Background: As of 28 July 2021, 60% of adults in the United States had been fully vaccinated against COVID-19, and more than 34 million cases had been reported. Given the uncertainty regarding undocumented infections, the population level of immunity against COVID-19 in the United States remains undetermined.

Objective: To estimate the population immunity, defined as the proportion of the population that is protected against SARS-CoV-2 infection due to prior infection or vaccination.

Design: Statistical and simulation modeling to estimate overall and age-specific population immunity.

Setting: United States.

Participants: Simulated age-stratified population representing U.S. demographic characteristics.

Measurements: The true number of SARS-CoV-2 infections in the United States was inferred from data on reported deaths using age-specific infection-fatality rates (IFRs). Taking into account the estimates for vaccine effectiveness and protection against reinfection, the overall population immunity was determined as the sum of protection levels in vaccinated persons and those who were previously infected but not vaccinated.

Methods

Model Overview

We used statistical and simulation modeling as well as reported deaths in the United States to estimate the total number of infections since the start of the pandemic. We then combined these with estimates of vaccination coverage and real-world vaccine effectiveness to calculate the overall and age-stratified population immunity in the United States. Simulation codes for reproducibility can be found at https://github.com/affans/c19popimmunity.

Data Sources

We used reported daily COVID-19 deaths for each age group (4, 5) in the United States from 21 January 2020 to 15 July 2021. Information on coverage of first and second vaccine doses, stratified by age groups, was obtained from the Centers for Disease Control and Prevention (CDC) (6). Infection-fatality rate (IFR) estimates and other data inputs were obtained from published literature and the CDC (Tables 1 to 3).

Estimating Total Infections

The total number of infections was estimated from reported daily COVID-19 deaths in the United States (4, 5) using 500 Monte Carlo simulations. Specifically, the reported death total $D_t$ on each day $t$ was used to impute the number of infections, using estimates of the IFR for COVID-19. The IFR, by definition, represents the ratio of the cumulative...
Deaths attributable to the disease to the total number of infections (both symptomatic and asymptomatic). To account for the uncertainty in IFR estimates, we sampled IFR values for model inputs from age-specific log-normal distributions, with the mean parameters informed by estimates from 3 sources: the CDC (7), an analysis of the early stages of the pandemic in China (8), and a multinational study combining death data from 45 countries and the results of 22 seroprevalence studies (9) (Table 1). The scale parameter of the log-normal distributions was set to 0.1, which allowed for the range of IFR estimates to be considered in the model sampling. The number of infections $N_f$ in each age group was then calculated as $D_t / \text{IFR}$.

### Analysis of the Lag Time Between Infection and Death

For each inferred infection, whether it resulted in recovery or death, we estimated the lag time $t'$ so that the infection occurred $t - t'$ days before the reported death on day $t$. The lag time was calculated as the sum of 2 independent intervals: the incubation period, which estimates the time from infection to onset of symptoms, and the time from symptom onset to death (Table 2). These intervals were sampled for each infection, accounting for the uncertainty of parameters in their corresponding distributions. The number of infections $N_f$ on any day $t$ was then calculated as the sum of all infections associated with reported deaths at time $t > \tau$ and all infection lag times $t'$ such that $t - t' = \tau$.

### Adjustment for Double-Counting

To prevent double-counting in our estimates of population immunity, we derived the proportion $P$ of previously infected persons who were vaccinated (Table 3). In the absence of information on prior infection for vaccinated persons, we probabilistically calculated the proportion $P$ for each age group via Monte Carlo replications. In each replication, we randomly assigned infections to a computational vector of population size for each age group, corresponding to the cumulative number of infections inferred from the process described earlier. Similarly, fully vaccinated persons were assigned randomly to a vector of the same size, according to the reported vaccination coverage of the age group. We then determined the overlap between these vectors to identify persons who were both vaccinated and infected. We iterated this process 1000 times to obtain the mean and the lower and upper bounds of the proportion $P$ and to determine the number of persons who were infected but not vaccinated. Given the CDC’s recommendation to delay vaccination by 3 months after infection, this part of the analysis was done for those who were infected 90 days before 15 July 2021 (16).

### Sensitivity Analyses

Given the variation in IFRs due to population heterogeneity and seroprevalence studies, we also considered IFR estimates from 2 other studies to inform our calculations of the overall population immunity (Table 1). First, we used IFR estimates derived from individual-level data for patients who died of COVID-19 in Hubei, China, during the early stages of the pandemic (8). We then used more recent IFR values that relied on age-specific data on COVID-19-attributable death from national-level seroprevalence surveys (9).

### Table 1. Age-Specific IFR Estimates Used in Analyses of Population Immunity

| Age Group, y | Mean IFR (95% CI), %* | Median IFR (Range), %† | IFR, %‡ |
|--------------|----------------------|------------------------|-------|
| 0–9          | 0.00161 (0.000185–0.0249) | 0.002 (0.001–0.006) | 0.002 |
| 10–19        | 0.00695 (0.00149–0.0502) | 0.003 (0.001–0.005) | 0.002 |
| 20–29        | 0.0309 (0.0138–0.0923) | 0.013 (0.004–0.025) | 0.05 |
| 30–39        | 0.0844 (0.0408–0.185) | 0.04 (0.012–0.077) | 0.05 |
| 40–49        | 0.161 (0.0746–0.323) | 0.12 (0.036–0.227) | 0.05 |
| 50–59        | 0.595 (0.34–1.28) | 0.32 (0.097–0.605) | 0.6 |
| 60–69        | 1.93 (1.11–3.89) | 1.07 (0.323–2.016) | 9 |
| 70–79        | 4.28 (2.45–8.44) | 3.2 (0.961–6.005) | 9 |
| ≥80          | 7.80 (3.80–13.3) | 8.29 (4.288–15.547) | 9 |

IFR = infection-fatality rate.
* From reference 8.
† From reference 9.
‡ From reference 7.

### Table 2. Model Parameters Used in Analyses of Naturally Acquired Protection Level and Vaccine-Induced Population Immunity

| Model Parameters | Mean | Distribution |
|------------------|------|--------------|
| Incubation period* | 5.2 d (SD, 3.8) | Log-normal |
| Time from symptom onset to death† | 16.1 d (SD, 8.7) | Gamma |
| Vaccine effectiveness 14 d after first dose‡ | | |
| Moderna | 51.4% (95% CI, 16.3%–71.8%) | Uniform |
| Pfizer-BioNTech | 51.4% (95% CI, 16.3%–71.8%) | Uniform |
| Vaccine effectiveness 14 d after second dose§ | | |
| Moderna | 93.3% (95% CI, 85.7%–97.4%) | Uniform |
| Pfizer-BioNTech | 86.1% (95% CI, 82.4%–89.1%) | Uniform |
| Naturally acquired protection against reinfection¶ | 80.5% (95% CI, 75.4%–84.5%) | Uniform |

* From reference 11.
† From reference 12.
‡ From reference 13.
§ From references 14 and 15.
¶ From references 6 and 10.
To account for uncertainty in vaccine effectiveness against infection and naturally acquired protection against reinfection, we performed 500 Monte Carlo simulations that accounted for confidence intervals around these estimates. The credible intervals (CrIs) around population immunity levels were obtained by performing 500 Monte Carlo simulations to account for uncertainty around input parameters.

Role of the Funding Source

The funding sources had no role in the design of the study, collection and analysis of the data, interpretation of the results, or the decision to publish the manuscript.

RESULTS

Primary Analysis

Using IFR estimates reported by the CDC (Table 1), we calculated that as of 15 July 2021, 114.9 (95% CrI, 103.2 to 127.4) million people had been infected with SARS-CoV-2 in the United States (Figure 1), which is 3.38 times higher than the reported number of cases (18). Accounting for an average of 80.5% protection against reinfection (6, 10) and vaccine-specific protections against infection in partially and fully vaccinated persons (8, 9), we estimated a mean overall population immunity of 62.0% (CrI, 58.4% to 66.4%) (Figure 2). Adults aged 65 years or older were estimated to have the highest level of immunity (77.2% [CrI, 76.2% to 78.6%]), followed by adults aged 18 to 49 years (72.7% [CrI, 67.6% to 79.0%]) and those aged 50 to 64 years (70.6% [CrI, 67.7% to 74.3%]). Immunity was substantially lower among adolescents aged 12 to 17 years (37.9% [CrI, 34.8% to 41.9%]) and children younger than 12 years (17.9% [CrI, 14.4% to 21.9%]) (Figure 2).

Sensitivity Analyses

When considering IFR values derived from the early stages of the pandemic in the model (8), we obtained 91.48 (CrI, 82.94 to 100.77) million infections, which is 26% lower than the number calculated by applying the CDC IFR estimates in the model (Appendix Figure 1, available at Annals.org). The highest and lowest immunity were associated with persons aged 65 years or older and children younger than 12 years, respectively. The overall population immunity was estimated to be 58.3% (CrI, 55.3% to 62.1%). However, when IFR values estimated from multiple seroprevalence studies were used (9), the total number of infections in the United States was 163.68 (CrI, 147.47 to 180.78) million, and the overall population immunity increased to 66.7% (CrI, 62.3% to 71.7%) (Appendix Figure 2, available at Annals.org). The patterns of age-specific immunity remained similar to those derived using IFR estimates from the CDC, with the highest immunity in persons aged 65 years or older (82.5% [CrI, 79.9% to 85.7%]) and the lowest immunity among children younger than 12 years (17.9% [CrI, 14.5% to 21.8%]).

These estimates were calculated using the reported mean vaccine effectiveness and the level of naturally acquired immunity.

Table 3. Proportion P of Previously Infected Persons Who Were Vaccinated and Vaccination Coverage by Age Group as of 20 June 2021

| Age Group, y | Mean P (Range)* | ≥1 Dose, % | Fully Vaccinated, % |
|--------------|-----------------|------------|---------------------|
| 0-11         | 0.997 (0.987-1) | 0.45       | 0.26                |
| 12-17        | 0.75 (0.65-0.85)| 35.0       | 25.1                |
| 18-49        | 0.53 (0.48-0.58)| 55.9       | 42.3                |
| 50-64        | 0.34 (0.25-0.42)| 74.0       | 66.0                |
| ≥65          | 0.21 (0.09-0.34)| 88.3       | 79.3                |

* Derived from Monte Carlo replications.

Figure 1. Estimated number of SARS-CoV-2 infections in the United States as of 15 July 2021, using estimates of IFR from the CDC (7).

The solid curve indicates the estimated cumulative number of infections, and the shaded area represents the 95% credible interval. The dotted curve represents the cumulative infections reported by the CDC as of 15 July 2021. CDC = Centers for Disease Control and Prevention; IFR = infection-fatality rate.
protection against reinfection. When we accounted for the confidence intervals of these protection levels and considered the 3 sources of IFR values, the average population immunity ranged from 52.1% to 71.0% (Figure 3).

**DISCUSSION**

Using age-stratified IFR estimates reported by the CDC, we calculated that as of 15 July 2021, 114.9 (CrI, 103.2 to 127.4) million SARS-CoV-2 infections had occurred in the United States. The range of our estimate overlaps the most recent CDC estimate of 120.2 million infections, which was derived from a statistical model applied to confirmed COVID-19 cases (19). When we considered all sources of IFR values in our analysis and used mean estimates of vaccine-induced and naturally acquired protection levels, the average population immunity ranged from 58.3% to 66.7%. Variants of SARS-CoV-2 with higher transmissibility, such as the Delta variant (20), will inevitably increase levels of naturally acquired immunity, but at the cost of potentially increased severe health outcomes. Improving daily vaccination rates would accelerate the increase in population immunity and reduce hospitalizations and deaths, even if infection occurs with lower vaccine effectiveness against immune-evading variants (21, 22). Given the large geographic variation and clustering in immunity, regions with low vaccination coverage will continue to experience local outbreaks and continuously disseminate infections to other regions, prolonging the COVID-19 pandemic in the United States.

Our study has limitations. First, our estimates of population immunity are based on the reported effectiveness of the Pfizer-BioNTech and Moderna vaccines, which constitute 96% of all doses administered in the United States (17). However, we note that the single-dose Johnson & Johnson vaccine authorized for emergency use in the United States has slightly lower effectiveness than the Pfizer-BioNTech and Moderna vaccines (23, 24). Second, we assumed that all persons with prior infection have a mean protection level of 80.5% against reinfection (6, 10) and performed sensitivity analyses on the estimated range (Figure 3). However, this protection can be age-dependent, decreasing to approximately 47% for persons aged 65 years or older (6). Third, to determine the proportion of previously infected persons who were also vaccinated, we assumed that the decision to be vaccinated was independent of prior infection. Fourth, vaccine-induced protection may be lower for immune-evading variants, such as the Delta variant, than for the original SARS-CoV-2 strain (22). Finally, we used daily deaths reported by the CDC to calculate the total number of infections. Recent estimates indicate that 24% of the total deaths attributable to COVID-19 in the United States have...
not been reported (25, 26), of which more than a third have occurred among persons aged 80 years or older. Although the unreported deaths may skew our estimates toward lower population immunity, IFR estimates are also subject to similar constraints, minimizing the potential bias in our estimates. Moreover, any underestimation of population immunity would further underscore the gap between population immunity and herd immunity thresholds that must be achieved to reverse the pandemic trajectory. In addition, vaccination and infection rates vary by location. For example, many states in the Northeast have achieved high vaccination coverage and low infection rates, whereas certain areas of the South and Midwest have relatively low vaccination and high infection rates.

Our study highlights the need to accelerate vaccination to prevent additional waves of COVID-19 and the evolution of novel variants and to shorten timelines for pandemic control in the United States. Adherence to nonpharmaceutical interventions, such as face masks and proactive testing, should be encouraged at least until population immunity is sufficiently high to contain the pandemic.

From York University, Toronto, Ontario, Canada (S.M.M.); Yale School of Public Health, New Haven, Connecticut (P.S., A.S., A.P.G.); and The University of Texas at Austin, Austin, Texas, and Santa Fe Institute, Santa Fe, New Mexico (L.A.M.).

Grant Support: Dr. Galvani received funding from NSF Expeditions grant 1918784, National Institutes of Health grant 1R01AI151176-01, National Science Foundation grant RAPID-2027755, and the Notsew Orm Sands Foundation. Dr. Moghadas was supported by the Canadian Institutes of Health Research (OV4 – 1590), and the Notsew Orm Sands Foundation. Dr. Galvani received funding from NSF Expeditions grant 1R01AI151176-01, National Science Foundation grant RAPID-2027755, and the Notsew Orm Sands Foundation. Dr. Moghadas was supported by the Canadian Institutes of Health Research (OV4 - 170643, COVID-19 Rapid Research) and the Natural Sciences and Engineering Research Council of Canada, Emerging Infectious Disease Modelling, Mathematics for Public Health grant.

Disclosures: Authors have reported no disclosures of interest. Forms can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M21-2721.

Reproducible Research Statement: Study protocol and data set: Available from Dr. Galvani (e-mail, alison.galvani@yale.edu). Statistical code: Available at https://github.com/affans/c19popimmunity.

Corresponding Author: Alison P. Galvani, PhD, Yale Center for Infectious Disease Modeling and Analysis, Yale School of Public Health, 135 College Street, Suite 200, New Haven, CT 06510; e-mail, alison.galvani@yale.edu.

Current author addresses and author contributions are available at Annals.org.

References
1. Cohen J. Covid-19 Herd Immunity Looks Like A Mirage, But Is Worth Pursuing. Forbes. 8 May 2021. Accessed at www.forbes.com/sites/joshuacohen/2021/05/08/covid-19-herd-immunity-looks-like-a-mirage-but-is-worth-pursuing on 28 July 2021.
2. Health Essentials. How Much of the Population Will Need to Be Vaccinated Until the Pandemic Is Over? 5 May 2021. Accessed at https://health.clevelandclinic.org/how-much-of-the-population-will-need-to-be-vaccinated-until-the-pandemic-is-over on 29 July 2021.
3. Russell RS. Herd immunity against COVID-19: more questions than answers [Editorial]. Viral Immunol. 2021;34:211-2. [PMID: 33999721] doi:10.1089/vim.2021.0075
4. GitHub. covid-19-data: an ongoing repository of data on coronavirus cases and deaths in the U.S. Accessed at https://github.com/nytimes/covid-19-data on 28 July 2021.
5. Centers for Disease Control and Prevention. Weekly Updates by Select Demographic and Geographic Characteristics. Accessed at www.cdc.gov/nchs/nvss/covid_weekly/index.htm on 28 July 2021.
6. Hansen CH, Michlmayr D, Gubbels SM, et al. Assessment of protection against reinfection with SARS-CoV-2 among 4 million PCR-tested individuals in Denmark in 2020: a population-level observational study. Lancet. 2021;397:1204-12. [PMID: 33743221] doi:10.1016/S0140-6736(21)00575-4
7. Centers for Disease Control and Prevention. COVID-19 Pandemic Planning Scenarios. Updated 19 March 2021. Accessed at www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html on 28 July 2021.
8. Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis.Lancet Infect Dis. 2020;20:669-77. [PMID: 32240634] doi:10.1016/S1473-3099(20)30243-7
9. 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. doi:10.1038/s41586-020-2918-0
10. Hall VJ, Fouldes S, Charlett A, et al; SIREN Study Group. SARS-CoV-2 infection rates of antibody-positive compared with antibody-negative health-care workers in England: a large, multicentre, prospective cohort study (SIREN). Lancet. 2021;397:1459-69. [PMID: 33844963] doi:10.1016/S0140-6736(21)00675-9
11. Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med. 2020;382:1199-207. [PMID: 31995857] doi:10.1056/NEJMoa2001316
12. Sanchez S, Lin YT, Xu C, et al. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. Emerg Infect Dis. 2020;26:1470-7. [PMID: 32255761] doi:10.3201/eid2607.200282
13. Chadick G, Tene L, Patalon T, et al. Assessment of effectiveness of 1 dose of BNT162b2 vaccine for SARS-CoV-2 infection 13 to 24 days after immunization. JAMA Netw Open. 2021;4:e2115985. [PMID: 34097044] doi:10.1001/jamanetworkopen.2021.15985
14. Centers for Disease Control and Prevention. Science Brief: COVID-19 Vaccines and Vaccination. Updated 27 July 2021. Accessed at www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/fully-vaccinated-people.html on 21 June 2021.
15. Pawlowski C, Lenehan P, Puranik A, et al. FDA-authorized mRNA COVID-19 vaccines are effective per real-world evidence synthesized across a multi-state health system. Med (N Y). 2021;2:979-992.e8. [PMID: 34223401] doi:10.1016/j.medj.2021.06.007
16. Centers for Disease Control and Prevention. Interim Clinical Considerations for Use of COVID-19 Vaccines Currently Authorized in the United States. Accessed at www.cdc.gov/vaccines/covid-19 /clinical-considerations/covid-19-vaccines-us.html on 21 June 2021.
17. Centers for Disease Control and Prevention. COVID Data Tracker. Accessed at https://covid.cdc.gov/covid-data-tracker on 21 June 2021.
18. GitHub. covid-19-data: an ongoing repository of data on coronavirus cases and deaths in the U.S. Accessed at https://github.com/nytimes/covid-19-data on 31 July 2021.
19. Centers for Disease Control and Prevention. Estimated COVID-19 Burden. Updated 27 July 2021. Accessed at www.cdc.gov/coronavirus/2019-ncov/cases-updates/burden.html on 30 July 2021.
20. Centers for Disease Control and Prevention. Delta Variant: What We Know About the Science. Updated 19 August 2021. Accessed at www.cdc.gov/coronavirus/2019-ncov/variants/delta-variant.html on 16 August 2021.
21. Barmann J. CDC confirms that viral loads in vaccinated people with delta may be infectious, so masks are necessary. SFist. 27 July 2021. Accessed at https://sfist.com/2021/07/27/cdc-confirms-that-viral-loads-in-vaccinated-people-with-delta-are-indistinguishable-from-unvaccinated on 29 July 2021.
22. Brown CM, Vostok J, Johnson H, et al. Outbreak of SARS-CoV-2 infections, including COVID-19 vaccine breakthrough infections, associated with large public gatherings – Barnstable County, Massachusetts, July 2021. MMWR Morb Mortal Wkly Rep. 2021;70:1059-62. [PMID: 34351882] doi:10.15585/mmwr.mm7031e2
23. Sadoff J, Gray G, Vandeboch A, et al; ENSEMBLE Study Group. Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. N Engl J Med. 2021;384:2187-201. [PMID: 33882225] doi:10.1056/NEJMoa2101544
24. U.S. Food and Drug Administration. Vaccines and Related Biological Products Advisory Committee February 26, 2021 Meeting Briefing Document- FDA. Accessed at www.fda.gov/media/146217 on 21 June 2021.
25. Iuliano AD, Chang HH, Patel NN, et al. Estimating under-recognized COVID-19 deaths, United States, March 2020–May 2021 using an excess mortality modelling approach. Lancet Reg Health Am. 2021:100019. [PMID: 34386789] doi:10.1016/j.lana.2021.100019
26. Moghadas SM, Galvani AP. The unrecognized death toll of COVID-19 in the United States. Lancet Reg Health Am. 2021:100033. [PMID: 34396362] doi:10.1016/j.lana.2021.100033
Current Author Addresses: Dr. Moghadas: Agent-Based Modelling Laboratory, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada.
Drs. Sah, Shoukat, and Galvani: Yale Center for Infectious Disease Modeling and Analysis, Yale School of Public Health, 135 College Street, Suite 200, New Haven, CT 06510.
Dr. Meyers: Department of Integrative Biology, The University of Texas at Austin, 1 University Station C0990, Austin, TX 78712.

Author Contributions: Conception and design: S.M. Moghadas, P. Sah, A. Shoukat, A.P. Galvani. Analysis and interpretation of the data: S.M. Moghadas, P. Sah, A. Shoukat, A.P. Galvani. Drafting of the article: S.M. Moghadas, P. Sah, A. Shoukat. Critical revision of the article for important intellectual content: S.M. Moghadas, P. Sah, A. Shoukat, A.P. Galvani. Final approval of the article: S.M. Moghadas, P. Sah, A. Shoukat, L.A. Meyers, A.P. Galvani. Statistical expertise: S.M. Moghadas, A. Shoukat. Obtaining of funding: S.M. Moghadas, A. Shoukat, A.P. Galvani. Collection and assembly of data: S.M. Moghadas, A. Shoukat.
Appendix Figure 1. Estimated number of infections and population immunity against SARS-CoV-2 infection in the United States as of 15 July 2021, using early estimates of infection-fatality rate (8).

Top. Cumulative number of infections (solid curve) and 95% credible intervals (shaded area). The dotted curve represents the cumulative infections reported by the Centers for Disease Control and Prevention. Bottom. Estimates of age-specific and overall population immunity. The bars represent 95% credible intervals.
Appendix Figure 2. Estimated number of infections and population immunity against SARS-CoV-2 infection in the United States as of 15 July 2021, using recent estimates of infection-fatality rate (9).

Top. Cumulative number of infections (solid curve) and 95% credible intervals (shaded area). The dotted curve represents the cumulative infections reported by the Centers for Disease Control and Prevention. Bottom. Estimates of age-specific and overall population immunity. The bars represent 95% credible intervals.