



Original Investigation | Infectious Diseases

Estimation of US SARS-CoV-2 Infections, Symptomatic Infections, Hospitalizations, and Deaths Using Seroprevalence Surveys

Frederick J. Angulo, DVM, PhD; Lyn Finelli, DrPH, MS; David L. Swerdlow, MD

Abstract

IMPORTANCE Estimates of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) disease burden are needed to help guide interventions.

OBJECTIVE To estimate the number of SARS-CoV-2 infections, symptomatic infections, hospitalizations, and deaths in the US as of November 15, 2020.

DESIGN, SETTING, AND PARTICIPANTS In this cross-sectional study of respondents of all ages, data from 4 regional and 1 nationwide Centers for Disease Control and Prevention (CDC) seroprevalence surveys (April [n = 16 596], May, June, and July [n = 40 817], and August [n = 38 355]) were used to estimate infection underreporting multipliers and symptomatic underreporting multipliers. Community serosurvey data from randomly selected members of the general population were also used to validate the underreporting multipliers.

MAIN OUTCOMES AND MEASURES SARS-CoV-2 infections, symptomatic infections, hospitalizations, and deaths. The median of underreporting multipliers derived from the 5 CDC seroprevalence surveys in the 10 states that participated in 2 or more surveys were applied to surveillance data of reported coronavirus disease 2019 (COVID-19) cases for 5 respective time periods to derive estimates of SARS-CoV-2 infections and symptomatic infections, which were summed to estimate SARS-CoV-2 infections and symptomatic infections in the US. Estimates of infections and symptomatic infections were combined with estimates of the hospitalization ratio and fatality ratio to derive estimates of SARS-CoV-2 hospitalizations and deaths. External validity of the surveys was evaluated with the April CDC survey by comparing results to 5 serosurveys (n = 22 118) that used random sampling of the general population. Internal validity of the multipliers from the 10 specific states was assessed in the August CDC survey by comparing multipliers from the 10 states to all states. A sensitivity analysis was conducted using the interquartile range of the multipliers to derive a high and low estimate of SARS-CoV-2 infections and symptomatic infections. The underreporting multipliers were then used to adjust the reported COVID-19 infections to estimate the full SARS-CoV-2 disease burden.

RESULTS Adjusting reported COVID-19 infections using underreporting multipliers derived from CDC seroprevalence studies in April (n = 16 596), May (n = 14 291), June (n = 14 159), July (n = 12 367), and August (n = 38 355), there were estimated medians of 46 910 006 (interquartile range [IQR], 38 192 705-60 814 748) SARS-CoV-2 infections, 28 122 752 (IQR, 23 014 957-36 438 592) symptomatic infections, 956 174 (IQR, 782 509-1 238 912) hospitalizations, and 304 915 (IQR, 248 253-395 296) deaths in the US through November 15, 2020. An estimated 14.3% (IQR, 11.6%-18.5%) of the US population were infected by SARS-CoV-2 as of mid-November 2020.

(continued)

Key Points

Question Accounting for underreporting, what is the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) disease burden in the US?

Findings In this cross-sectional study using data from public health surveillance of reported coronavirus disease 2019 cases and seroprevalence surveys, an estimated 46 910 006 SARS-CoV-2 infections, 28 122 752 symptomatic infections, 956 174 hospitalizations, and 304 915 deaths occurred in the US through November 15, 2020.

Meaning Findings of this study suggest that although more than 14% of the US population was infected with SARS-CoV-2 by mid-November, a substantial gap remains before herd immunity can be reached.

Author affiliations and article information are listed at the end of this article.

Open Access. This is an open access article distributed under the terms of the CC-BY-NC-ND License.

Abstract (continued)

CONCLUSIONS AND RELEVANCE The SARS-CoV-2 disease burden may be much larger than reported COVID-19 cases owing to underreporting. Even after adjusting for underreporting, a substantial gap remains between the estimated proportion of the population infected and the proportion infected required to reach herd immunity. Additional seroprevalence surveys are needed to monitor the pandemic, including after the introduction of safe and efficacious vaccines.

JAMA Network Open. 2021;4(1):e2033706. doi:10.1001/jamanetworkopen.2020.33706

Introduction

Estimates of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections are needed to understand how interventions can be titrated to reopen society.¹ Seroprevalence data provide an estimate of the proportion of the population who has been infected, and these data can be used for monitoring progress toward herd immunity. The Centers for Disease Control and Prevention (CDC) indicates that there have been 10 846 373 reported coronavirus disease 2019 (COVID-19) cases and 244 810 deaths in the US through November 15, 2020 with 1 037 962 reported cases within the last 7 days of that date (an average of 148 280 reported cases per day).² The number of reported cases is an underestimate of the true number of persons with infection because many persons with symptomatic COVID-19 either do not seek medical care or are not tested and therefore are not included in tallies of COVID-19 infections reported to public health authorities.³ Furthermore, an estimated 40% of individuals with SARS-CoV-2 infection are asymptomatic and unlikely to be tested and reported.⁴ In this study, data from seroprevalence surveys were used to adjust for underreporting of COVID-19 infections and thereby derive estimates of the number of SARS-CoV-2 infections, symptomatic infections, hospitalizations, and deaths in the US as of November 15, 2020.

Methods

Data were used from 2 types of cross-sectional seroprevalence studies that tested blood specimens for SARS-CoV-2 antibodies: community serosurveys with blood samples collected in April 2020 from randomly selected members of the general population⁵⁻⁹ and 5 seroprevalence surveys conducted by CDC that tested residual diagnostic blood specimens from large commercial laboratories.^{10,11} The CDC seroprevalence surveys used a SARS-CoV-2-specific enzyme-linked immunosorbent assay that has a reported specificity of more than 99% and sensitivity of 96% for detection of antibodies against the prefusion-stabilized form of SARS-CoV-2 spike protein.¹² This study did not involve primary data collection or patient interviews and was determined not to involve human subjects; therefore, per US Department of Health and Human Services regulation under the 45 CFR 46 Common Rule, the study was exempt from institutional review. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for cross-sectional studies.

Statistical Analysis

The first 4 CDC seroprevalence surveys were conducted with specimens from persons in 10 specific states (California, Connecticut, Florida, Louisiana, Minnesota, Missouri, New York, Pennsylvania, Utah, and Washington); the fifth seroprevalence survey was conducted nationwide. For each community serosurvey and for each state participating in the CDC seroprevalence surveys, the seroprevalence estimate was multiplied by the population to estimate the number of infections, and the proportion of reported infections was calculated by dividing the reported cases by the estimated infections. The infection underreporting multiplier is the inverse of the proportion of infections that were reported. The CDC estimated symptomatic proportion (60%) was used to derive the

symptomatic underreporting multiplier for each seroprevalence survey.⁴ The median and interquartile range (IQR) of the underreporting multipliers were calculated for the community serosurveys and each CDC seroprevalence survey.

Both external and internal validity of the multipliers derived from the CDC seroprevalence surveys were assessed. The median and IQR of the underreporting multipliers from the April community serosurveys were compared with the April CDC seroprevalence survey to assess external validity. Underreporting multipliers derived using data from the August CDC seroprevalence survey restricted to the 10 specific states were compared with multipliers derived using the data from the August survey from all states to assess internal validity.

The median underreporting multipliers derived from seroprevalence survey results from the 10 specific states in each of the 5 CDC seroprevalence surveys were used for 5 time periods for the pandemic in the US that aligned with dates of the 5 surveys conducted in 2020: January 21 to April 30, May 1 to May 31, June 1 to June 30, July 1 to July 31, and August 1 to November 15. The underreporting multipliers derived from the 10 specific states were multiplied by the number of COVID-19 cases reported in the US during the respective time periods to derive estimates of the number of persons with SARS-CoV-2 infection and symptomatic SARS-CoV-2 infection for each period, which were then summed to estimate the overall number of infections and symptomatic infections during the pandemic in the US.

The number of COVID-19 hospitalizations was estimated by multiplying the estimated symptomatic infections by the CDC estimated symptomatic case hospitalization ratio of 3.4%, and the number of COVID-19 deaths was estimated by multiplying the infection fatality ratio by 0.65%.⁴ A sensitivity analysis was conducted using the IQR of the underreporting multipliers (rather than the median) from each CDC seroprevalence survey to derive high and low estimates of the underreporting multipliers and high and low estimated numbers of SARS-CoV-2 infections and symptomatic SARS-CoV-2 infections.

Results

Community serosurveys were conducted in California, Florida, Georgia, Indiana, and New York with blood samples collected from 22 118 randomly selected participants from April 10 to May 3, 2020 (**Table 1**). The median infection underreporting multiplier for the 5 community serosurveys was 10.6× (IQR, 9.1× to 16.9×), and the median symptomatic underreporting multiplier was 6.4× (IQR, 5.5× to 10.1×).

Table 1. Seroprevalence Estimates From SARS-CoV-2 Community Surveys in the US, April 10 to May 3, 2020

Location	Population (millions) ^a	Sample size, No.	SARS-CoV-2 seroprevalence (date of specimen collection), %	Estimated No. of persons with SARS-CoV-2 infection	Cumulative reported COVID-19 cases (date) ^b	Underreporting multiplier	
						Infection	Symptomatic
California (Los Angeles) ⁵	9.4	863	4.6 (April 10-11)	434 031	13 069 (April 11)	33.2×	19.9×
Florida (Miami-Dade) ⁶	2.7	1800	6.0 (April 15-20)	163 016	9657 (April 20)	16.9×	10.1×
Georgia (DeKalb-Fulton) ⁷	1.8	696	2.5 (April 28-May 3)	45 581	5500 (May 3)	8.3×	5×
Indiana (statewide) ⁸	6.7	3658	2.8 (April 25-29)	188 502	17 756 (April 29)	10.6×	6.4×
New York (statewide) ⁹	19.4	15 101	14.0 (April 14-28)	2 723 498	299 691 (April 28)	9.1×	5.5×
All	NA	Total: 22 118	NA	NA	NA	Median: 10.6× (IQR, 9.1× to 16.9×)	Median: 6.4× (IQR, 5.5× to 10.1×)

Abbreviations: COVID-19, coronavirus disease 2019; IQR, interquartile range; NA, not applicable; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

^b Cumulative reported COVID-19 cases derived from county and state health department websites.

^a US Census Bureau quick factors (all-age population).

The 5 CDC seroprevalence surveys tested residual diagnostic blood specimens collected from March 23 to May 3 (survey 1), April 20 to June 7 (survey 2), May 19 to June 27 (survey 3), July 3 to July 17 (survey 4), and July 9 to August 12 (survey 5), 2020 (**Table 2**). Of the 10 specific states, 7 participated in 5 surveys, 8 in 4 surveys, 9 in 3 surveys, and 10 in 2 surveys. In the first 4 surveys, there was a mean of 1689 (range, 824-3264) residual blood specimens tested per state in each survey. Survey 1 included 16 596 blood specimens from the 10 specific states, survey 2 included 14 291 from 9 of the 10 specific states, survey 3 included 14 159 from 8 of the 10 specific states, and survey 4 included 12 367 from 7 of the 10 specific states. Survey 5 of the CDC seroprevalence included 38 355 blood specimens collected in all states except Hawaii, South Dakota, and Wyoming from July 9 to August 12, 2020; 40 states, including 9 of the 10 specific states, collected specimens from July 28 to August 12, 2020. There was a median of 850 specimens (range, 107-1005 specimens) collected per state in the 47 states.

Three of the 5 community serosurveys were conducted in states (California, Florida, and New York) that were among the 10 specific states. The median infection multiplier from the community serosurveys in April was 10.6× (IQR, 9.1× to 16.9×), and the median symptomatic multiplier from the community serosurveys for the same time was 6.4× (IQR, 5.5× to 10.1×), similar to estimates of the infection multiplier (10.8× [IQR, 9.4× to 11.7×]) and the symptomatic multiplier (6.5× [IQR, 5.6× to 7.0×]) from the first CDC seroprevalence survey conducted predominantly in April.

When looking at the multipliers across the other CDC seroprevalence surveys, the median infection multiplier in survey 2 conducted predominantly in May was 4.5× (IQR, 3.9× to 9.9×) and the symptomatic multiplier was 2.7× (IQR, 2.4× to 5.9×). In survey 3 conducted predominantly in June, the median infection multiplier was 5.4× (IQR, 3.7× to 6.5×) and the symptomatic multiplier was 3.2× (IQR, 2.2× to 3.9×). In survey 4 conducted in July, the median infection multiplier was 3.9× (IQR, 3.4× to 5.5×) and the symptomatic multiplier was 2.4× (IQR, 2.1× to 3.3×). For the 10 specific states in survey 5 conducted predominantly in August, the median infection multiplier was 3.2× (IQR, 2.5× to 4.0×) and the symptomatic multiplier was 1.9× (IQR, 1.5× to 2.4×). The infection and symptomatic multipliers for all 47 states that participated in survey 5 were 4.0× and 2.4×, respectively, similar to the median multipliers for the 10 specific states in survey 5. Across the 5 seroprevalence surveys, there was a decrease in the underreporting multipliers and an increase in uniformity of the underreporting multipliers between the states later in the pandemic.

We used the underreporting multipliers from the 5 CDC seroprevalence surveys to adjust public health surveillance data of reported COVID-19 reported cases for 5 time periods (**Table 3**). There were an estimated 46 910 006 SARS-CoV-2 infections, 28 122 752 symptomatic infections, 956 174 hospitalizations, and 304 915 deaths in the US through November 15, 2020 (**Figure**); within the last 7 days of that date, there were an estimated 3 321 478 infections. In the sensitivity analysis using the IQR of the multipliers, the ranges of estimates were 38 192 705 to 60 814 748 SARS-CoV-2 infections, 23 014 957 to 36 438 591 symptomatic infections, 782 509 to 1 238 912 hospitalizations, and 248 253 to 395 296 deaths. These data indicate that 14.3% (range, 11.6%-18.5%) of the US population (ie, 328 239 523) was infected with SARS-CoV-2 and 8.6% (range, 7.0%-11.1%) had a symptomatic infection, with an infection hospitalization ratio of 2.0% (range, 1.6%-2.5%) and symptomatic fatality ratio of 1.1% (range, 0.8%-1.3%) through November 15, 2020.

Discussion

By mid-November of 2020, a substantial proportion of the US population was infected with SARS-CoV-2. Given that the critical proportion for herd immunity for SARS-CoV-2 (ie, the proportion of the population that needs to have SARS-CoV-2 antibodies to disrupt ongoing transmission) is approximately 60% based on an estimated SARS-CoV-2 reproduction number of 2.5,⁴ the US population remains a long way from herd immunity even with millions of new infections each week. The number of estimated COVID-19 deaths is also remarkably more than the reported deaths in the

Table 2. Seroprevalence Estimates From Centers for Disease Control and Prevention SARS-CoV-2 Seroprevalence Surveys in the US

Location	Population (millions) ^a	Sample size, No.	SARS-CoV-2 seroprevalence (date of specimen collection), %	Estimated No. of persons with SARS-CoV-2 infection	Cumulative reported COVID-19 cases (date)	Underreporting multiplier	
						Infection	Symptomatic
Survey 1 ^b : March 23-May 3, 2020							
California (6 counties in San Francisco Bay Area)	6.5	1224	1.0 (April 23-27)	65 000	7151 (April 27)	9.1×	5.5×
Connecticut	3.6	1431	4.9 (April 26-May 3)	176 000	29 300 (May 3)	6×	3.6×
Florida (4 southern counties)	6.5	1742	1.8 (April 6-10)	117 000	10 525 (April 10)	11.1×	6.7×
Louisiana	4.6	1184	5.8 (April 1-8)	267 000	17 030 (April 8)	15.7×	9.4×
Minnesota (19 central countries)	3.8	1431	2.4 (April 30-May 12)	91 000	8800 (May 12)	10.3×	6.2×
Missouri	6.1	1882	2.7 (April 20-26)	162 000	6794 (April 26)	23.8×	14.3×
New York (5 counties in New York City)	9.3	2482	6.9 (March 23-April 1)	642 000	53 803 (April 1)	11.9×	7.2×
Pennsylvania (7 metropolitan Philadelphia counties)	4.9	824	3.2 (April 13-25)	157 000	22 987 (April 25)	6.8×	4.1×
Utah	2.2 adults	1132	2.2 (April 20-May 3)	47 000	4493 (May 3)	10.5×	6.3×
Washington (5 western counties)	4.4	3264	1.1 (March 23-April 1)	48 000	4308 (April 1)	11.1×	6.7×
Survey 2 ^c : April 20-June 7, 2020							
California (6 counties in San Francisco Bay Area)	6.5	1539	0.7 (May 19-27)	47 000	11 913 (May 27)	4.5×	2.7×
Connecticut	3.6	1800	5.2 (May 21-26)	185 000	41 234 (May 26)	3.9×	2.4×
Florida (4 southern counties)	6.5	1280	2.9 (April 20-24)	181 000	18 286 (April 24)	9.9×	5.9×
Minnesota (19 central counties)	3.8	1323	2.2 (May 25-June 7)	27 883	8400 (June 7)	3.3×	2.0×
Missouri	6.1	1831	2.8 (May 25-30)	171 000	12 956 (May 30)	13.2×	7.9×
New York (8 counties including New York City)	12.2	1116	23.2 (April 25-May 6)	2 833 000	281 670 (May 6)	10.1×	6.0×
Pennsylvania (12 metropolitan Philadelphia counties)	6.8	1743	3.6 (May 26-30)	245 000	56 318 (May 30)	4.4×	2.6×
Utah	2.2 adults	1940	1.1 (May 25-June 5)	25 000	11 330 (June 5)	2.2×	1.3×
Washington (19 western counties)	5.8	1719	2.1 (April 27-May 11)	122 000	13 098 (May 11)	9.3×	5.6×
Survey 3 ^d : May 19-June 27, 2020							
Connecticut	3.6	1798	6.3 (June 15-17)	223 000	45 347 (June 17)	4.9×	3.0×
Florida (4 southern counties)	6.5	1790	4.2 (May 19-27)	265 000	30 055 (May 27)	8.8×	5.3×
Minnesota (19 central counties)	3.8	1667	4.3 (June 15-27)	167 000	26 528 (June 27)	6.3×	3.8×
Missouri	6.1	1850	0.8 (June 15-20)	51 000	17 588 (June 20)	2.9×	1.7×
New York (8 counties including New York City)	12.2	1581	19.5 (June 15-21)	2 376 000	329 418 (June 21)	7.2×	4.3×
Pennsylvania (12 metropolitan Philadelphia counties)	6.8	1694	3.8 (June 14-20)	255 000	64 095 (June 20)	4.0×	2.4×
Utah	2.2 adults	1976	1.5 (June 15-24)	33 000	17 180 (June 24)	1.9×	1.2×
Washington (19 western counties)	5.8	1803	1.7 (June 15-20)	100 000	17 185 (June 20)	5.4×	3.5×
Survey 4 ^e : July 3-July 17, 2020							
Connecticut	3.6	1802	5.2 (July 3-6)	184 000	46 975 (July 6)	3.9×	2.4×
Minnesota (19 central counties)	3.8	1677	6.1 (July 6-17)	235 000	34 027 (July 17)	6.9×	4.1×
Missouri	6.1	1914	1.4 (July 5-9)	87 000	25 762 (July 9)	3.4×	2×
New York (8 counties including New York City)	12.2	1602	17.6 (July 7-11)	2 152 000	338 224 (July 11)	6.4×	3.8×
Pennsylvania (12 metropolitan Philadelphia counties)	6.8	1751	5 (July 6-11)	335 000	71 746 (July 11)	4.7×	2.8×
Utah	2.2 adults	1824	2.7 (July 6-15)	58 000	28 303 (July 15)	2×	1.2×
Washington (19 western counties)	5.8	1797	1.3 (July 6-7)	75 000	21 299 (July 7)	3.5×	2.1×
Survey 5 (10 specific states) ^f : July 16-August 12, 2020							
California	39.1	879	5.6 (July 30-August 5)	2 192 000	532 260	4.1×	2.5×
Connecticut	3.6	992	3.3 (July 30-August 3)	118 000	49 815	2.4×	1.4×
Florida	20.6	978	4.1 (July 31-August 3)	845 000	491 224	1.7×	1.0×
Louisiana	4.7	986	10.8 (July 16-August 12)	504 000	134 068	3.8×	2.3×
Minnesota	5.5	854	3.9 (July 29-August 12)	216 000	62 173	3.5×	2.1×

(continued)

Table 2. Seroprevalence Estimates From Centers for Disease Control and Prevention SARS-CoV-2 Seroprevalence Surveys in the US (continued)

Location	Population (millions) ^a	Sample size, No.	SARS-CoV-2 seroprevalence (date of specimen collection), %	Estimated No. of persons with SARS-CoV-2 infection	Cumulative reported COVID-19 cases (date)	Underreporting multiplier	
						Infection	Symptomatic
Missouri	6.1	979	2.8 (July 28-August 12)	171 000	59 932	2.9×	1.7×
New York	19.6	846	22.5 (July 31-August 11)	4 414 000	422 003	10.5×	6.3×
Pennsylvania	12.8	575	8.9 (July 31-August 11)	1 138 000	120 279	9.5×	5.7×
Utah	3.0	880	3.1 (July 30-August 11)	94 000	44 769	2.1×	1.3×
Washington	7.3	688	2.4 (July 29-August 11)	175 000	63 939	2.7×	1.6×
Survey 5 (47 states): July 9-August 12, 2020							
Nationwide (except Hawaii, South Dakota, Wyoming)	319.0	38 355	NA (July 9-August 12) ^a	19 592 000	4 880 149	4.0×	2.4×

Abbreviations: COVID-19, coronavirus disease 2019; IQR, interquartile range; NA, not applicable; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

^a US Census Bureau quick factors (all-age population).

^b Survey 1 summary total: 16 596; median infection underreporting multiplier: 10.8× (IQR, 9.4× to 11.7×); and median symptomatic underreporting multiplier: 6.5× (IQR, 5.6× to 7.0×).

^c Survey 2 summary total: 14 291; median infection underreporting multiplier: 4.5× (IQR, 3.9× to 9.9×); and median symptomatic underreporting multiplier: 2.7× (IQR, 2.4× to 5.9×).

^d Survey 3 summary total: 14 159; median infection underreporting multiplier: 5.4× (IQR, 3.7× to 6.5×); and median symptomatic underreporting multiplier: 3.2× (IQR, 2.2× to 3.9×).

^e Survey 4 summary total: 12 367; median infection underreporting multiplier: 3.9× (IQR, 3.4× to 5.5×); and median symptomatic underreporting multiplier: 2.4× (IQR, 2.1× to 3.3×).

^f Survey 5 (10 specific states) summary total: 8652; median infection underreporting multiplier: 3.2× (IQR, 2.5× to 4×); and median symptomatic underreporting multiplier: 1.9× (IQR, 1.5× to 2.4×).

^g Summary data available on website does not enable estimation of nationwide seroprevalence.

Table 3. Estimated SARS-CoV-2 Infections, Symptomatic Infections, Hospitalizations, and Deaths by Time Period, 2020

Time period	Reported cases, No.	Infection (symptomatic) underreporting multiplier	Estimated, No.			
			Infections	Symptomatic infections	Hospitalizations	Deaths
January 21-April 30	1 062 446	10.8× (6.5×)	11 474 417	6 905 899	234 801	74 584
May 1-May 31	725 234	4.5× (2.7×)	3 263 553	1 958 132	66 576	21 213
June 1-June 30	837 193	5.4× (3.2×)	4 520 842	2 679 018	91 087	29 385
July 1-July 31	1 917 706	3.9× (2.4×)	7 479 053	4 602 494	156 485	48 614
August 1-November 15	6 303 794	3.2× (1.9×)	20 172 141	11 977 209	407 225	131 119
Total	10 846 373	NA	46 910 006	28 122 752	956 174	304 915

Abbreviations: NA, not applicable; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

US through November 15, 2020, supporting the conclusion that approximately 35% of COVID-19 deaths are not reported.¹³

Reported COVID-19 cases do not represent the full SARS-CoV-2 disease burden.¹⁴ Case reports are dependent on patients seeking health care, availability and type of care (eg, telemedicine), and testing availability. Using data from seroprevalence surveys and surveillance is a common strategy for estimating underreporting and disease burden. This approach was used successfully throughout the 2009 novel influenza A pandemic.¹⁵ For example, in a simple model combining several sources of surveillance data, the CDC estimated 55 million symptomatic infections of 2009 pandemic influenza A (pH1N1) had occurred in the US by December 2009.¹⁶ This finding was consistent with an estimate from a seroprevalence survey for same period estimating 59 million infections (including asymptomatic infections).¹⁷

There are several methodologic issues, including time between infection, antibody development, antibody waning, and reporting of laboratory-confirmed infections, that must be considered when using seroprevalence surveys to derive underreporting multipliers to adjust surveillance data to estimate disease burden. In the New York statewide serosurvey, Rosenberg et al⁹ used the number of reported cases 1 week before the start of the survey to estimate the underreporting multiplier. In the Indiana statewide serosurvey, Menachemi et al⁸ included

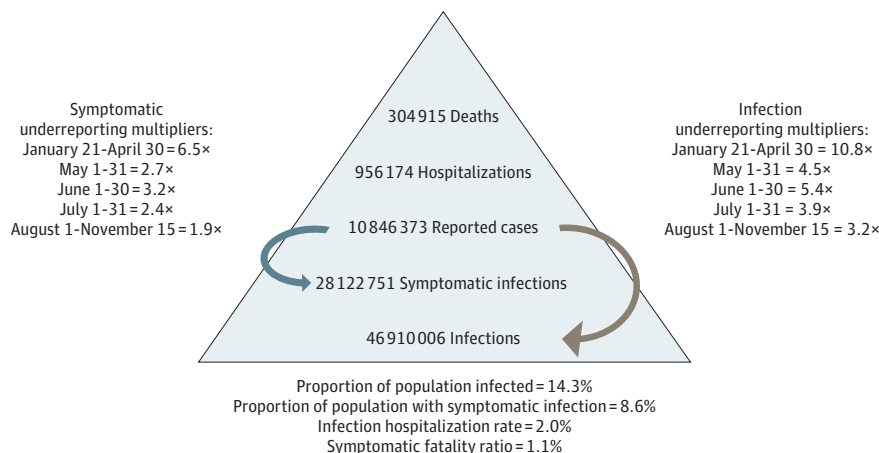
polymerase chain reaction test results to account for recent infections. Neither of these approaches (using an earlier reported number of cases to account for the antibody development lag or using current reported number with seroprevalence and polymerase chain reaction results to account for the antibody development lag and recent infections) accounts for the inherent delays in reporting cases. A sensitivity analysis of the CDC seroprevalence surveys suggested that using the number of reported cases at the end of the survey period provides a useful estimate of the underreporting multipliers, particularly early in the pandemic.¹⁰ Therefore, the number of reported cases on the last day of the seroprevalence survey was used in the COVID-19 disease burden estimation in this study.

Multipliers will change over time as the proportion of persons with infection tested, diagnosed, and reported changes, so a conservative approach was used in this disease burden estimation. In the 10 specific states that participated in the early and late CDC seroprevalence surveys, the median infection underreporting multiplier declined from 10.8× to 3.2×, and the symptomatic underreporting multiplier declined from 6.5× to 1.9×. Therefore, 5 different sets of multipliers based on the 5 CDC seroprevalence surveys were used in this estimation. Using the early (and larger) underreporting multipliers later in the pandemic would have resulted in an overestimation of the number of infected and symptomatic people and increased the estimated number of hospitalizations and deaths.

Limitations

This study has limitations. All seroprevalence surveys should be evaluated for selection bias.¹⁸ Conducting serosurveys with a random sampling design of the general population may be difficult to do in a pandemic, but a random sampling design of the general population yields a seroprevalence estimate least likely to be affected by selection bias. The CDC seroprevalence surveys used residual diagnostic blood specimens from large commercial diagnostic laboratories. The CDC seroprevalence surveys therefore rely on convenience sampling, which can contribute to selection bias; for example, more severely ill people (those hospitalized or visiting health care professionals) may be more likely to be tested. The external validity of the CDC seroprevalence surveys was evaluated by comparing the multipliers derived from the first seroprevalence survey with available contemporaneous community serosurveys that used random sampling designs of the general population that minimize selection bias. Of note, there were 3 states in common between the 5 community serosurveys and the 10 states that participated in the first CDC seroprevalence survey. The similarity of the median and IQR of the multipliers derived from the CDC seroprevalence survey to those derived from the community serosurveys (ie, symptomatic underreporting multiplier 6.4× [IQR, 5.5× to 10.1×]

Figure. Coronavirus Disease 2019 Burden in the US as of November 15, 2020



compared with 6.5× [IQR, 5.6× to 7.0×]) supports the conclusion of limited selection bias in the CDC seroprevalence surveys.

An additional limitation of the approach to estimate the COVID-19 disease burden in the US using underreporting multipliers derived from the 10 specific states is that seroprevalence results from these states may not be nationally representative. The internal validity of using the 10 specific states to represent national data was therefore evaluated using the fifth CDC seroprevalence survey. The national infection and symptomatic underreporting multipliers derived from the 47 states that participated in the fifth survey (4× and 2.4×, respectively) were similar to, although higher than, the median multipliers derived from the 10 specific states in the fifth survey (3.2× and 1.9×, respectively), suggesting that the multipliers derived from the 10 specific states are nationally representative and conservative (resulting in a lower disease burden).

Another limitation of this analysis is that the derivation of the multipliers used summary data from the seroprevalence surveys; the data sets of the seroprevalence surveys were not available. For each seroprevalence survey, medians of the multipliers of the states that participated in the survey were used to calculate the number of infections. Standard errors of the medians were too large to warrant derivation of CIs around the median. To explore the impact of the variation of the multipliers in each survey on the estimates of infections, a sensitivity analysis was conducted using the multipliers' IQR (rather than median multiplier). A key finding of the sensitivity analysis is that the multipliers' IQR became increasingly narrow as the pandemic progressed. This finding, coupled with the progressively diminishing multipliers with each passing month, probably reflects more widespread access to SARS-CoV-2 testing (and therefore smaller underreporting multipliers with less geographic variation) in later months. The narrowing of the multipliers' IQR later in the pandemic is particularly important during the final time period of this study (eg, the IQR for the symptomatic underreporting multiplier is 1.5× to 2.4× after August 1) because 58% of the almost 11 million reported COVID-19 infections in the US by mid-November occurred after August 1. The small variability in the underreporting multipliers derived from the 10 specific states in the August seroprevalence survey used to adjust 58% of the reported infections provides increased confidence in the overall disease burden estimates.

Finally, using seroprevalence surveys to derive estimates of underreporting of COVID-19 infections assumes that all persons infected with SARS-CoV-2 will have detectable antibodies at the time of the seroprevalence survey. The CDC seroprevalence surveys used a laboratory test with a high accuracy for detection of SARS-CoV-2 antibodies,¹² reducing the likelihood of failing to identify antibodies in persons infected. Furthermore, the seroprevalence results were adjusted for laboratory test characteristics.¹⁰ However, the CDC seroprevalence surveys may have underestimated the number of infected persons due to waning antibodies. Although the longevity of antibodies in persons infected with SARS-CoV-2 is not fully understood, waning antibodies have been reported in some persons infected,^{19,20} indicating that seroprevalence surveys may not detect some persons previously infected, particularly later in the pandemic, which would result in underestimation of the COVID-19 disease burden in this report.

Conclusions

In this cross-sectional study, estimates of underreporting multipliers were derived and combined with surveillance data to adjust reported surveillance data for underreporting. Results suggest that although more than 14% of the US population may have been infected with SARS-CoV-2 as of mid-November 2020, there remains a substantial gap between the estimated proportion of the population infected and the proportion infected that is required for herd immunity. Additional seroprevalence surveys are warranted to monitor the pandemic, including after the development of safe and efficacious vaccines.

ARTICLE INFORMATION

Accepted for Publication: November 24, 2020.

Published: January 5, 2021. doi:10.1001/jamanetworkopen.2020.33706

Open Access: This is an open access article distributed under the terms of the [CC-BY-NC-ND License](#). © 2021 Angulo FJ et al. *JAMA Network Open*.

Corresponding Author: Frederick J. Angulo, DVM, PhD, Medical Development and Scientific/Clinical Affairs, Pfizer Vaccines, 4024 NE Alameda St, Portland, OR 97212 (frederick.angulo@pfizer.com).

Author Affiliations: Medical Development and Scientific/Clinical Affairs, Pfizer Vaccines, Collegeville, Pennsylvania (Angulo, Swerdlow); Center for Observational and Real-World Evidence, Merck & Co Inc, Kenilworth, New Jersey (Finelli).

Author Contributions: Dr Angulo had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: All authors.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: All authors.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: All authors.

Obtained funding: Angulo.

Administrative, technical, or material support: Angulo, Finelli.

Supervision: All authors.

Conflict of Interest Disclosures: Dr Angulo reported being employed by Pfizer Vaccines and owning stock and stock options in Pfizer. Dr Finelli reported being employed by Merck & Co Inc and may own stock in the company. Dr Swerdlow reported being employed by Pfizer Vaccines and owning stock and stock options in Pfizer, as well as providing overviews of severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome epidemiology to a consulting firm for a minimal honorarium.

Funding/Support: This work was supported by Pfizer and Merck.

Role of the Funder/Sponsor: Pfizer and Merck had a role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; and approved the decision to submit the manuscript for publication.

REFERENCES

1. Angulo FJ, Finelli L, Swerdlow DL. Reopening society and the need for real-time assessment of COVID-19 at the community level. *JAMA*. 2020;323(22):2247-2248. doi:10.1001/jama.2020.7872
2. Centers for Disease Control and Prevention. Cases in the US. Accessed November 12, 2020. <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html>
3. Krantz SG, Rao ASRS. Level of underreporting including underdiagnosis before the first peak of COVID-19 in various countries: preliminary retrospective results based on wavelets and deterministic modeling. *Infect Control Hosp Epidemiol*. 2020;41(7):857-859. doi:10.1017/ice.2020.116
4. Centers for Disease Control and Prevention. COVID-19 pandemic planning scenarios. Updated July 1, 2020. Accessed November 12, 2020. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios-h.pdf>
5. 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(23):2425-2427. doi:10.1001/jama.2020.8279
6. Office of the Mayor, Miami-Dade County. Second round of COVID-19 community testing completed; Miami-Dade County and the University of Miami Miller School of Medicine announce initial findings. Published April 24, 2020. Accessed November 12, 2020. <https://www.miamidade.gov/releases/2020-04-24-sample-testing-results.asp>
7. 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(29):965-970. doi:10.15585/mmwr.mm6929e2
8. 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(29):960-964. doi:10.15585/mmwr.mm6929e1
9. 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. doi:10.1016/j.annepidem.2020.06.004

10. 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(12):1576-1586. doi:10.1001/jamainternmed.2020.4130
11. Centers for Disease Control and Prevention. Commercial laboratory seroprevalence survey data. Accessed November 12, 2020. <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/commercial-labs-interactive-serology-dashboard.html>
12. Freeman B, Lester S, Mills L, et al. Validation of a SARS-CoV-2 spike ELISA for use in contact investigations and serosurveillance. *bioRxiv*. Preprint posted online April 25, 2020. doi:10.1101/2020.04.24.057323
13. Woolf SH, Chapman DA, Sabo RT, Weinberger DM, Hill L, Taylor DDH. Excess deaths from COVID-19 and other causes, March-July 2020. *JAMA*. 2020;324(15):1562-1564. doi:10.1001/jama.2020.19545
14. Reed C, Chaves SS, Daily Kirley P, et al. Estimating influenza disease burden from population-based surveillance data in the United States. *PLoS One*. 2015;10(3):e0118369. doi:10.1371/journal.pone.0118369
15. Reed C, Angulo FJ, Swardlow DL, et al. Estimates of the prevalence of pandemic (H1N1) 2009, United States, April-July 2009. *Emerg Infect Dis*. 2009;15(12):2004-2007. doi:10.3201/eid1512.091413
16. Shrestha SS, Swardlow DL, Borse RH, et al. Estimating the burden of 2009 pandemic influenza A (H1N1) in the United States (April 2009-April 2010). *Clin Infect Dis*. 2011;52(52)(suppl 1):S75-S82. doi:10.1093/cid/ciq012
17. Reed C, Katz JM, Hancock K, Balish A, Fry AM; H1N1 Serosurvey Working Group. Prevalence of seropositivity to pandemic influenza A/H1N1 virus in the United States following the 2009 pandemic. *PLoS One*. 2012;7(10):e48187. doi:10.1371/journal.pone.0048187
18. Accorsi EK, Qiu X, Rumpler E, et al. How to detect and reduce potential sources of biases in epidemiologic studies of SARSCoV-2. *OSF Preprints*. doi:10.31219/osf.io/46am5
19. Patel MM, Thornburg NJ, Stubblefield WB, et al. Change in antibodies to SARS-CoV-2 over 60 days among health care personnel in Nashville, Tennessee. *JAMA*. 2020;324(17):1781-1782. doi:10.1001/jama.2020.18796
20. Ibarrondo FJ, Fulcher JA, Goodman-Meza D, et al. Rapid decay of anti-SARS-CoV-2 antibodies in persons with mild Covid-19. *N Engl J Med*. 2020;383(11):1085-1087. doi:10.1056/NEJMc2025179