



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

SARS-CoV-2 active infection prevalence and seroprevalence in the adult population of St. Louis County



Charles W. Goss, PhD^{a,1,*}, Brett B. Maricque, PhD^{b,1}, Victoria V. Anwuri, MPH^c, Rachel E. Cohen, MPH^d, Kate Donaldson, MPH^d, Kimberly J. Johnson, PhD^e, William G. Powderly, MD^{c,f}, Kenneth B. Schechtman, PhD^a, Spring Schmidt, BA^d, Jeannette Jackson Thompson, PhD^g, Anne M. Trolard, MPH^c, Jinli Wang, MS^a, Elvin H. Geng, MD, MPH^f

^a Division of Biostatistics, Washington University School of Medicine, St. Louis, MO^b McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO^c Institute for Public Health, Washington University, St. Louis, MO^d St. Louis County Department of Public Health, Berkeley, MO^e Brown School, Washington University in St. Louis, St. Louis, MO^f Division of Infectious Diseases, Department of Medicine, Washington University School of Medicine, St. Louis, MO^g Health & Behavioral Risk Research Center, University of Missouri School of Medicine Dept. of Health Management & Informatics, Columbia, MO

ARTICLE INFO

Article history:

Received 30 October 2021

Revised 25 February 2022

Accepted 2 March 2022

Available online 8 March 2022

Keywords:

Covid-19

Probability survey

Seroprevalence

ABSTRACT

Background: The true prevalence of COVID-19 is difficult to estimate due to the absence of random population-based testing. To estimate current and past COVID-19 infection prevalence in a large urban area, we conducted a population-based survey in St. Louis County, Missouri.

Methods: The population-based survey of active infection (PCR) and seroprevalence (IgG antibodies) of adults (≥ 18 years) was conducted through random-digit dialing and targeted sampling of St. Louis County residents with oversampling of Black residents. Infection prevalence of residents was estimated using design-based and raking weighting.

Results: Between August 17 and October 24, 2020, 1245 residents completed a survey and underwent PCR testing; 1073 residents completed a survey and underwent PCR and IgG testing or self-reported results. Weighted prevalence estimates of residents with active infection were 1.9% (95% CI, 0.4%–3.3%) and 5.6% were ever infected (95% CI, 3.3%–8.0%). Overall infection hospitalization and fatality ratios were 4.9% and 1.4%, respectively.

Conclusions: Through October 2020, the percentage of residents that had ever been infected was relatively low. A markedly higher percentage of Black and other minorities compared to White residents were infected with COVID-19. The St. Louis region remained highly vulnerable to widespread infection in late 2020.

© 2022 Elsevier Inc. All rights reserved.

Introduction

Although the global COVID-19 pandemic has led to almost six million deaths globally [1] and over 900,000 in the United States

Conflicts of interest statement: This study was funded through the St. Louis County Department of Public Health (CARES Act) and was authorized by S. Schmidt; R. Cohen and K. Donaldson are also employees of this department.

* Corresponding author. Division of Biostatistics, Washington University School of Medicine, 660 S. Euclid Ave, St. Louis, MO 63110. Tel.: +1 314-362-4438.

E-mail address: cwgoss@wustl.edu (C.W. Goss).

¹ These authors contributed equally: Charles W. Goss, Brett B. Maricque

alone [2], fundamental epidemiological characteristics of the epidemic remain incompletely understood. In particular, in the first year of the pandemic clinical manifestations of disease revealed a remarkable spectrum of clinical severity (as many as 50% of infected people are asymptomatic, and 90% are not severely ill) [3–5], routine public health surveillance based on diagnosed cases in routine care vastly underestimate both the prevalence of active infection at any given time, and the cumulative incidence over time. The limited supply of tests (early in the pandemic) [6–9] and the variation in limitations by geography [10], time, and racial and socioeconomic groups [11–13] all make it inappropriate to use any

single “correction” factor. Without an estimate of true incidence and prevalence of infection, however, the actual burden of disease, and differences in burden by geography, racial groups, and socioeconomic status remain incompletely known. In addition, calculating the risk of hospitalization and death as well as the reach of municipal testing efforts all require the estimates of true prevalence of infection.

Few existing studies that attempt to assess the prevalence of active as well as cumulative COVID-19 infections [14–18] have been based on survey sampling methods in the United States (US). A number of studies estimating SARS-CoV-2 seroprevalence have been conducted globally [19,20], but the majority of these studies were conducted outside of the US and therefore are of uncertain relevance to the US epidemic. Most of the surveys that have been conducted in the US are based on convenience samples [but see 21,22] which are more likely to generate biased prevalence estimates compared to population-based probability samples.

St. Louis County is among the 50 most populated counties in the US and the most populated county in Missouri, with about one quarter identifying as Black and two thirds identifying as White [23]. Early on in the pandemic (April–May 2020) estimates of seroprevalence in the St. Louis metropolitan area [24] and the Missouri region [25] indicated that about 3% of individuals had been infected with SARS-CoV-2 at some point; however, these studies relied on convenience serum samples collected during routine health care and accurate estimates of infection that are critical to estimating key epidemiological measures (e.g., risk of death and hospitalization) and to informing public health decisions were largely lacking in Missouri during the early stages of the pandemic.

In the fall of 2020, we undertook a complex survey sampling study to assess the prevalence of active COVID-19 disease (as measured by PCR assay), and the prevalence of infection to that point (using both PCR and antibody assays) in St. Louis County. We stratified the sample to enable examination of prevalence by racial group and geography to identify variable disease burden in subpopulations in the region. We also used the estimate of true infection prevalence to assess regional penetration of testing (number with active infections compared to number tested) as well as hospitalization and fatality rates (number ever infected compared to cumulative hospitalizations and deaths). This formal application of a survey sampling approach offers additional insights into the extent of the COVID-19 pandemic, regional variability in burden of disease, as well as the impact of public health activities.

Methods

Study design and participants

Our target population consisted of St. Louis County, Missouri residents greater than or equal to 18 years that were not in long-term care facilities and who could be reached by landline or cell phone. We sampled individuals from this population for our study through random digit dialing. Telephone numbers were obtained from Marketing Systems Group (MSG), a commercial vendor who supplies contact information for a range of survey purposes (e.g., public health surveillance, marketing, political campaigns). We initially sought an equal number of Black and White respondents to obtain estimates of comparable precision in each group. Because ~24% of the county residents identify as Black, but the racial/ethnic identity of phone owners is not known a priori, we oversampled phone numbers linked to geographical areas where the majority of the residents are Black. Initially, fully random digit dialing numbers based on Behavioral Risk Factor Surveillance System [26] protocols were used to recruit St. Louis County residents for testing (eligible population ~777,067). Due to slow ramp-up of calling activities and time constraints, we later supplemented

the sample with listed cell and landline phone numbers provided by MSG. Use of listed numbers increases operational efficiency because the sample excludes non-working and business numbers and is also accompanied by socio-demographic information, but sacrifices a pure probability sample since mechanisms of selection into listed samples are not comprehensively known to the vendor.

Study procedures

Residents reached by phone were offered SARS-CoV-19 testing and invited to participate in a 15-minute survey. PCR and/or antibody testing was conducted at one of seven locations distributed throughout St. Louis County. The Health and Behavioral Risk Research Center at the University of Missouri Columbia was contracted to complete calling, phone interviews, and scheduling of tests. We offered gift card incentives for completing the survey and participating in PCR and/or antibody testing; we also provided roundtrip rides via Uber Health and cab vouchers to participants who needed transportation. The surveys administered were primarily based on the 2020 Behavioral Risk Factor Surveillance System and included newly-developed custom COVID-19 pandemic related questions.

SARS-CoV-2 PCR testing

At each of seven specimen collection sites, nasopharyngeal swabs were collected. PCR tests were offered to all participants in this study. Presence or absence of infection with SARS-CoV-2 was determined using FDA-approved RT-PCR assays on the Roche Cobas® platform [27]. All PCR testing was performed in CAP/CLIA certified laboratories.

SARS-CoV-2 IgG antibody testing

At five of seven specimen collection sites, 5 mL venous blood samples were collected from those participants that agreed to an IgG antibody test. Due to personnel limitations, during the last 2 weeks off the study antibody tests were only performed at two of the sites. Presence or absence of SARS-CoV-2 specific IgG antibodies was determined using FDA approved Abbott Architect chemiluminescent microparticle immunoassays [28]. All IgG antibody testing was performed in CAP/CLIA certified laboratories.

County surveillance data: rates of infection, hospitalization and fatality

To obtain county-level estimates of the number of individuals infected at the time of testing, we used the estimated prevalence from our study and St. Louis County census and testing data from the Missouri Department of Health & Senior Services. The percentage of individuals that were tested in the county was determined by tabulating the total number of residents that had PCR-positive tests at the midpoint of the study testing period (September 20, 2020) and compared these totals to the total expected individuals with active infection. We used + or – 10 days of PCR positivity around date of positive PCR test. Given the sensitivity of estimates to assumed duration of infection and the uncertainty of duration of PCR positivity in non-hospitalized patients, we explore effects on estimates assuming ± 7 and ± 14 days [29]. Repeat positive tests and congregate facilities (e.g., nursing homes) were removed from the testing dataset.

The infection hospitalization ratio (IHR) and infection fatality ratio (IFR) were determined by tabulating the cumulative number of COVID-19 related hospitalizations and deaths (excluding congregate facilities such as nursing homes) from March 2020 through the midpoint of our testing period (hospitalization data from the

Missouri Hospital Association, and death data from the Missouri Department of Health & Senior services). These totals were then divided by the number of residents estimated to have ever been infected to obtain IHR and IFR estimates. The lower and upper confidence bounds from the ever-infected prevalence estimates were multiplied by the county census estimates to obtain the denominator for the 95% confidence bounds for IHR and IFR. Estimates are reported for overall and stratified by age, race, and sex. Instances where race was unknown were imputed using a hot-deck imputation approach [30].

Statistical analysis

Our analytical approach entailed imputation and weighting of the survey data followed by a weighted frequency analysis. Sociodemographic variables used in the weighting process (sex, geographic area, age, race, ethnicity, education, income, smoking history, COVID-19 testing history, depression and adults with reduced contact due to epidemic) as well as other key variables from the survey were imputed using a hierarchical hot-deck imputation approach [30]. The probability sample ($n = 951$) was weighted using a combination of design-based weights as well as raking based on geodemographic variables. For the targeted non-probability sample ($n = 1363$) in addition to geodemographic weighting variables, a set of calibration variables (depression, smoking history, testing history, and reduced adult contact during pandemic) were used in the weighting process for this subset to better integrate these data with the probability-based sample and obtain the final weights for the total survey sample ($n = 2314$). Individuals with active infection were determined as the subset of individuals who completed the survey and took a PCR test ($n = 1245$). Individuals that completed a survey and took a PCR and IgG test or did not test with us but reported a recent positive or negative test ($n = 31$) as the main reason for not wanting to test in our study were included in our estimates of “ever infected” group (total $n = 1073$). Data for both outcomes were re-weighted using design-based weights and an iterative raking approach. More details regarding the imputation and weighting approach can be found in the supplement (see Supplementary Methods).

Active infection (PCR) and ever infected (PCR, IgG or self-reported [previous positive or negative test]) prevalence was estimated using Taylor series linearization [31] to adjust the variance estimates for survey weights. In addition to overall prevalence, we estimated the prevalence and associations between infection and key variables collected from the survey that included: age (18–39, 40–60, ≥ 60 years), race (White, Black/other), sex (F, M), county division (Central, Inner North, Outer North, South, West), prior COVID-19 testing history, mask wearing frequency (always vs. not always), COVID-19 related symptoms, income level ($< \$35,000$ vs. $\geq \$35,000$), and education (attended college vs. did not attend college). Results are reported as n (%), mean \pm SD, and weighted prevalence (%) with 95% confidence intervals. As sensitivity analyses, we estimated unweighted the unweighted prevalence for active (PCR), ever infected (PCR, IgG, self-reported), and ever infected with self-reported results excluded. All analyses were conducted using SAS version 9.4 (SAS Institute Inc., Cary, NC) and P -values $< .05$ were considered significant.

Results

Calling summary and cohort description

Over the course of the study, a total of 121,423 persons were reached and 4994 were eligible to be tested (Fig. 1). Of these, a total of 1245 took a PCR test and 1073 took both PCR and IgG tests or refused to test with us but reported having a recent prior negative or positive test result. Sixty three percent of respondents were female and 63%, 35%, and 2% of respondents were White, Black or

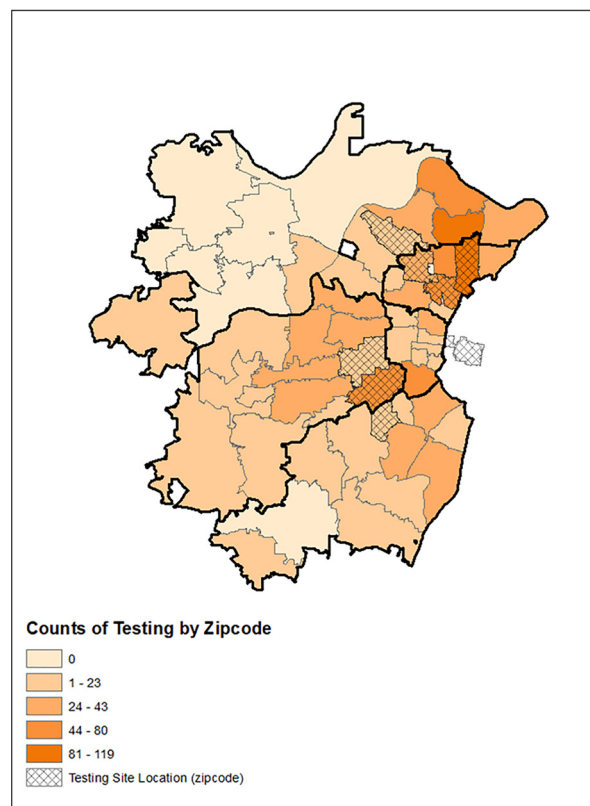


Fig. 1. Testing density by zip code and testing locations for residents who were tested in the study. “(For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)”

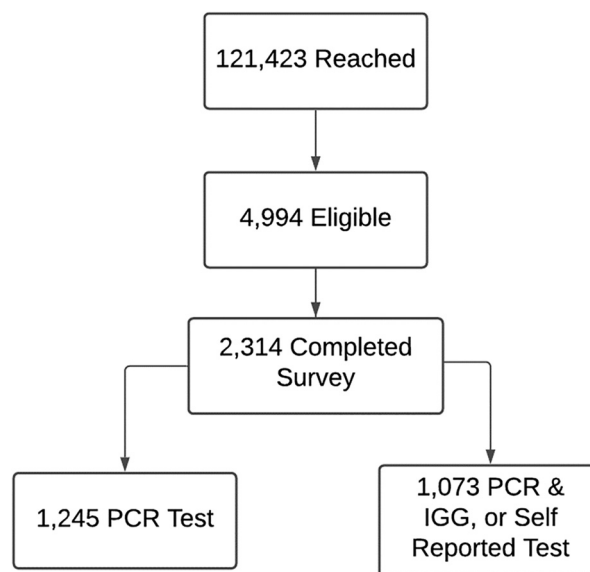


Fig. 2. Flow diagram of participants included in the analysis. “(For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)”

another minority group, respectively. The mean age for the cohort was 60 ± 15.5 years. The highest percentage of respondents were in the northern (51%) and western parts of the county (23%), with the southern (13%) and central parts (13%) of the county comprising less than 30% of the total (Fig. 2). Additional characteristics are provided in Table 1.

Table 1
Sociodemographic statistics for individuals who received a PCR test.

Characteristic	n (%) (Total N = 1245)
Age group	
18–39	164 (13.2%)
40–60	370 (29.7%)
>60	711 (57.1%)
Race	
Black/Other	463 (37.2%)
White	782 (62.8%)
Sex	
Female	786 (63.1%)
Male	459 (36.9%)
County Geographic Area	
Central	163 (13.1%)
Inner North	343 (27.6%)
Outer North	285 (22.9%)
South	166 (13.3%)
West	288 (23.1%)
Education	
Attended college	1025 (82.3%)
Did not attend college	220 (17.7%)
COVID-19 testing history	
No previous test	847 (68%)
Previous test	398 (32%)
COVID-19 Symptoms?	
No symptoms	696 (55.9%)
Symptoms	549 (44.1%)
Always wears mask?	
Always wears a mask	1098 (88.2%)
Does not always wear a mask	147 (11.8%)
Income <\$35,000?	
<\$35,000	320 (25.7%)
≥\$35,000	925 (74.3%)

Active infection rates as indicated by PCR testing

We estimated the prevalence of infection as detected by PCR assay (Table 2) to be 1.9% (95% CI, 0.9% to 3.9%) of the population in St. Louis County during the testing period (August 17 to October 24). Black and other minority residents had a prevalence of 4.1% which was 4.6 times higher compared to White residents (0.9%) ($P = .026$). Individuals that reported having been tested for COVID-19 previously (regardless of result) had over three times higher prevalence (3.7%) compared to those that did not have a previous test (1.0%) ($P = .057$). The prevalence in females was slightly higher than in males (2.6% vs. 1.0%) although not significantly different ($P = .207$). By age group, the highest prevalence was for the 18–39 years age group (2.0%), followed by 40–60 years group (2.4%) and residents >60 (1.2%) ($P = .797$). There was small (non-significant; $P = .312$) differences in prevalence estimates among different regions in the county with the highest prevalence in the northern areas (Inner North, 3.9%; Outer North, 2.5%), followed by South (2.0%) and West (1.0%); Central County did not have anyone that had a PCR positive test. Symptom status, mask wearing, education level and income level were not significantly associated with having active infection (Table 2). When we compared the weighted results to the unweighted results, the unweighted absolute prevalence and relative differences between groups were smaller compared to the weighted results but the direction was largely the same (Supplemental Table 1).

Current or past infection as indicated by PCR or IgG antibody testing results

We estimated (Table 3) the prevalence of ever having been infected (as measured by either a + antibody test, + PCR or self-reported positive test) of 5.6% (95% CI, 3.7% to 8.5%). Prevalence estimates for Black residents and other minorities was 10.5% which

Table 2
Weighted prevalence estimates based on testing for current (PCR) infection.

Characteristic	Weighted Prevalence (95% CI)	P-value
Overall	1.9% (0.9%–3.9%)	NA
Race		.026
White	0.9% (0.3%–2.5%)	
Black/Other	4.1% (1.6%–10.4%)	
Sex		.207
Male	1% (0.3%–3.3%)	
Female	2.6% (1.1%–6.2%)	
Age Group		.797
18–39	2% (0.4%–8.8%)	
40–60	2.4% (1.1%–4.9%)	
>60	1.2% (0.3%–4.5%)	
County Geographic Area		.312
Central*	0% (0%–2.3%)	
Inner North	3.9% (0.8%–16.4%)	
Outer North	2.5% (1%–6%)	
South	2% (0.6%–7.1%)	
West	1% (0.3%–3%)	
COVID-19 Testing History		.057
Previous test	3.7% (1.3%–9.9%)	
No previous test	1% (0.4%–2.6%)	
Household Income		.276
≥\$35,000	1.4% (0.7%–2.9%)	
<\$35,000	3.7% (0.8%–15.3%)	
COVID-19 Symptoms?		.681
No symptoms	1.6% (0.4%–5.7%)	
Symptoms	2.2% (1%–4.6%)	
Education		.182
Did not attend college	3.4% (1%–10.9%)	
Attended college	1.3% (0.6%–2.5%)	
Always wears mask?		.469
Does not always wear a mask	3% (0.9%–9.6%)	
Always wears a mask	1.7% (0.7%–4%)	

* The Central area had zero individuals with a PCR positive result. To obtain a group P-value a randomly selected individual was set to positive and given a weight of .001.

Table 3
Weighted prevalence estimates based on testing for current (PCR) or past infection (IgG antibody test or self-reported prior infection).

Characteristic	Prevalence (95% CI)	P-value
Overall	5.6% (3.7%–8.5%)	NA
Race		.008
White	3.6% (1.9%–6.5%)	
Black/Other	10.5% (6%–17.8%)	
Sex		.539
Male	4.8% (2.2%–9.9%)	
Female	6.3% (3.9%–10.2%)	
Age Group		.691
18–39	7% (3.1%–14.7%)	
40–60	5.3% (2.8%–9.7%)	
>60	4.6% (2.6%–8.1%)	
County Geographic Area		.460
Central	1.1% (0.2%–4.5%)	
Inner North	7% (3.2%–14.8%)	
Outer North	6.7% (3.6%–12.2%)	
South	7.4% (3.3%–15.9%)	
West	4.8% (1.8%–12.1%)	
COVID-19 Testing History		.020
Previous test	9.9% (6%–15.9%)	
No previous test	3.7% (1.8%–7.3%)	
Household Income		.590
≥\$35,000	5.3% (3.3%–8.6%)	
<\$35,000	6.9% (3.2%–14.4%)	
COVID-19 Symptoms?		.539
No symptoms	5% (2.6%–9.4%)	
Symptoms	6.4% (3.8%–10.6%)	
Education		.166
Did not attend college	8.4% (4%–16.7%)	
Attended college	4.5% (2.8%–7.2%)	
Always wears mask?		.331
Does not always wear a mask	9% (3.2%–22.9%)	
Always wears a mask	5.1% (3.3%–7.7%)	

Table 4
Estimated percentage of active infections tested by St. Louis County during the study.

# Of days from point of testing assumed to have active infection	# Of active infections based on routine county PCR testing	Percentage of true infections detected (95% CI)
± 13 days	3248	22.6% (10.8%, 47.8%)
± 10 days	2516	17.5% (8.3%, 37%)
± 7 days	1813	12.6% (6%, 26.7%)

The true active infection prevalence assumed to be 1.9% by PCR. The total number of individuals tested on September 20, 2020 (midpoint of the study testing period).

was almost three times higher compared to White residents (3.6%) ($P = .008$). Those reporting a prior test (any test result) had an estimated prevalence of 9.9% compared to 3.7% for those that reported not having tested previously ($P = .020$). The 18–39 and 40–60 age groups had elevated infection prevalence compared to greater than 60-year-old group but the difference was not significant ($P = .691$). Female prevalence was slightly elevated compared to males (6.3% vs. 4.8%) although not significant ($P = .539$). The northern and southern portions of the county had the greatest prevalence with prevalence estimates near or above 7%, and western and central areas of the county had lower prevalence with estimates of 4.8% and 1.1%, respectively ($P = .460$). Mask wearing, education, COVID-19 symptoms, and income level were not strongly associated with ever being infected (Table 3). Sensitivity analyses revealed that, in general, the unweighted and weighted results generated similar prevalence estimates (Supplemental Table 2). When we excluded those that reported prior positive or negative tests, we found that the overall unweighted prevalence estimates were just over 1% lower compared to the weighted estimates (overall prevalence, 4.4%); and while there were absolute differences in prevalence estimates for the different groups, the relative differences were generally similar (Supplemental Table 3).

County surveillance: rates of testing, hospitalization and fatality for infected residents

Based on prevalence of true active infections estimated in our survey (as indicated by PCR), we estimated that the percentage of true infections captured through routine testing ranged from 12.6%–22.6%, depending on how long individuals are assumed to have active infection (Table 4). The overall infection hospitalization ratio (IHR) was 4.9% (95% CI, 3.3%–7.5%) and the infection fatality ratio (IFR) was 1.4% (95% CI, 0.9%–2.2%). IHR and IFR estimates were similar between race and gender groups (Table 5); however, there were strong differences between age groups with the youngest age group (18–39 years) having the lowest IHR (1.5%) and IFR (~0%), the 40–60 years-old group having intermediate IHR (5%) and IFR (0.4%) estimates, and the oldest age group (>60 years) had more than double the nearest age group with 10.6% IHR estimate and 5% IFR estimate.

Discussion

Here, we report population-based prevalence estimates of active and ever-infected residents with SARS-CoV-2 in St. Louis County. During the study period (August 17th through October 24th 2020), 1.9% of St. Louis County residents had active SARS-CoV-2 infection and 5.6% of residents either had an active infection or had been infected in the past. These data indicate that there has been a substantial burden of disease in the St. Louis region and highlight the considerable portion of the population that was still at risk of SARS-CoV-2 infection at the time of this study in the fall of 2020. This is consistent with the rapid rise in cases in the region

during December 2020 and January 2021. Results from our study provided critical and timely information to St. Louis County health officials to make decisions regarding COVID policy in this region. These data are also consistent with other studies in regions that were relatively heavily affected that have led to similar estimates of disease burden around the same calendar time [32].

Although we recorded data on several factors thought to be associated with infection prevalence, race was the only factor with consistently significant associations for both active and ever-infected outcomes. Communities of color have been disproportionately affected by SARS-CoV-2 infection and COVID illness [10–12, 33]. Consistent with nationwide trends and local testing data, we observed a significant disparity in overall infection prevalence between Black residents and White residents, in which Black residents were nearly three times more likely to have been infected with SARS-CoV-2 than White residents. These disparities are consistent with longstanding race-based disparities in health [34–37], increased representation in essential work settings [38], and lack of access to effective COVID testing resources [10,11]. Age, sex, and income level have been shown to be associated with infection prevalence in a recent global meta-analysis of infection prevalence [19]. In our study we found numerical differences in the expected direction for two out of three of these factors (higher prevalence for younger ages and lower incomes), but none of the differences were significant. This suggests that our study was either underpowered to detect differences across these groups and/or that these factors are not important in the region that we studied. Nevertheless, the data presented here (esp. racial disparities) are critical for prioritizing resource allocation, modifying testing approaches and improving rates of testing in communities that have been hardest hit by COVID-19.

Results from the current study indicate that routine testing procedures in St. Louis likely detected one in five COVID-19 cases, demonstrating that the majority of infections were not identified as cases. A wide range of factors may contribute to this, including a substantial percentage of asymptomatic infections, lack of widespread surveillance testing, decentralized recommendations for when testing was indicated and accessibility issues. While COVID-19 is a disease with remarkably heterogeneous clinical manifestations [39], the low case-ascertainment rate and mild and asymptomatic infections underscores the challenges to public health. Low detection rates severely limit the value of public health strategies such as quarantine of detected cases or contact tracing [40], strategies whose impact is directly related to the proportion of cases that are rapidly diagnosed and acted upon. Indeed, in most regions in the United States, these measures were unable to keep up with and had minimal impact on the epidemic in part because of limited penetrance of testing. Rapid development of strategies to optimize testing, in addition to the test characteristics themselves, will be critical to the next pandemic response as well as this present one.

Consistent with previous estimates, our work suggests approximately 5% of persons with COVID-19 were hospitalized. Population-based studies in Indiana [41] and Connecticut [42] from early on in the pandemic (March through June 2020) report overall IHR estimates that bracket the estimate in our study with IHR estimates of 2.1% and 6.9%, respectively. This figure is relatively sensitive to the age structure of those infected (which is dependent on the age structure of the population) and our results show a strong increase in IHR as a function of age which mirrors these studies. The incidence of hospitalizations among all infected is an important quantity for anticipating the burden on the health system, and therefore for allocating and optimizing resources. The relationship is also of high interest for modeling exercises: because the threshold for hospitalization is relatively uniform, hospitalizations can be

Table 5

Cumulative COVID-19 hospitalization data from St. Louis County from March 1, 2020 through September 20, 2020 (midpoint of the study).

Characteristic	Total # infected at any point	Cumulative COVID-19 related hospitalizations (%)	IHR (95% CI)	Cumulative COVID-19 related deaths (%)	IFR (95% CI)
Overall	43,714 (100%)	2143 (100%)	4.9% (3.3%, 7.5%)	619 (100%)	1.4% (0.9%, 2.2%)
Race					
White	19,652 (45%)	847 (40%)	4.3% (2.4%, 8.0%)	301 (49%)	1.5% (0.8%, 2.8%)
Black/Other	23,750 (54%)	1296 (60%)	5.5% (3.2%, 9.5%)	318 (51%)	1.3% (0.8%, 2.3%)
Age group					
18–39	19,170 (44%)	282 (13%)	1.5% (0.7%, 3.3%)	3 (0%)	0% (0%, 0%)
40–60	13,853 (32%)	689 (32%)	5% (2.7%, 9.3%)	56 (9%)	0.4% (0.2%, 0.8%)
>60	11,098 (25%)	1172 (55%)	10.6% (6%, 18.8%)	560 (90%)	5% (2.9%, 9%)
Sex					
Male	17,249 (39%)	1040 (49%)	6% (2.9%, 12.8%)	307 (50%)	1.8% (0.9%, 3.8%)
Female	26,328 (60%)	1103 (51%)	4.2% (2.6%, 6.8%)	312 (50%)	1.2% (0.7%, 1.9%)

Total number infected derived from the combined PCR, IgG and self-reported prevalence estimate of 5.6% and county-level census estimates for the different characteristics.

used to estimate true burden of infection when diagnosed cases are known to capture only a fraction of true cases.

Infection fatality ratio (IFR) estimates from our study indicated that just over 1% of infected individuals died with a strong positive association between age and fatality rates with essentially 0% of the youngest group (18–39 years) dying and 5% of infections in the oldest age group (>60 years) resulted in death. As with the IHR estimates, our overall and positive association of IFR with age aligns with previously reported IFR estimates from early on in the pandemic [42]. These IFR estimates are also within the range of results from a recent mathematical modeling study in England [43]. Results from the current study to contribute to our understanding of the severity of disease once infected, and confirm that despite the heterogeneity of clinical manifestations, SARS-CoV-2 infection can have severe clinical consequences in a small but substantial proportion of individuals.

This study has several limitations. While our study was in progress, we modified the sampling approach to ensure an adequate number of tests for Black residents at a cost of decreasing representativeness by including listed phone numbers to obtain targeted samples. Furthermore, the relatively low response rates (especially among younger participants) could have created additional bias in our estimates. Closing some lesser-used testing sites and having limited evening and weekend testing availability may have deterred some individuals from testing. PCR-based prevalence estimates are limited by the timing of testing relative to infection. An infected individual is most likely to test positive in PCR-based assays within 4 days of the onset of infection. Finally, IgG-based estimates are limited by decreasing antibody assay sensitivity as the time between SARS-CoV-2 infection and testing increases [44].

Our sample did not include residents in long-term care facilities. This group is particularly vulnerable to COVID-19 infection and hospitalization or death which may have led to underestimates of IHR and IFR. For our “ever tested” prevalence estimates, we included self-reported results for a subset of 31 individuals that did not test during the study. The self-reported test results were not confirmed with outside sources (e.g., health care records) and potentially, and only a subset of the patients indicated this as the main reason for not testing during the study.

Conclusion

In St. Louis County through October 2020, there was a relatively low prevalence of residents who had ever been infected. Notably, a higher percentage of Black and other minority residents were currently or had ever been infected with COVID-19 compared to White residents. The St. Louis region remained highly vulnerable to widespread infection in late 2020. Routine testing did not identify most true cases, implying that mitigation strategies that depend on identification (e.g., contact tracing) were unlikely to be effective.

In understanding the epidemiology of COVID-19, researchers must triangulate among different sources of data to develop an accurate understanding of the transmission dynamics and burden when all methods have shortcomings, and this paper contributes to that understanding.

Acknowledgments

The authors would like to acknowledge interviewers and other staff of the Health and Behavioral Risk Research Center.

This work was supported by the St. Louis County Department of Public Health through the 2020 Coronavirus Aid, Relief, and Economic Security (CARES) Act.

References

- [1] Johns Hopkins University, COVID-19 Global Map [<https://coronavirus.jhu.edu/map.html>].
- [2] Johns Hopkins University, COVID-19 U.S. Map. <https://coronavirus.jhu.edu/us-map>.
- [3] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395(10223):497–506.
- [4] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020;395(10223):507–13.
- [5] Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis* 2020;20(6):669–77.
- [6] Coronavirus Testing Falls Woefully Short as Trump Seeks to Reopen. U.S. - The New York Times.
- [7] RNA Extraction Kits for COVID-19 Tests Are in Short Supply in US | The Scientist Magazine®.
- [8] Péré H, Podglajen I, Wack M, Flamarion E, Mirault T, Goudot G, et al. Nasal swab sampling for SARS-CoV-2: a convenient alternative in times of nasopharyngeal swab shortage. *J Clin Microbiol* 2020;58(6):58.
- [9] van Kasteren PB, van der Veer B, van den Brink S, Wijsman L, de Jonge J, van den Brandt A, et al. Comparison of seven commercial RT-PCR diagnostic kits for COVID-19. *J Clin Virol* 2020;128:104412.
- [10] Hu T, Yue H, Wang C, She B, Ye X, Liu R, et al. Racial segregation, testing site access, and COVID-19 incidence rate in Massachusetts, USA. *Int J Environ Res Public Health* 2020;17(24):17.
- [11] Escobar GJ, Adams AS, Liu VX, Soltesz L, Chen Y-FI, Parodi SM, et al. Racial disparities in COVID-19 testing and outcomes: retrospective cohort study in an integrated health system. *Ann. Intern. Med.* 2021;174(6).
- [12] Mody A, Pfeifauf K, Bradley C, Fox B, Hlatshwayo MG, Ross W, et al. Understanding drivers of COVID-19 racial disparities: a population-level analysis of COVID-19 testing among black and white populations. *Clin Infect Dis* 2020;73(9):E2921–31.
- [13] Rentsch CT, Kidwai-Khan F, Tate JP, Park LS, Jr JTK, Skanderson M, et al. Patterns of COVID-19 testing and mortality by race and ethnicity among United States veterans: a nationwide cohort study. *PLoS Med.* 2020;17(9):e1003379.
- [14] Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS ONE* 2020;15(3):e0230405.
- [15] Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med* 2020;26(6):855–60.

- [16] Kucharski AJ, Klepac P, Conlan AJK, Kissler SM, Tang ML, Fry H, et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *Lancet Infect Dis* 2020;20(10):1151–60.
- [17] Kyrychko YN, Blyuss KB, Brovchenko I. Mathematical modelling of the dynamics and containment of COVID-19 in Ukraine. *Sci Rep* 2020;10(1):19662.
- [18] Tsay C, Lejarza F, Stadtherr MA, Baldea M. Modeling, state estimation, and optimal control for the US COVID-19 outbreak. *Sci Rep* 2020;10(1):10711.
- [19] Rostami A, Sepidarkish M, Fazlzadeh A, Mokdad AH, Sattarnezhad A, Esfandiyari S, et al. Update on SARS-CoV-2 seroprevalence: regional and worldwide. *Clin Microbiol Infect* 2021;27(12):1762–71.
- [20] Rostami A, Sepidarkish M, Leeflang MMG, Riahi SM, Shiadeh MN, Esfandiyari S, et al. SARS-CoV-2 seroprevalence worldwide: a systematic review and meta-analysis. *Clin Microbiol Infect* 2021;27(3):331–40.
- [21] Bruckner TA, Parker DM, Bartell SM, Vieira VM, Khan S, Noymner A, et al. Estimated seroprevalence of SARS-CoV-2 antibodies among adults in Orange County, California. *Sci Rep* 2021;11(1).
- [22] Sood N, Simon P, Ebner P, Eichner D, Reynolds J, Bendavid E, 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):2425–7.
- [23] The United States Census Bureau. URL: <https://www.census.gov>.
- [24] Smith BK, Janowski AB, Danis JE, Harvey IB, Zhao H, Dai YN, et al. Seroprevalence of SARS-CoV-2 Antibodies in Children and Adults in St. Louis, Missouri, USA. *mSphere* 2021;6(1).
- [25] Havers FP, Reed C, Lim T, Montgomery JM, Klena JD, Hall AJ, 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.
- [26] CDC - Behavioral Risk Factor Surveillance System (BRFSS). URL: <https://www.cdc.gov/brfss/>
- [27] Roche's cobas SARS-CoV-2 Test to detect novel coronavirus receives FDA emergency use authorization and is available in markets accepting the CE mark.
- [28] Bryan A, Pepper G, Wener MH, Fink SL, Morishima C, Chaudhary A, et al. Performance characteristics of the Abbott architect SARS-CoV-2 IgG Assay and Seroprevalence in Boise, Idaho. *J Clin Microbiol* 2020;58(8).
- [29] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395(10229):1054–62.
- [30] Mukhopadhyay P.K.: Survey data imputation with PROC SURVEYIMPUTE.26.
- [31] Woodruff RS. A simple method for approximating the variance of a complicated estimate. *J Am Stat Assoc* 1971;66(334):411.
- [32] Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-Olmeda M, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet* 2020;396(10250):535–44.
- [33] Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/coronavirus/2019-ncov/community>.
- [34] Bibbins-Domingo K. This time must be different: disparities during the COVID-19 pandemic. *Ann Intern Med* 2020;173(3):233–4.
- [35] Jones CP. Invited commentary: “race,” racism, and the practice of epidemiology. *Am J Epidemiol* 2001;154(4) 299–304; discussion 305–306.
- [36] Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and Racial/Ethnic disparities. *JAMA* 2020;323(24):2466–7.
- [37] Chowkwanyun M, Reed AL. Racial health disparities and Covid-19 - caution and context. *N Engl J Med* 2020;383(3):201–3.
- [38] Rogers TN, Rogers CR, VanSant-Webb E, Gu LY, Yan B, Qeadan F. Racial disparities in COVID-19 mortality among essential workers in the United States. *World Med Health Policy* 2020;12(3):311–27.
- [39] Kasper MR, Geibe JR, Sears CL, Riegodedios AJ, Luse T, Thun AMV, et al. An outbreak of Covid-19 on an aircraft carrier. *N Engl J Med* 2020;384(10):976.
- [40] Aleta A, Martín-Corral D, Pastore y Piontti A, Ajelli M, Litvinova M, Chinazzi M, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat Hum Behav* 2020;4(9):964–71.
- [41] Menachemi N, Dixon BE, Woos-Kaloustian KK, Yiannoutsos CT, Halverson PK. How many SARS-CoV-2-infected people require hospitalization? Using random sample testing to better inform preparedness efforts. *J Public Health Manag Pract* 2021;27(3):246–50.
- [42] Mahajan S, Caraballo C, Li SX, Dong Y, Chen L, Huston SK, et al. SARS-CoV-2 infection hospitalization rate and infection fatality rate among the non-congregate population in Connecticut. *Am J Med* 2021;134(6):812–816.e2.
- [43] Chen S, Flegg JA, White LJ, Aguas R. Levels of SARS-CoV-2 population exposure are considerably higher than suggested by seroprevalence surveys. *PLoS Comput Biol* 2021;17(9):e1009436.
- [44] Peluso MJ, Takahashi S, Hakim J, Kelly JD, Torres L, Iyer NS, et al. SARS-CoV-2 antibody magnitude and detectability are driven by disease severity, timing, and assay. *Sci Adv* 2021;7(31):eabh3409.