Covid-19 Seroprevalence - Do the Results Advance the Aims?

Authors: Gal Almogy, PhD

Author affiliation: Flurensics, Tel Aviv
Contact email: dr.galmogy@gmail.com

Abstract:

Seroprevalence studies suggest that the number of PCR-confirmed COVID-19 cases is significantly smaller than the true number of infections. I study logintidual seroprevalence data from 7 sites across the US, from early April 2020 to June 27. I show that not only COVID-19 seroprevalence does not seem to increase over time, there is no clear association between the number of cases reported during a period and the change in seroprevalence during the same time. I conclude that as they are, seroprevalence studies can only be used in the qualitative sense and distinguish between populations with no COVID-19 exposure, to those populations where the virus had already started spreading.

Introduction

Seroprevalence studies suggest that the number of PCR-confirmed COVID-19 cases in the US, especially at the early stages of the outbreak when testing capacity was more limited, likely represents only a fraction of the total number of infections (1-4).

Seroprevalence studies are useful in estimating the proportion of the population already exposed to the virus. However it is difficult to compare the results of different serology surveys due to differences in study designs, as well as potentially large heterogeneity in the cumulative COVID-19 incidence across different regions, and even within regions (5-6).

Data from the first round of a large seroprevalence study covering multiple sites across the US suggested higher than reported covid-19 incidence. Havers et al (7) found that 6.9% and 1.1% of the population in NY and WA produced antibodies against covid-19 by April 2020, suggesting only 10% of all COVID cases in those regions were identified in real-time.

Here I examine the covid-19 seroprevalence in the two following CDC serology surveys, conducted in May and in June (8), in an effort to better understand the relationship between the seroprevalence and increased covid-19 incidence. Methods
Data was retrieved from the CDC website (8), including the adjusted seroprevalence rates, estimated number of infections, and reported number of cases per site.

The estimated number of new infections per survey round is derived by subtracting the estimated number of infections in the previous round. I.e, the estimated number of *new* infections in round 3 is the estimated number in round 3 minus the estimate in round 2 as provided by the CDC:

C_SP (round 1) = seroprevalence (round 1)
C_SP (round N) = seroprevalence (round N) - seroprevalence (round N-1)

And similarly for the change in reported cases:
C_RC (round N) = reported cases (round N) - reported cases (round N - 1)

In some states (marked with * in tables 1 and 2) more counties were surveyed in rounds 2 and 3, complicating the comparison between rounds 1 and 2, and hence comparisons made regarding the change from round 1 to round 2 surveys should be treated with caution.

All statistical calculation, confidence intervals etc were directly adopted from CDC published results.

Results

To better understand the relationship between the reported seroprevalence rates and the number of covid-19 cases I compare the change in seroprevalence and the change in incidence for the 3 survey rounds.

Part 1: CDC method

The expected number of infections given the seroprevalence, and the reported incidence at the end of each survey period are presented in table 1 (data from CDC). The ratio of estimated to reported cases (in brackets in table 1) decreased over time in all sites, which is consistent with the increase in covid-19 testing capacity over the course of the outbreak. In the first round the ratio was highest in MO, where the number of estimated infections was 24 times the reported number of cases and minimal in CT, where seroprevalence data suggested 6-times more infections than reported. The highest ratio in the second round was again in MO, where the number of estimated infections was 13-times the number of reported cases. In the third round the ratio ranged between 2 (UT) and 7 (NY).

There are however two facts to note in context. First, with the exception of the increase in NY (first to second survey rounds), *the change in seroprevalence over time, in all other survey sites, was not statistically significant* (see CDC for details). Second, while it is not exactly known how many of those infected develop a humoral response and how long the response lasts, the current data suggests that in those who develop a response, antibodies can be detected for at least 2-3 months and likely longer (9-11). Therefore the difference in the estimated number of infections between two survey rounds represents the number of new infections during that period.

Part 2 - the change in seroprevalence vs the change in the number of cases

Assuming covid-19 seropositivity does not significantly decay over the course of two months (9), then changes in seroprevalence can be taken to represent the increase in the number of people that were exposed to the virus between survey rounds.

In table 2 I present the changes in the number of estimated infections based on serology results, C_SP, and the change in the number of reported cases, C_RC (defined in the methods) for the three survey rounds. For the change on the first round I assume seropositivity was zero before the outbreak began.

In the unrealistic scenario where both serology and PCR tests are 100% accurate and that every person is tested daily by PCR, then the change in seroprevalence between two time points should accurately represent the number of new infections, and should equal the number of confirmed cases *added* during the period.

Even if we ignore the cases where seroprevalence decreased (marked red in table 2), it is clear that the reduction in the ratio of estimated infections to confirmed cases seen in table 1 is limited to the period between rounds 1 and 2 - in fact, out of the 4 sites where seroprevalence increased from round 2 to round 3, in two the ratio increased, from <1 to 9 in CT, and from 10 to 16 in MN.

In conclusion, the round-to-round reduction in estimated infections to confirmed cases (table 1) does not indicate that health authorities are identifying a higher proportion of new cases in real time, but rather that the number of cases is substantially increasing, whereas seroprevalence levels remain unchanged.

Discussion

Results from the first round (April to early May) of an ongoing longitudinal serology survey were taken to indicate that the number of infections in multiple sites across the US was much greater

than reported cases. Similarly high ratios of seroprevalence to the number of confirmed cases were found in several other surveys in the US and elsewhere (1-8).

However, while the increase in seroprevalence observed in April-May, compared to pre-pandemic period, is reported to be significant, and may reflect widespread undetected transmission that occurred before the first survey period, seroprevalence did not increase in a statistically significant manner (95% confidence) in any of the test sites in the later rounds (except for NY). Moreover, a comparison of the increase in seropositivity and the number of new cases added between survey rounds (table 2) shows that the change in seroprevalence was not a reliable predictor of the number of new infections added during a period.

Even if not immediately helpful in determining the true number of infections in a population, the rigorous study design and the scale of the CDC survey make it an excellent reference point for interpreting results of other studies in context. Specifically, it seems there are three rough epidemic stages that can be inferred from seroprevalence studies:

Pre-pandemic: seroprevalence = 0
Community transmission: seroprevalence > 1%
Post-outbreak: seroprevalence > 20%

More refined predictions, e.g "true infection rate was X-times higher than reported" (1-8 and others), should be taken with a large amount of salt.

References

1. Sandri, Maria Teresa, et al. "IgG serology in health care and administrative staff populations from 7 hospital representative of different exposures to SARS-CoV-2 in Lombardy, Italy." medRxiv (2020).
2. Ioannidis, John. "The infection fatality rate of COVID-19 inferred from seroprevalence data." medRxiv (2020).
3. Arora, Rahul K., et al. "SeroTracker: a global SARS-CoV-2 seroprevalence dashboard." The Lancet Infectious Diseases (2020).
4. Bendavid, Eran, et al. "COVID-19 Antibody Seroprevalence in Santa Clara County, California." MedRxiv (2020).
5. Hallal, Pedro, et al. "Remarkable variability in SARS-CoV-2 antibodies across Brazilian regions: nationwide serological household survey in 27 states." medRxiv (2020).
6. Song, Suk-Kyoon, et al. "IgG Seroprevalence of COVID-19 among Individuals without a History of the Coronavirus Disease Infection in Daegu, Korea." Journal of Korean Medical Science 35.29 (2020).
7. Havers, Fiona P., et al. "Seroprevalence of antibodies to SARS-CoV-2 in 10 sites in the United States, March 23-May 12, 2020." JAMA Internal Medicine (2020).

8. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/about-serology-surveillance.html
9. Xiao, Ai Tang, Chun Gao, and Sheng Zhang. "Profile of specific antibodies to SARS-CoV-2: the first report." The Journal of infection (2020).
10. Blairon, Laurent, et al. "Large-scale, molecular and serological SARS-CoV-2 screening of healthcare workers in a 4-site public hospital in Belgium after COVID-19 outbreak." Journal of Infection (2020).
11. Sandri, Maria Teresa, et al. "IgG serology in health care and administrative staff populations from 7 hospital representative of different exposures to SARS-CoV-2 in Lombardy, Italy." medRxiv (2020).

Table 1

State	Round 1 expected (ratio)	Round 2 expected (ratio)	Round 3 expected (ratio)	Round 1 reported	Round 2 reported	Round 3 reported
NY	642000 (12)	2832000 (10)	2376000 (7)	53803	281670	329418
WA	48000 (11)	122000 (9)	100000 (6)	4308	13098	17185
MO	162000 (24)	171000 (13)	51000 (3)	6794	12956	17588
PA	157000 (7)	245000 (4)	255000 (4)	22987	56318	64095
UT	47000 (11)	25000 (2.5)	33000 (2)	4493	10310	17180
CT	176000 (6)	185000 (4.5)	223000 (5)	29287	41234	45347
MN	91000 (10)	84000 (4)	167000 (6)	8880	21345	26528

Legend: the expected number of infections for each survey round, as reported by the CDC. The (rounded) ratio of the number of expected number of infections to the number of cases reported by the end of the survey period is in brackets. Red indicates seroprevalence was lower than on the previous round. The number of cases reported by the end of each survey period is presented in the last three columns.

Table 2

State	C_SP (round 1) (ratio)	C_SP (round 2) (ratio)	C_SP (round 3) (ratio)	C_RC (round 1)	C_RC (round 2)	C_RC (round 3)
NY	642000 (12)	2190000 (10)	-456000	53803	227867	47748
WA	48000 (11)	74000 (8)	-22000	4308	8790	4087
MO	162000 (24)	9000 (1.5)	-120000	6794	6162	4632
PA	157000 (7)	88000 (3)	10000 (1.5)	22987	33331	7777
UT	47000 (11)	-22000	8000 (1)	4493	5817	6870
CT	176000 (6)	9000 (<1)	38000 (9)	29287	11947	4113
MN	91000 (10)	-7000	83000 (16)	8880	12465	5183

Legend: the expected change in the number of cases (C_SP) between the survey rounds as defined in the methods. Negative values (red) are the result of seroprevalence decreasing between rounds; e.g C_SP (round 3) in NY or C_SP (round 2) in UT. The (rounded) ratio of C_SP to the change in the number of cases (C_RC) over the 3 rounds is in brackets. The ratio was not calculated for negative C_SP values.