51 141 237	21.1	632	13.6	50 719 007	20.9	52 980 427	20.8
South	2275	48.9	90 171 242	37.1	2531	54.4	90 429 763	37.2	97 108 254	38.1
West	1312	28.2	58 625 304	24.1	1422	30.6	58 672 440	24.2	60 633 214	23.8

Weighted N is the sum of the weights of participants.

Abbreviations: AN, anterior nares swab for polymerase chain reaction testing/severe acute respiratory syndrome coronavirus 2 detection; Ig, total immunoglobulin (IgA, IgM, or IgG) to nucleocapsid protein; N, total participants.

[a]2019 bridged-race estimates (National Vital Statistics System).

Table 2. Unweighted and Weighted Severe Acute Respiratory Syndrome Coronavirus 2 Antibody Prevalence for a Probability Sample of 4654 US Households and Weighted Results and Prevalence Ratios, United States, 2020

Characteristic	Unweighted			Weighted							
	n	N	Prevalence	n	N	Prevalence	95% CI[a]		Prevalence Ratio	95% CI	
Overall	229	4654	4.9	12 722 882	242 875 582	5.24	4.14	6.60	n/a		
Sex											
Male	92	1927	4.8	5 983 835	115 613 214	5.18	3.59	7.41	Reference		
Female	137	2727	5.0	6 739 047	127 262 368	5.30	3.93	7.10	1.02	.64	1.64
Race/Ethnicity											
Hispanic	51	607	8.4	4 631 941	40 277 007	11.50	7.54	17.16	3.11	1.83	5.28
Non-Hispanic Black	71	683	10.4	2 200 979	27 643 982	7.96	4.73	13.11	2.15	1.17	3.97
Non-Hispanic White	104	3063	3.4	5 692 713	153 881 404	3.70	2.67	5.10	Reference		
Other	3	301	1.0	197 250	21 073 189	0.94	0.27	3.24	0.25	.06	1.01
Age, years											
18–34	72	1013	7.1	4 558 387	67 946 989	6.71	4.44	10.01	2.70	1.18	6.18
35–44	53	777	6.8	2 963 168	40 347 844	7.34	4.65	11.41	2.96	1.26	6.93
45–54	33	765	4.3	1 911 289	39 524 761	4.84	2.59	8.84	1.95	.75	5.06
55–64	37	926	4.0	1 963 111	41 638 646	4.71	2.76	7.94	1.90	.77	4.66
65+	34	1173	2.9	1 326 927	53 417 341	2.48	1.22	4.98	Reference		
US census region											
Northeast	20	476	4.2	2 619 466	42 937 799	6.10	3.54	10.32	1.67	.79	3.55
Midwest	19	591	3.2	2 027 923	51 141 237	3.97	2.24	6.92	1.09	.50	2.36
South	149	2275	6.6	5 934 236	90 171 242	6.58	4.66	9.21	1.80	.97	3.36
West	41	1312	3.1	2 141 257	58 625 304	3.65	2.18	6.07	Reference		
Urbanicity (zip code)											
Micropolitan/Small town/Rural	20	468	4.3	728 649	32 292 975	2.26	1.20	4.20	Reference		
Metropolitan	209	4186	5.0	11 994 233	210 582 607	5.70	4.46	725	2.52	1.27	5.00
Education											
High School/GED or less	47	698	6.7	5 598 377	85 965 483	6.51	4.23	9.91	1.63	.94	2.82
Some college/Associate's degree	71	1409	5.0	3 727 595	69 226 861	5.38	3.67	784	1.35	.81	2.24
Bachelor's degree	68	1430	4.8	2 228 895	55 756 279	4.00	2.85	5.57	ref		
Graduate degree	43	1117	3.9	1 168 014	31 926 958	3.66	2.23	5.93	0.92	.50	1.67
Annual income											
$0–$24 999	39	721	5.4	1 165 276	29 566 723	3.94	2.32	6.62	0.79	.40	1.57
$25 000–$49 999	56	916	6.1	3 276 418	41 443 877	7.91	4.89	12.53	1.59	.84	3.03
$50 000–$99 999	69	1445	4.8	3 638 036	73 211 031	4.97	3.23	757	Reference		
$100 000–$199 999	55	1125	4.9	3 435 662	67 795 060	5.07	3.26	779	1.02	.55	1.89
$200 000+	10	447	2.2	1 207 490	30 858 891	3.91	1.61	9.18	0.79	.29	2.15
Health insurance											
None	19	263	7.2	1 243 547	13 358 208	9.31	4.35	18.83	1.88	.83	4.28
Medicare/Medicaid/Other government plan	60	1352	4.4	2 887 942	66 230 875	4.36	2.70	6.98	0.88	.50	4.28
Private/Parent's plan	135	2734	4.9	7 286 120	147 299 448	4.95	3.65	6.67	Reference		

Table 2. Continued

Characteristic	Unweighted			Weighted							
	n	N	Prevalence	n	N	Prevalence	95% CI[a]		Prevalence Ratio	95% CI	
Don't know	15	305	4.9	1 305 273	15 987 051	8.16	3.73	16.94	1.65	0.71	3.84
Comorbidities											
Diabetes	27	438	6.2	683 580	22 485 621	3.04	1.08	8.26	0.56	0.19	1.67
Heart condition	11	325	3.4	430 691	16 727 097	2.57	1.03	6.32	0.47	0.18	1.26
Chronic lung disease	16	389	4.1	1 274 183	21 451 947	5.94	2.44	13.77	1.15	0.45	2.94
Hypertension	50	1045	4.8	1 175 196	46 383 405	2.53	1.54	4.15	0.43	0.25	0.76
Symptoms since 1 January 2020											
Cold/Flu	149	1917	7.8	8 053 479	98 083 444	8.21	6.14	10.90	2.55	1.57	4.13
Loss of taste or smell	103	272	37.9	5 396 043	13 179 352	40.94	30.94	51.75	12.84	8.50	19.37
Any other symptom[b]	202	3803	5.3	11 222 678	196 089 280	5.72	4.46	7.31	1.78	0.86	3.70
Symptoms in past 30 days											
Loss of taste or smell	25	85	29.4	2 185 030	4 449 757	49.10	32.16	66.26	11.11	7.06	17.49
Any other symptom[b]	131	2816	4.7	7 920 970	144 955 397	5.46	4.04	7.35	1.11	0.69	1.80
Month of sample collection											
August	36	1195	3.0	4 100 580	98 937 128	4.14	2.67	6.37	Reference		
September	23	406	5.7	1 981 937	33 460 432	5.92	3.33	10.31	1.43	0.69	2.95
October	27	812	3.3	2 675 819	55 101 083	4.86	2.90	8.02	1.17	0.60	2.31
1 November–8 December	143	2241	6.4	3 964 546	53 376 939	7.16	4.86	10.44	1.73	0.96	3.10

n is the weighted number of cases; weighted N is the sum of the weights of participants.

Abbreviations: CI, confidence interval; N, total participants.

[a]Confidence intervals are calculated using the modified Wilson method.

[b]Symptoms include cough, itchy eyes, shortness of breath, runny/stuffy nose, fever, headache, chills, diarrhea, muscle pain, sore throat, vomiting, or nausea.
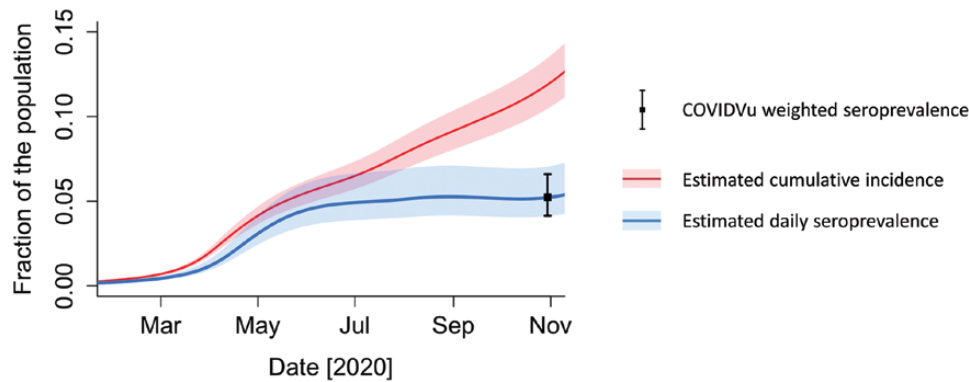
Figure 2. Estimated cumulative incidence of severe acute respiratory syndrome coronavirus 2 infection adjusted for waning antibodies and daily seroprevalence, United States, 2020. Abbreviation: COVIDVu, coronavirus disease 2019.

selection bias for people concerned about exposure or symptoms [6, 23]. Third, many SARS-CoV-2 infections may be asymptomatic; asymptomatic or paucisymptomatic persons are unlikely to seek diagnostic testing and be reported as cases. Fourth, reporting systems for COVID-19 had to be established very quickly by public health institutions, and there was substantial underreporting of demographic data, including race/ethnicity, needed to describe relative impacts of the epidemic [24, 25]. Finally, naturally acquired antibodies to SARS-CoV-2 wane over time, and antibodies directed toward different antigenic targets might wane at different rates [26]. As a result, seroprevalence estimates alone are not a reliable indicator of cumulative incidence, even over the short history of the US epidemic. Our study addressed many of these challenges by collecting data from randomly selected US households (minimizing selection bias), oversampling to achieve a diverse sample, and using statistical methods to account for waning antibodies.

Previously reported US seroprevalence studies have featured varying degrees of probability sampling methods and convenience sampling. One study constructed a demographically and geographically representative sample from a sampling frame of screened volunteers [27]. However, to our knowledge, no study has reported national data from a probability sample of US households [28]. A synthesis of population-based samples and remnant clinical samples yielded a seroprevalence of 14.3% by mid-November 2020 but did not consider waning antibodies and called for additional serosurvey data [29]. A study of US plasma donors reported seroprevalence of 8.0% in July 2020, but dialysis patients tend to be significantly older than US adults overall [30]. Other seroprevalence studies have used various strategies to minimize bias, including the use of proprietary sampling frames (4% in Los Angeles April 2020 [23]), use of remnant blood specimens from blood donors (1.8% prevalence in June 2020 –August 2020 [31]) or specimens submitted for other laboratory testing (range of 1.0%–6.9% across 10 US sites in March 2020–May 2020 [32]), and flow sampling through grocery stores (12.5% in New York City in March 2020 [33]). The CDC publishes state-specific seroprevalence estimates from commercial laboratory samples, which was >20% in many states as of February 2021 [34]. The CDC reported results from local population-based household samples in metropolitan Atlanta, Georgia (2.5% in April 2020–May 2020 [35]), and Indiana (seroprevalence 1.0% in May 2020–June 2020 [36]). Reports of previous surveys have recognized the limitations of seroprevalence studies alone to estimate cumulative incidence and have called for representative surveys to minimize sampling bias [37].

Our crude antibody prevalence was adjusted in 2 ways. First, we applied sampling weights to our observed data to account for the sampling process, resulting in a small increase in the seroprevalence estimate. Second, we accounted for waning antibodies [7]. Although studies conducted in the first half of 2020 might have been minimally impacted by waning antibodies, serology studies that collected data in the second half of 2020 were subject to substantial misclassification bias, perhaps differentially by symptomatology [38, 39]. In a period prevalence survey that spanned several months, people with a previous SARS-CoV-2 infection might lose detectable antibodies and be misclassified; on the other hand, in periods of high incidence (eg, December 2020), people with positive PCR tests indicating infection might be misclassified as not being a cumulative incident case because antibodies had not yet developed. These potentially misclassified statuses are temporally varying during the beginning of an epidemic: misclassification due to waning antibodies will be a more prominent bias in later months, and misclassification of infection status by antibody measurement will be greater during periods of high incidence. The combined effect of these biases was likely large through the fall of 2020. In Figure 2, daily seroprevalence stabilized even as cumulative incidence rose: each day some people acquired a new detectable antibody result, and others lost detectable antibodies).

Our estimate of the reported fraction is higher than estimates from some previous reports. Based on projections from remnant blood donors and clinical samples, the CDC estimated in June 2020 that only 10% of cumulative SARS-CoV-2 infections had been reported [40]. It might be that the reported fraction has increased as testing capacity has increased. Our data confirm that the reported disproportionate impact on Black [41-45] and Hispanic [45-48] people also persists in the representative sample, as did previously reported associations of higher positivity with lower age and metropolitan residence [37]. Establishing these associations in a representative study is important because measures of relative impact developed using reported data are impacted by differences in testing availability by race or urbanicity [49]. Others have reported disparities by race, residence and age based on diagnosed cases; we found that these disparities are also observed in a representative sample of US respondents corrected for waning, which indicates that these previously reported disparities were not an artifact of a higher a risk of symptoms or testing in certain groups. Our data also suggest that the geographic areas of higher burden have shifted toward the South since earlier in the epidemic [50, 51].

Our study is subject to limitations. We used a representative sampling frame, but our response rate was 12.6%, which is low but typical for mailed surveys using address-based sampling frames [52]. The CDC's 2 household samples, conducted as a door-to-door offer of enrollment, also had low response rates (23.6%–23.7% [35]). Weighting for nonresponse addresses selection bias for some traits known for households, but residual selection bias exists. Our results are likely subject to differential response bias; we addressed this by oversampling specific groups (eg, Black and Hispanic households) with lower response rates and by weighting for nonresponse of households. We were only able to address differential nonresponse using characteristics of the population that were available to us on the frame (eg, population distributions by race/ethnicity or household income levels). Characteristics that may be associated with COVID-19 risk but not available at the population level, such as higher general propensity to take risks, were not available for extrapolation to the underlying population and therefore may contribute to uncorrected selection bias. Our laboratory results were subject to misclassification based on the latent period for seroconversion and waning antibodies. Unlike most other studies reported to date, we accounted for these biases through our modeling approach.

We conducted additional testing to quantify potential biases associated with our choice of an antibody test targeting the NC protein, which is more subject to waning; the results indicated minimal bias toward misclassifying true antibody-positive tests as negative. We were also at risk for misclassification because DBS cards have less biological material available for use in assays. As part of our CLIA validation, DBS vs venipuncture specimens for both serology assays showed 100% sensitivity and specificity for DBS tests compared with a serum gold standard (n = 30 positives and 30 negatives, unpublished results, available upon request).

Our study furthers previous seroprevalence surveys by estimating cumulative incidence in a national probability sample of US households, addressing many of the limitations of previous estimates of SARS-CoV-2 burden in the United States. We found somewhat higher estimates of reported fraction than others, which have ranged from 4%–16% [32, 37]. Our findings suggest substantially higher cumulative incidence than has been reported in previous studies that did not adjust for waning antibodies [53]. A related finding is that our estimate of IFR is somewhat lower than had been suggested by studies that did not include waning-adjusted estimates of cumulative incidence (0.85% vs 1.39% [54]); the timing of analyses likely also influenced these differences. Representative population-based samples provide minimally biased data as a contextual framework for other types of studies. Adjusting for waning antibodies is critical to developing credible estimates of cumulative incidence and will become increasingly important over time.

Supplementary Data

Notes

References

1. Self WH, Tenforde MW, Stubblefield WB, et al; CDC COVID-19 Response Team; IVY Network. Decline in SARS-CoV-2 antibodies after mild infection among frontline health care personnel in a multistate hospital network—12 states, April-August 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:1762–6.
2. Dan JM, Mateus J, Kato Y, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. Science **2021**; doi:10.1126/science.abf4063.
3. Stadlbauer D, Tan J, Jiang K, et al. Repeated cross-sectional sero-monitoring of SARS-CoV-2 in New York City. Nature **2021**; 590:146–50.
4. Ward H, Cooke G, Atchison CJ, et al. Declining prevalence of antibody positivity to SARS-CoV-2: a community study of 365,000 adults. MedRxiv **2020**. [Preprint]. Available from: https://www.medrxiv.org/content/10.1101/2020.10.26.20219725v1.abstract.

5. Fenwick C, Croxatto A, Coste AT, et al. Changes in SARS-CoV-2 spike versus nucleoprotein antibody responses impact the estimates of infections in population-based seroprevalence studies. J Virol 2021; 95. doi:10.1128/JVI.01828-20.

6. 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. Ann Epidemiol 2020; 49:50–60.

7. Shioda K, Lau MSY, Kraay ANM, et al. Estimating the cumulative incidence of SARS-CoV-2 infection and the infection fatality ratio in light of waning antibodies. Epidemiology 2021; 32:518–24.

8. Lavange LM, Kalsbeek WD, Sorlie PD, et al. Sample design and cohort selection in the Hispanic community health study/study of Latinos. Ann Epidemiol 2010; 20:642–9.

9. Chido-Amajuoyi OG, Yu RK, Agaku I, Shete S. Exposure to court-ordered tobacco industry antismoking advertisements among US adults. JAMA Netw Open 2019; 2:e196935.

10. Cerel J, Maple M, van de Venne J, Moore M, Flaherty C, Brown M. Exposure to suicide in the community: prevalence and correlates in one U.S. state. Public Health Rep 2016; 131:100–7.

11. Sullivan PS, Sailey C, Guest JL, et al. Detection of SARS-CoV-2 RNA and antibodies in diverse samples: protocol to validate the sufficiency of provider-observed, home-collected blood, saliva, and oropharyngeal samples. JMIR Public Health and Surveillance 2021; 6:e19054.

12. Guest JL, Sullivan PS, Valentine-Graves M, et al. Suitability and sufficiency of telehealth clinician-observed, participant-collected samples for SARS-CoV-2 testing: the iCollect cohort pilot study. JMIR Public Health Surveill 2020; 6:e19731.

13. US Food and Drug Administration. EUA authorized serology test performance. Available at: https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/eua-authorized-serology-test-performance. Accessed 28 June 2021.

14. Sempos CT, Tian L. Adjusting coronavirus prevalence estimates for laboratory test kit error. Am J Epidemiol 2021; 190:109–15.

15. DiCiccio TJ, Efron B. Bootstrap confidence intervals. SSO Schweiz Monatsschr Zahnheilkd 1996; 11:189–228.

16. Andridge RR, Little RJ. A review of hot deck imputation for survey non-response. Int Stat Rev 2010; 78:40–64.

17. American Community Survey. Available at: http://methods.sagepub.com/reference/encyclopedia-of-survey-research-methods/n16.xml. Accessed 30 June 2020.

18. National Academies of Sciences, Engineering, Medicine. Rapid expert consultations on the COVID-19 pandemic: March 14, 2020–April 8, 2020. National Academies Press, 2020.

19. Long QX, Tang XJ, Shi QL, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. Nat Med 2020; 26:1200–4.

20. Centers for Disease Control and Prevention, COVID-19 Response. COVID-19 case surveillance public data access, summary, and limitations (version date: December 31, 2020). 2020. Available at: https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf. Accessed 2 February 2021.

21. Eberly LE, Casella G. Estimating Bayesian credible intervals. J Stat Plan Inference 2003; 112:115–32.

22. Dyer O. Covid-19: US testing ramps up as early response draws harsh criticism. BMJ 2020; 368:m1167.

23. Sood N, Simon P, Ebner P, et al. Seroprevalence of SARS-CoV-2-specific antibodies among adults in Los Angeles County, California, on April 10–11, 2020. JAMA 2020; 323:2425–7.

24. Stokes EK, Zambrano LD, Anderson KN, et al. Coronavirus disease 2019 case surveillance—United States, January 22–May 30, 2020. MMWR Morb Mortal Wkly Rep 2020; 69:759–65.

25. Killerby ME, Link-Gelles R, Haight SC, et al; CDC COVID-19 Response Clinical Team. Characteristics associated with hospitalization among patients with COVID-19 - metropolitan Atlanta, Georgia, March–April 2020. MMWR Morb Mortal Wkly Rep 2020; 69:790–4.

26. Stephens DS, McElrath MJ. COVID-19 and the path to immunity. JAMA 2020; 324:1279–81.

27. Kalish H, Klumpp-Thomas C, Hunsberger S, et al. Mapping a pandemic: SARS-CoV-2 seropositivity in the United States. medRxiv 2021. doi:10.1101/2021.01.27.21250570.

28. Lai CC, Wang JH, Hsueh PR. Population-based seroprevalence surveys of anti-SARS-CoV-2 antibody: an up-to-date review. Int J Infect Dis 2020; 101:314–22.

29. Angulo FJ, Finelli L, Swerdlow DL. Estimation of US SARS-CoV-2 infections, symptomatic infections, hospitalizations, and deaths using seroprevalence surveys. JAMA Netw Open 2021; 4:e2033706.

30. Anand S, Montez-Rath M, Han J, et al. Prevalence of SARS-CoV-2 antibodies in a large nationwide sample of patients on dialysis in the USA: a cross-sectional study. Lancet 2020; 396:1335–44.

31. Dodd RY, Xu M, Stramer SL. Change in donor characteristics and antibodies to SARS-CoV-2 in donated blood in the US, June–August 2020. JAMA 2020; doi:10.1001/jama.2020.18598.

32. 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; 180:1576–86.

33. Rosenberg ES, Tesoriero JM, Rosenthal EM, et al. Cumulative incidence and diagnosis of SARS-CoV-2 infection in New York. Ann Epidemiol 2020; 48:23–29.e4.

34. Centers for Disease Control and Prevention. COVID data tracker: nationwide commercial laboratory seroprevalence survey. Available at: https://covid.cdc.gov/covid-data-tracker/#national-lab. Accessed 11 February 2021.

35. Biggs HM, Harris JB, Breakwell L, et al; CDC Field Surveyor Team. Estimated community seroprevalence of SARS-CoV-2 antibodies—two Georgia counties, April 28–May 3, 2020. MMWR Morb Mortal Wkly Rep 2020; 69:965–70.

36. Menachemi N, Yiannoutsos CT, Dixon BE, et al. Population point prevalence of SARS-CoV-2 infection based on a statewide random sample—Indiana, April 25–29, 2020. MMWR Morb Mortal Wkly Rep 2020; 69:960–4.

37. Bajema KL, Wiegand RE, Cuffe K, et al. Estimated SARS-CoV-2 seroprevalence in the US as of September 2020. JAMA Intern Med 2020; doi:10.1001/jamainternmed.2020.7976.

38. Perreault J, Tremblay T, Fournier MJ, et al. Waning of SARS-CoV-2 RBD antibodies in longitudinal convalescent plasma samples within 4 months after symptom onset. Blood 2020; 136:2588–91.

39. Choe PG, Kang CK, Suh HJ, et al. Waning antibody responses in asymptomatic and symptomatic SARS-CoV-2 infection. Emerg Infect Dis 2021; doi:10.3201/eid2701.203515.

40. CDC says COVID-19 cases in U.S. may be 10 times higher than reported. 2020. Available at: https://www.nbcnews.com/health/health-news/cdc-says-covid-19-cases-u-s-may-be-10-n1232134. Accessed 29 December 2020.

41. Millett GA, Jones AT, Benkeser D, et al. Assessing differential impacts of COVID-19 on black communities. Ann Epidemiol 2020; 47:37–44.

42. Poulson M, Geary A, Annesi C, et al. National disparities in COVID-19 outcomes between Black and White Americans. J Natl Med Assoc 2020; doi:10.1016/j.jnma.2020.07.009.

43. Holtgrave DR, Barranco MA, Tesoriero JM, Blog DS, Rosenberg ES. Assessing racial and ethnic disparities using a COVID-19 outcomes continuum for New York State. Ann Epidemiol 2020; 48:9–14.

44. Egede LE, Walker RJ. Structural racism, social risk factors, and Covid-19—a dangerous convergence for black Americans. N Engl J Med 2020; 383:e77.

45. Moore JT, Ricaldi JN, Rose CE, et al; COVID-19 State, Tribal, Local, and Territorial Response Team. Disparities in incidence of COVID-19 among underrepresented racial/ethnic groups in counties identified as hotspots during June 5–18, 2020—22 States, February–June 2020. MMWR Morb Mortal Wkly Rep 2020; 69:1122–6.

46. Rodriguez-Diaz CE, Guilamo-Ramos V, Mena L, et al. Risk for COVID-19 infection and death among Latinos in the United States: examining heterogeneity in transmission dynamics. Ann Epidemiol 2020; 52:46–53.e2.

47. Macias Gil R, Marcelin JR, Zuniga-Blanco B, Marquez C, Mathew T, Piggott DA. COVID-19 pandemic: disparate health impact on the Hispanic/Latinx population in the United States. J Infect Dis 2020; 222:1592–5.

48. Bui DP, McCaffrey K, Friedrichs M, et al. Racial and ethnic disparities among COVID-19 cases in workplace outbreaks by industry sector—Utah, March 6–June 5, 2020. MMWR Morb Mortal Wkly Rep 2020; 69:1133–8.

49. Tao R, Downs J, Beckie TM, Chen Y, McNelley W. Examining spatial accessibility to COVID-19 testing sites in Florida. Ann GIS 2020; 26:319–27.

50. Oster AM, Kang GJ, Cha AE, et al. Trends in number and distribution of COVID-19 hotspot counties—United States, March 8–July 15, 2020. MMWR Morb Mort Wkly Rep 2020; 69:1127–32.

51. Oster AM, Caruso E, DeVies J, Hartnett KP, Boehmer TK. Transmission dynamics by age group in COVID-19 hotspot counties—United States, April–September 2020. MMWR Morb Mortal Wkly Rep 2020; 69:1494–6.

52. Fahimi M, Link M, Schwartz DA, Levy P, Mokdad A. Tracking chronic disease and risk behavior prevalence as survey participation declines: statistics from the behavioral risk factor surveillance system and other national surveys. Prev Chronic Dis 2008;07_0097a.

53. O'Driscoll M, Ribeiro Dos Santos G, Wang L, et al. Age-specific mortality and immunity patterns of SARS-CoV-2. Nature 2021; 590:140–5.

54. 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. Lancet Infect Dis 2021; 21:203–12.