

SUPPLEMENTAL APPENDIX

Seroprevalence of SARS-CoV-2-Specific IgG Antibodies Among Adults Living in Connecticut Between March 1 and June 1, 2020: Post-Infection Prevalence (PIP) Study

List of Contents

eMethods 1. Calculation of sample size.

eMethods 2. Details of data source and participant enrollment.

eMethods 3. Details of weighting of study sample.

eFigure 1. Flow chart showing sample selection for the state-level estimate.

eFigure 2. Distribution of the timing of the blood draws between June 10 and July 6, 2020 for the state-level estimate.

eTable 1. Comparison of demographics (age-group, race/ethnicity, and geographic region) of those who completed the survey with those who completed the survey and the serology test.

eTable 2. Comparison of prevalence of symptomatic illness, risk factors for possible exposure, and adherence to social-distancing behaviors since March 1, 2020, among those who completed only the survey with those who completed the survey and the serology test.

eTable 3. Unweighted and weighted seroprevalence of SARS-CoV-2-specific IgG antibodies among adults in Connecticut, by sociodemographic and clinical subgroups.

eMethods 1. Calculation of sample size.

In designing the study, sample size calculation has an important role to detect an effect and to achieve the desired precision in estimates of the prevalence of SARS-CoV-2. Early research has shown that SARS-CoV-2 is more prevalent in minority groups and in older populations, which needs to be accounted for while estimating sample size requirements. For the purposes of baselining, we estimated having to recruit a total of 1460 respondents to go through testing. The assumptions underlying the state-level estimate is shown below:

- *For state-level estimates:* required sample size = 609 assuming state-level prevalence of 10%, precision requirement of 2%, and a confidence level of 90%.

Currently, we are in phase 2 of the project where we are oversampling Black and Hispanic subpopulations. These oversamples have not been folded into the state sample yet since data is still being collected.

eMethods 2. Details of data source and participant enrollment.

Data source: To achieve the required sample size for generating state-level estimates of seroprevalence, Gallup used a dual-frame Random Digit Dial (RDD)⁹ methodology to select a random sample of adults residing in non-congregate settings (i.e. excluding individuals living in long-term care facilities, assisted living facilities, nursing homes, and prisons or jails), aged 18 years and older, living in Connecticut. Briefly, this involves drawing a random sample of landline and cell phone numbers from among all potential landline and cell phone numbers with valid area codes assigned to Connecticut. The source of the telephone numbers was the North American Numbering Plan that specifies which area codes and exchanges are assigned to Connecticut.²² A simple random sample of telephone numbers was drawn from the landline and cellphone frame with a goal to complete roughly 30% of the sample from the landline frame and the remaining 70% from the cellphone frame. The proportion in which landline and cell phone interviews were completed was designed to maximize coverage and representation of those with access to landline telephone only, cell phone only, and those having access to both landline and cell phones. Interviews were available to be conducted in either English or Spanish depending on the preference of the respondent.

To minimize the under coverage resulting from residents having cell phone numbers that are not assigned to Connecticut, Gallup supplemented the RDD sample with a random sample of listed cellular numbers of residents with an address in Connecticut but a cell phone number with area code that is not assigned to Connecticut. This was done to ensure that residents of Connecticut who moved from another area in the country and ported their cellphone number were still eligible to participate in this survey. The combination of the different sample types was designed to ensure high coverage and representativeness of the state population and subgroups in an efficient manner.

Currently, we are in phase 2 of the project where we are oversampling Black and Hispanic subpopulations. To ensure adequate sample sizes for subpopulations of interest (Black and Hispanic), additional RDD samples were drawn from the cell phone frame, with numbers assigned to rate centers or geographies to which cell phone numbers are associated. These oversamples have not been folded into the state sample yet since data is still being collected.

Participant enrollment: Within households reached by landline, a random adult 18 or older from among all eligible adults in that household was selected using the “next birthday” method. For those reached via cell phone, we confirmed that we are speaking with an adult 18 years or older, living in Connecticut, and for the oversample that they meet the race and ethnicity requirements. No respondent selection was done in the case of cell phone respondents, where one user per number was assumed.

To maximize the opportunity for the inclusion of harder to reach individuals, a multi-call design was implemented whereby up to five (5) attempts were made to each randomly selected telephone number, spread over different days of the week, including the weekend and different times of day to ensure we achieved a representative sample of adults. Scheduled “call-backs” to complete an interview with an eligible respondent at a later time or date agreed upon between the interviewer and respondent were also executed.

If after 5 attempts, we were unable to make a human contact or encountered a refusal to participate, the number was retired. New sample replicates were released as numbers were retired and the selection of the respondent was started afresh with the new number.

For the state-level sample, out of the 7272 participants contacted, 1842 were via landline and 5430 via cell phone. Of the 727 participants who completed the survey, 218 were contacted via landline and 509 via cell phone. Out of 505 participants who completed the survey and the blood draw, 183 were contacted via landline and 342 via cell phone.

Response rate: We contacted a total of 7272 respondents at the state-level between June 4 and June 23 of 2020, and successfully completed 727 interviews resulting in a combined dual-frame Response Rate of 7.3%. Response rate calculated was RR3 as specified by the American Association of Public Opinion Research standard definition:

https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf, page 61.

Response rate formula was defined as ‘completes’ divided by ‘eligible’ plus ‘presumed eligible’ participants, and as such, the denominator contained more individuals than those actually contacted. Among those who we could not contact, we estimate a certain proportion to be eligible based on what we observed among those we contacted. So, the actual denominator is more than 7272. We estimate, had we been able to continue calling indefinitely, we would have found a total of 9959 individuals eligible to participate in the study, given the number of phone numbers we were dialing.

eMethods 3. Details of weighting of study sample.

The sample data were weighted to minimize bias and thereby ensure that the results were representative of the target population of all adults in the state of Connecticut. The base weight (or selection probability weight) assigned to each completed survey was derived as the inverse of the probability of selection of that respondent in the sample. As described in the section on Sampling Methodology, a dual-frame (landline and cell phone) sample design was employed for this study. In addition, Gallup supplemented the RDD sample with a random sample of listed cellular numbers (hereafter referred to as the listed sample) of residents with an address in Connecticut but a cell phone number with area code that is not assigned to Connecticut.

The calculation of selection probabilities for each respondent was done taking into consideration his/her access to landline and/or cell phones. If someone had access to landline only, then selection probability was based on the landline frame count and the sample size drawn from that frame. If someone had access to cell phone only, similar calculations based on cell frame count and sample size were used. For dual users (with access to both landline and cell phones), their chances of selection from both frames were taken into consideration. If a respondent from the listed cell sample had access to landline phones, the selection probability calculation for that respondent included his/her chances of selection from the landline frame also. In other words, the selection probability was dependent on all possible ways that person could be selected in the sample and not based on how that respondent was actually selected in this sample.

The next step involved post-stratification weighting adjustments to account for survey non-response and to match the weighted sample estimates to known population characteristics for the state of Connecticut. Post-stratification weighting was carried out using raking (or Iterative Proportional Fitting) procedures to adjust for demographic variables such as age, gender, race/ethnicity and education. The different categories for these variables that were used for post-stratification were as follows: age-group (18-39, 40-49, 50-59, 60-69 and 70+), gender (male, female) race/ethnicity (Hispanic, non-Hispanic White/Other, non-Hispanic Black), and education (high school or less, some college/no 4 year college degree, college grad, post-graduate degree). The distribution of the final weights (obtained after the post-stratification weighting) were examined and some trimming of weights (5th percentile at the bottom and 97th percentile at the top) was carried out to avoid extreme weights and thereby reduce the effect of such weights on sampling variance.

The margin of error (MOE) for a simple random sample, also known as precision for estimating the unknown population proportion 'P' at the 90% confidence level can be derived based on the following formula:

$$\text{MOE} = \pm 1.645 * \sqrt{P * (1 - P) / n}$$

where 'n' is the sample size (i.e. the number of individuals who completed the survey and the blood test).

The weighting process resulted in unequal weights to correct for household selection with unequal probability of selection and non-response adjustments through post-stratification weighting. This introduces a design effect that needs to be taken into account while computing precision of estimates. The design effect is defined as the ratio of the design-based sample variance to the sample variance obtained from a simple random sample of the same size. Overall study design effect as estimated by the Kish₁ approximation equals 1.78. In addition, the design

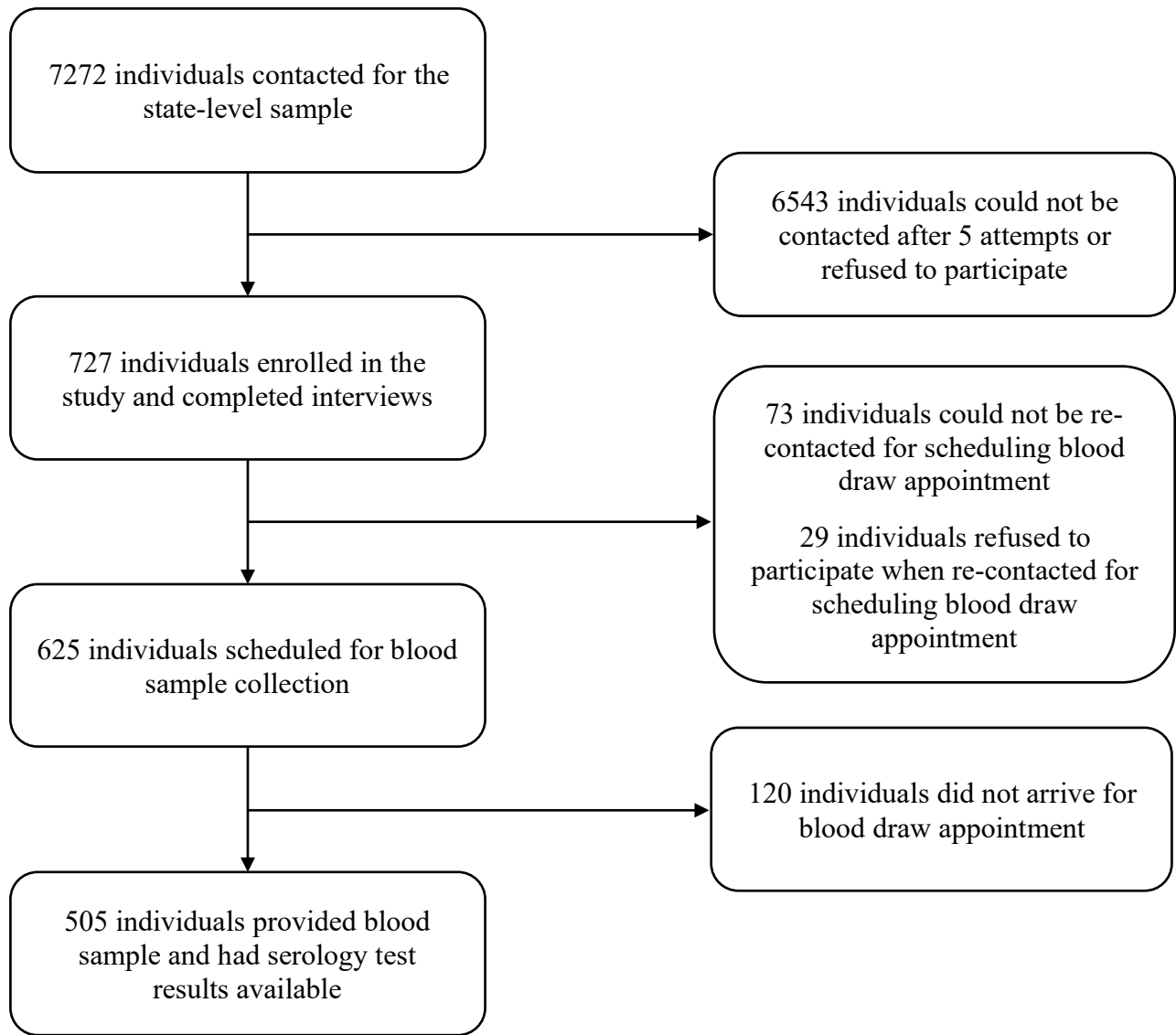
effect for every estimate is calculated and used in the construction of 90% confidence intervals using the below formula.

$$\text{MOE} = \pm 1.645 * \sqrt{P * (1 - P)/n} * \sqrt{\text{design effect}}$$

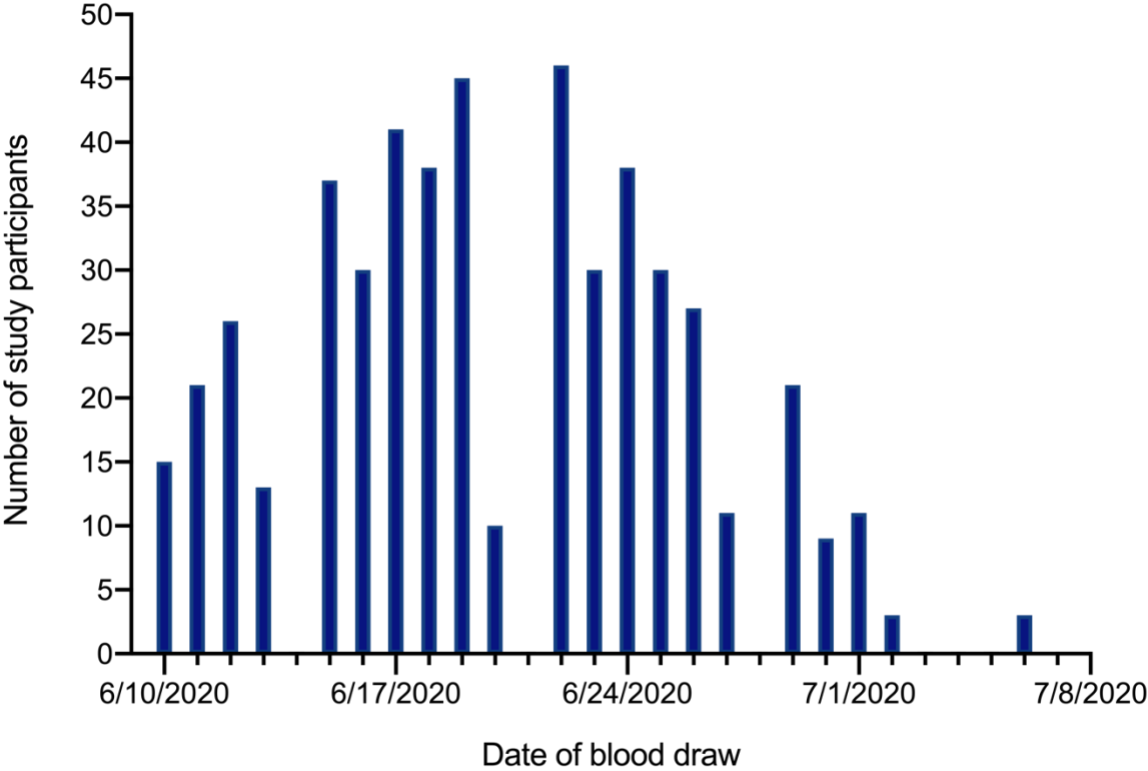
¹ The design effect was defined formally by Kish (1965), Section 8.2, p. 258) as “the ratio of the actual variance of a sample to the variance of a simple random sample of the same number of elements.” Based on Kish’s approximate formula:

$$\{\text{design effect} = n * \Sigma \cdot (w_t)^2 / ((\Sigma w_t)^2)\}$$

eFigure 1. Flow chart showing sample selection for the state-level estimate.



eFigure 2. Distribution of the timing of the blood draws between June 10 and July 6, 2020 for the state-level estimate.



eTable 1. Comparison of demographics (age-group, race/ethnicity, and geographic region) of those who completed the survey with those who completed the survey and the serology test.

	Completed survey but not blood test (N=222)		Completed survey and blood test (N=505)		P-value†
	N	Unweighted %	N	Unweighted %	
Region					
Total	222	-	505	-	-
Fairfield	46	20.7%	113	22.4%	0.30
Hartford	59	26.6%	137	27.1%	0.44
Litchfield	13	5.9%	39	7.7%	0.19
Middlesex	17	7.7%	29	5.7%	0.15
New Haven	48	21.6%	119	23.6%	0.28
New London	19	8.6%	37	7.3%	0.27
Tolland	10	4.5%	19	3.8%	0.33
Windham	6	2.7%	11	2.2%	0.34
Information missing	4	1.8%	1	0.2%	0.01
Race/Ethnicity					
Total*	241	-	510	-	
Hispanic	50	22.5%	38	7.5%	<0.001
Non-Hispanic White	144	64.9%	426	84.343.5%	<0.001
Non-Hispanic Black	39	17.6%	30	5.9%	<0.001
Non-Hispanic Asian	5	2.3%	9	1.8%	0.32
Non-Hispanic Other	1	0.5%	4	0.8%	0.32
Non-Hispanic Race DK/RF	2	0.9%	3	0.6%	0.32
Age					
Total	222	-	505	-	
18-29 years	39	17.6%	36	7.1%	<0.001
30-44 years	60	27.0%	76	15.0%	<0.001
45-54 years	39	17.6%	93	18.4%	0.40
55-64 years	16	7.2%	128	25.3%	<0.001
≥65 years	67	30.2%	170	33.7%	0.18
Refused	1	0.5%	2	0.4%	0.42
*Can have multi-response for race/ethnicity so this sum may be higher than the total N.					
†P-value at 95% confidence level					

eTable 2. Comparison of prevalence of symptomatic illness, risk factors for possible exposure, and adherence to social-distancing behaviors since March 1, 2020, among those who completed only the survey with those who completed the survey and the serology test.

Characteristics	Completed survey but not blood draw, N=222		Completed survey and blood draw, N=505		P-value†
	Unweighted N*	Unweighted %	Unweighted N*	Unweighted %	
Symptoms					
Fever	16	7.2%	41	8.1%	0.34
Cough	29	13.1%	73	14.5%	0.31
Sore throat	22	9.9%	48	9.5%	0.43
New loss of taste or smell	7	3.2%	19	3.8%	0.34
Diarrhea	23	10.4%	62	12.3%	0.23
Risk Factors/Behaviors					
Received coronavirus test	41	18.5%	78	15.4%	0.15
Tested positive for coronavirus	8	3.6%	9	1.8%	0.07
Anyone in household (other than respondent) had symptoms of coronavirus	14	6.3%	42	8.3%	0.18
Avoided going to public places, such as stores or restaurants	170	76.6%	376	74.5%	0.27
Avoided small gatherings of people, with family or friends	171	77.0%	377	74.7%	0.25
Worked from home (among all respondents, regardless of employment status)	57	25.7%	198	39.2%	<0.001
Worn a mask on your face when outside your home	211	95.0%	496	98.2%	0.01
Traveled by airplane	19	8.6%	33	6.5%	0.16
Traveled using public transportation, such as bus or train	15	6.8%	17	3.4%	0.02
*N for 'yes' response. †P-value at 95% confidence level					

eTable 3. Unweighted and weighted seroprevalence of SARS-CoV-2-specific IgG antibodies among adults in Connecticut, by sociodemographic and clinical subgroups.

Characteristics	Sample Size, N	Unweighted Seroprevalence, N (%)	Weighted Seroprevalence %, (MOE at 90% CI)	Weighted Seroprevalence %, (MOE at 95% CI)
Overall	505	18 (3.6%)	3.1% ($\pm 1.7\%$)	3.1% ($\pm 2.0\%$)
Age group, years				
18-29	36	1 (2.8%)	5.6% ($\pm 6.9\%$)	5.6% ($\pm 8.2\%$)
30-44	76	3 (3.9%)	2.4% ($\pm 3.5\%$)	2.4% ($\pm 4.1\%$)
45-54	93	7 (7.5%)	5.5% ($\pm 5.0\%$)	5.5% ($\pm 6.0\%$)
55-64	128	5 (3.9%)	2.0% ($\pm 2.7\%$)	2.0% ($\pm 3.2\%$)
≥ 65	170	2 (1.2%)	0.9% ($\pm 1.6\%$)	0.9% ($\pm 1.9\%$)
Sex				
Men	212	7 (3.3%)	2.1% ($\pm 2.1\%$)	2.1% ($\pm 2.5\%$)
Women	292	11 (3.8%)	4.1% ($\pm 2.6\%$)	4.1% ($\pm 3.1\%$)
Race/Ethnicity				
Hispanic	38	1 (2.6%)	6.2% ($\pm 7.9\%$)	6.2% ($\pm 9.4\%$)
Non-Hispanic White	426	13 (3.1%)	2.3% ($\pm 1.6\%$)	2.3% ($\pm 1.9\%$)
Non-Hispanic Black	30	3 (10.0%)	4.3% ($\pm 7.5\%$)	4.3% ($\pm 9.0\%$)
Non-Hispanic Asian	9	*	*	*
Non-Hispanic Other	4	*	*	*
Education level				
Less than high school	5	*	*	*
High school or GED	68	1 (1.5%)	2.4% ($\pm 3.3\%$)	2.4% ($\pm 4.0\%$)
Some college	117	4 (3.4%)	3.2% ($\pm 3.2\%$)	3.2% ($\pm 3.8\%$)
Bachelor's degree or more	314	13 (4.1%)	4.0% ($\pm 2.2\%$)	4.0% ($\pm 2.6\%$)
Income level				
Less than \$24,000	34	1 (2.9%)	3.6% ($\pm 6.3\%$)	3.6% ($\pm 7.5\%$)
\$24,000 to \$59,999	92	2 (2.2%)	3.2% ($\pm 3.9\%$)	3.2% ($\pm 4.7\%$)
\$60,000 to \$119,999	159	7 (4.4%)	3.3% ($\pm 3.1\%$)	3.3% ($\pm 3.7\%$)
\$120,000 or more	176	8 (4.5%)	3.5% ($\pm 2.9\%$)	3.5% ($\pm 3.5\%$)

Don't know/Refused	44	0 (0.0%)	0.0%	0.0%
Health insurance				
Yes	496	18 (3.6%)	3.3% ($\pm 1.8\%$)	3.3% ($\pm 2.1\%$)
No	8	*	*	*
Unknown	1	*	*	*
Employment status				
Employed full-time	231	10 (4.3%)	2.9% ($\pm 2.3\%$)	2.9% ($\pm 2.8\%$)
Employed part-time	48	5 (10.4%)	7.0% ($\pm 8.2\%$)	7.0% ($\pm 9.7\%$)
Unemployed	41	1 (2.4%)	3.6% ($\pm 6.1\%$)	3.6% ($\pm 7.3\%$)
Retired	139	1 (0.7%)	0.0%	0.0%
Homemaker	15	*	*	*
Student	6	*	*	*
Disabled	0	*	*	*
Unknown	25	*	*	*
Essential job (exempt from stay-at-home orders)				
Yes	115	5 (4.3%)	2.3% ($\pm 2.9\%$)	2.3% ($\pm 3.5\%$)
No	156	10 (6.4%)	4.1% ($\pm 3.4\%$)	4.1% ($\pm 4.1\%$)
Don't know/refused	33	1 (3.0%)	0.0%	4.0 ($\pm 12.2\%$)
Not Employed	201	2 (1.0%)		1.1% ($\pm 2.0\%$)
Region/County				
Fairfield	113	7 (6.2%)	6.6% ($\pm 5.2\%$)	6.6% ($\pm 6.1\%$)
Hartford	137	4 (2.9%)	2.8% (± 3.1)	2.8% ($\pm 3.7\%$)
Litchfield	39	2 (5.1%)	2.7% ($\pm 6.0\%$)	2.7% ($\pm 7.2\%$)
Middlesex	29	*	*	*
New Haven	119	3 (2.5%)	1.6% ($\pm 2.4\%$)	1.6% ($\pm 2.9\%$)
New London	37	1 (2.7%)	2.9% ($\pm 6.6\%$)	2.9% ($\pm 7.8\%$)
Tolland	19	*	*	*
Windham	11	*	*	*
Type of home				
Mobile home	2	*	*	*

Single family house/townhouse	400	16 (4.0%)	3.1% ($\pm 1.9\%$)	3.1% ($\pm 2.2\%$)
Apartment or condo	97	1 (1.0%)	2.9% ($\pm 3.6\%$)	2.9% ($\pm 4.3\%$)
Group facility	1	*	*	*
Unknown	5	*	*	*
Self-reported health status				
Excellent	160	3 (1.9%)	1.3% ($\pm 2.0\%$)	1.3% ($\pm 2.4\%$)
Very good	204	10 (4.9%)	3.5% ($\pm 2.8\%$)	3.5% ($\pm 3.4\%$)
Good	110	3 (2.7%)	1.6% ($\pm 2.4\%$)	1.6% ($\pm 2.9\%$)
Fair	26	*	*	*
Poor	4	*	*	*
Unknown	1	*	*	*
Chronic conditions				
Diabetes	52	1 (1.9%)	1.6% ($\pm 3.8\%$)	1.6% ($\pm 4.5\%$)
Asthma, COPD or another lung disease	52	1 (1.9%)	2.8% ($\pm 4.9\%$)	2.8% ($\pm 5.9\%$)
Heart disease	35	0 (0.0%)	0.0%	0.0%
Cancer	64	2 (3.1%)	7.8% ($\pm 7.8\%$)	7.8% ($\pm 9.3\%$)
High blood pressure	149	4 (2.7%)	2.1% ($\pm 2.7\%$)	2.1% ($\pm 3.2\%$)
Immune compromised	39	1 (2.6%)	9.1% ($\pm 10.8\%$)	9.1% ($\pm 12.8\%$)
Individual symptoms				
Fever	41	12 (29.3%)	22.6% ($\pm 14.4\%$)	22.6% ($\pm 17.2\%$)
Cough	73	10 (13.7%)	10.8% ($\pm 7.8\%$)	10.8% ($\pm 9.3\%$)
Sore throat	48	5 (10.4%)	10.5% ($\pm 9.7\%$)	10.5% ($\pm 11.5\%$)
New loss of taste or smell	19	*	*	*
Diarrhea	62	4 (6.5%)	7.1% ($\pm 6.7\%$)	7.1% ($\pm 8.0\%$)
Symptoms aggregate				
Asymptomatic	366	4 (1.1%)	0.6% ($\pm 0.9\%$)	0.6% ($\pm 1.1\%$)
1 or more symptoms	139	14 (10.1%)	8.0% ($\pm 4.9\%$)	8.0% ($\pm 5.8\%$)
2 or more symptoms	60	11 (18.3%)	13.0% ($\pm 9.5\%$)	13.0% ($\pm 11.3\%$)
Risk Factors/Behaviors				
Received coronavirus test	78	10 (12.8%)	13.0% ($\pm 8.6\%$)	13.0% ($\pm 10.3\%$)
Tested positive for coronavirus †	9	*	*	*

Anyone in household (other than respondent) had symptoms	42	8 (19.0%)	13.3% ($\pm 10.8\%$)	13.3% ($\pm 12.9\%$)
Anyone in household (other than respondent) tested positive for coronavirus	9	*	*	*
Avoided going to public places, such as stores or restaurants	376	13 (3.5%)	3.0% ($\pm 1.9\%$)	3.0% ($\pm 2.3\%$)
Avoided small gatherings of people, with family or friends	377	13 (3.4%)	3.2% ($\pm 2.0\%$)	3.2% ($\pm 2.4\%$)
Worked from home (among all respondents, regardless of employment status)	198	11 (5.6%)	4.0% ($\pm 2.9\%$)	4.0% ($\pm 3.5\%$)
Worn a mask on your face when outside your home	496	18 (3.6%)	3.3% ($\pm 1.8\%$)	3.3% ($\pm 2.1\%$)
Traveled by airplane	33	0 (0.0%)	0.0%	0.0%
Traveled using public transportation, such as bus or train	17	*	*	*
<p>* Sample size is <30 and too small to report † Though the sample size was too small to report seroprevalence estimates, all 9 of these individuals tested positive for SARS-Cov-2-specific IgG antibodies. Note 1: The rows highlighted in grey indicate estimates where we are confident at the 90% level that there is not a null result (estimate is not equal to 0). Though the results for the other rows are presented, the sample size is inadequate to be able to detect any significance. Abbreviations: COPD, Chronic Obstructive Pulmonary Disease; GED, General Educational Development test; MOE, Margin of Error</p>				