

ⓘ Anyone can publish on Medium per our [Policies](#), but we don't fact-check every story. For more info about the coronavirus, see [cdc.gov](https://www.cdc.gov).

Peer Review of “COVID-19 Antibody Seroprevalence in Santa Clara County, California”



Balaji S. Srinivasan [Follow](#)

Apr 18 · 14 min read ★

The high reported positive rate in this serosurvey may be explained by the false positive rate of the test and/or by sample recruitment issues.

A new Stanford preprint was released earlier today (PDF, SI). The authors claim that the true population prevalence of COVID-19 in Santa Clara County is 50–85X higher than the number of confirmed cases. They base their reasoning on a serosurvey of 3330 participants.

If true, this would actually be good news for society. It would mean that the virus had already widely spread, and thus had a lower fatality rate than previously expected, so the disease wasn't as severe as we thought. Indeed, the authors claim this explicitly, by putting rough caps on the number of deaths and the infection fatality rate (as distinct from the case fatality rate).

We can use our prevalence estimates to approximate the infection fatality rate from COVID-19 in Santa Clara County. As of April 10, 2020, 50 people have died of COVID-19 in the County, with an average increase of 6% daily in the number of deaths. If our estimates of 48,000-81,000 infections represent the cumulative total on April 1, and we project deaths to April 22 (a 3 week lag from time of infection to death²²), we estimate about 100 deaths in the county. A hundred deaths out of 48,000-81,000 infections corresponds to an infection fatality rate of 0.12-0.2%. If antibodies take longer than 3 days to appear, if the average duration from case identification to death is less than 3 weeks, or if the epidemic wave has peaked and growth in deaths is less than 6% daily, then the infection fatality rate would be lower. These

From Page 7 of the PDF

However, I am skeptical of this result for several reasons.

Before I begin, I want to say that (1) I'm very glad that the authors are doing serosurveys, (2) I offer this writeup in the spirit of a peer review from a fellow scientist and citizen, (3) I believe I am qualified to render a peer review in the field of molecular diagnostics and statistics, (4) I too would like to get people back to work without causing mass sickness, (5) I hope

the feedback from this round of review (even if I myself am mistaken!) allows us all to develop improved serosurveys.

OK, so why I am I skeptical of this paper?

1. The False Positive Rate of the Test is High

The most important raw result by the authors is the claim that 50 out of their 3330 participants tested positive on an antibody test, showing that they were already infected.

However, as we will see, because they reported 2 false positives out of 401 tested samples, there is a really wide range on what the actual false positive rate could be, and it could be significantly higher than 1.2%. So false positives could account for many if not all of the 50 reported positives in their study.



Above is page 16 of their PDF. A critical question is: **how many of these 50 positives might be false positives?** To answer this, we need to look at the calibration of the test against gold standard positive and negative results. That is, if you put in a known negative sample into the test, something known not to have COVID-19 antibodies, do you get out a negative result? Or do you get out a positive result? The former is called a “true negative” and the latter is a “false positive”, as shown below.





From their Results section and Supplementary Information, we get some crucial information on the false positive rate. Here are the key bits:



From their Results section, page 5



From their Supplementary Info. In this third scenario, they combine the 30/30 figure from their own validation and 369/371 figure from the manufacturer into a single 399/401 overall specificity.

They show that in their internal calibration, 30 out of 30 known negatives tested negative. In the manufacturer's calibration, however, 369 out of 371 known negatives tested negative. If we follow the authors and just combine these, **they had 2 false positives out of 401 total.** As Jeffrey Spence observed, this is actually a high enough rate to potentially mess up the whole study.

Jeffrey likely means $(371 + 30)$ known negatives here rather than $(371 + 35)$, and our confidence interval somewhat differs from his. But the conclusions are directionally the same.

To show this, we are going to go through three progressively more sophisticated ways to estimate the false positive rate:

- A simple point estimate: “the false positive rate is $2/401$ or $\sim 0.5\%$ ”
- A naive confidence interval based on a normality assumption: “the false positive rate is between 0% and 1.2% with 95% confidence”
- A less naive confidence interval based on the reality of non-normality: “the false positive rate is between 0% and a number greater than 1.2% with 95% confidence”

Let's start with the simple way.

A naive false positive rate is a point estimate: $2/401$ or 0.5%

Let's assume for a second that among the 3330 participants there were *actually* 0 that were positive. So with a perfect test we'd get 3330 true

negative results out and 0 true positive results out.

However, if we have a test with a false positive rate of $2/401$, then if we ran it on this population we'd get on average $2/401 * 3330 \approx 17$ *false* positives with the remainder being 3313 true negatives. That number of 17 false positives is roughly ***one third*** of the 50 total reported positives, and we'd get that in a sample with *zero* actual positives. That hints that a solid chunk of the reported positives could be false positives.

Might it be more than that?

Actually, yes, because when you have a low enough numerator like 2 out of 401, you can't just assume the false positive rate is exactly $2/401 = 0.5\%$. Because if you re-calibrated the assay on a different set of 401 known negative samples, you might get 0/401 false positives (for a false positive rate of 0%), or even 6/401 (for a false positive rate of 1.5%).

So, can we get a confidence interval for what the actual false positive rate could be, given that we saw 2 false positives out of 401 total tests?

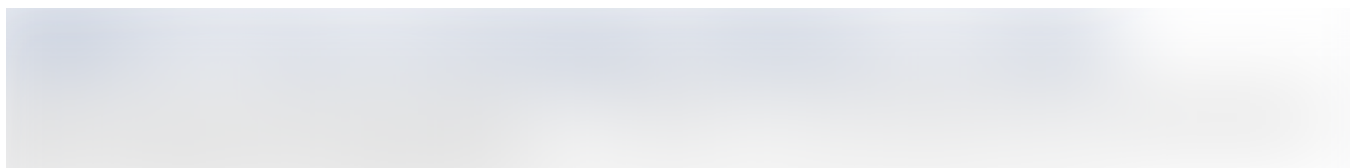
A less naive false positive rate is a confidence interval: [0%, 1.2%]

Beyond this point there will be some stats. You don't need to understand it all. The main point is that the false positive rate could be considerably higher than 0.5%. It might be more like 1.2% or even higher.

How did we come to that conclusion?

Basically, you can think of each of the $n=401$ calibration tests as a coin flip, with an unknown probability p of coming up heads (a false positive) that we're trying to put some bounds on. The exact model for the observed number of false positives given (n, p) is the binomial distribution. We divide that by the total number of samples (n) to get something called the *sampling distribution of the sample proportion*. It lets you put error bars on estimates of proportions that come from real data, like the proportion of false positives.

Under some circumstances (actually invalid in this case, but more on that in a second) you can approximate this distribution with a normal distribution. Here's a reference with a worked example:





So to recap: this formula is a way to get a confidence interval on a “sample proportion” like the $2/401$ number we have, the proportion of false positives. If we assume for the moment that the sample proportion is distributed normally (more on this in a second) you can then add two

standard deviations up and down to get a 95% confidence interval for the false positive rate. Plugging in numbers, we get:



Note in the last line we restrict the lower bound of the confidence interval for the false positive rate to be 0 because a proportion can't be negative (that is, we can't get less than 0 false positives out of 401 trials).

So under the normality assumption, the 95% confidence interval for the false positive rate is $[0, 0.012]$ or $[0\%, 1.2\%]$. The upper end of that range is high.

Let's return to that idealized population of 3330 participants where *zero* were actually positive and all 3330 were actually negative. If the false positive rate was actually 1.2% then you'd get $1.2\% * 3330 = \sim 40$ false positives.

If we compare to the 50 reported positives in the study, **at the upper end that kind of false positive rate would mean $40/50 = 80\%$ of the positives in the study could be false positives.**

You could make this calculation exact by starting with a population of 3330 participants, of which there would be a small number of actual positives (call it k) and a much larger number of actual negatives ($3330-k$). You would then do out the full 2x2 contingency table to figure out how many false positives, false negatives, true positives, and true negatives you get under various assumptions of false positive and false negative rates.

But when you do that out, you would see that so long as k is small (and k is likely less than 50 as that's the total reported positives in the study), you would find that if the false positive rate is anything like 1.2% then the false positives would dominate the share of reported positives.

An even less naive false positive rate is: [0%, >1.2%]

Now, remember that we assumed normality? That's not really applicable in this circumstance. When p is around 50%, a normal distribution is a reasonable approximation for the sample proportion.

But when n is large ($n = 401$) and p is small ($p \sim 2/401$), a Poisson distribution is a better approximation to the binomial than a normal distribution. There are different rules of thumb for when to use this, but Devore's says a binomial can be approximated by a Poisson distribution if $n > 50$ and $np < 5$, which are both satisfied here ($n=401$, $np=2$).

To get something similar to the sampling distribution for the sample proportion, you might start by looking at the distribution of the number of false positives if it was modeled by a Poisson distribution with parameter $\lambda = np = 2$, and then dividing by $n = 401$ to get a false positive rate. Here's a graph.

Now, just like when we were looking at the sampling distribution of the sample proportion, we don't actually know what the *actual* false positive rate (p) is. All we have to go on are two numbers: 2 false positives in 401 samples. But if the actual false positive rate is *around* $2/401$, then that graph of the Poisson distribution with $np=2$ is *close* to what the distribution of false positives will be given $n=401$ tests.

Here's the issue: this is bad news for the false positive rate. A Poisson distribution with low mean ($np = 2$) has a long right tail, so when you divide by $n=401$ to get a false positive rate the actual upper bound on the confidence interval for the false positive rate will be significantly higher than 1.2%.

This is just a sketch of an argument. You can throw a lot of math at this problem if you want to make it complicated. But at this point we're breaking a butterfly on a wheel. The only raw inputs after all were "2" and "401". And the main conclusion is that **we don't have high confidence that the false positive rate is low enough.**

TLDR: because the authors reported 2 false positives out of 401 tested samples, there is a really wide confidence interval on what our actual false

positive rate could be, and it could be significantly higher than 1.2%. This *could* account for many if not all of the 50 reported positives in their study.

This is one possible failure mode.

2. Were Participants Enriched for COVID-19 Cases?

Recall again that the authors reported 50 positives out of 3330 total participants. They recruited these participants via Facebook ads. They used these positive results to conclude that the population prevalence in Santa Clara County was much higher than expected, indeed 50–85X higher than you'd expect based on confirmed cases.



From page 7

Put another way, rather than 956 confirmed cases out of 1.928 million in Santa Clara County $\sim = 1/2000 = 0.05\%$ rate of infection, they think it's 2.49%-4.16%. That's how they get 50–85X.



From Table 2

But there's a problem: what if their group of participants was *enriched* for positives relative to the general population? What if their participants had a much higher rate of COVID-19 than normal? As a *reductio ad absurdum*, if you went to a hospital and tested all recovering patients for COVID-19 antibodies, you'd probably get a very high percentage of positives. You wouldn't be able to generalize from that to say that much of the general population had already gotten the illness. To their credit, the authors acknowledge this might be an issue:



They do try to correct for a possibly unrepresentative sample later in the paper with various renormalizations and so on. The general thrust of their argument is that even if their sample wasn't representative, it wouldn't produce a 50–85X increase in positives relative to baseline. However, there are two mechanisms that could do exactly that.

2A. Exposed people may have signed up for the study to get tested

The first mechanism that could significantly enrich the number of COVID-19 cases in the study is if symptomatic or exposed people used the study to get a test they could get no other way.





From the San Jose Mercury News on April 2, 2020 [around the time samples were collected]

After all, in the Bay Area in early April, it was really hard to get a test for people with mild symptoms or exposure. So people who thought they were exposed or symptomatic may have signed up for the study to get access to a free COVID-19 test they could get no other way. We only need ~50 out of 3330 to exhibit this behavior. And there are at least some who appear to have done just that.

2B. Exposed people may have recruited other exposed people for the study

The second mechanism that could significantly enrich the number of COVID-19 cases in the study is if symptomatic or exposed people recruited other symptomatic or exposed people to get tested.

Recall that recruitment of study participants was done through Facebook. People who thought they had symptoms or exposure could be sharing links to the study in private groups, WhatsApp chats, email threads, and the like. If one of those groups was for people who had COVID-19 symptoms or exposure, then it's game over: you could get a “super-recruiter” event where one person recruits N other enriched people into the study. That could significantly boost the number of positives beyond what you'd see in a random sample of Santa Clara.

2C. How plausible are these enrichment mechanisms?

Could symptomatic/exposed people have signed up to get tested? Or could such people have sent the link around to private groups enriched for other people with COVID-19 symptoms or exposure?

James Cham, a venture capitalist in the Bay Area, told me that "I am pretty sure these ads were forwarded widely on WhatsApp in the Chinese community around Silicon Valley. We saw it more than a couple of times."

Farhad Manjoo of the New York Times said:

And one of the study authors said:





<https://twitter.com/i/status/1251193282006310915>

So, this appears a bit contradictory, but the overall impression is that (a) individual links to the study from Facebook ads were unique but (b) anyone

could go and sign up for the study once they knew about it. This would mean highly non-random recruitment that could enrich for the number of positives.

3. The Study Would Imply Faster Spread than Past Pandemics

At the risk of paraphrasing incorrectly, the gist of the paper's argument is that "many people have already got it, they didn't have severe symptoms, so it's not as serious as you think, and the total number of deaths and fatalities will be low.". Here's a section of the paper that makes this argument explicitly.



Now, we don't know where the number of deaths will end up. It is possible that social distancing is bending the curve, though as of April 17 the COVIDtracking website doesn't yet seem to show a deceleration in daily new deaths for the state of California. I haven't seen more granular statistics for Santa Clara County. The premise of social distancing is that mass infection will cause mass sickness and death.

However, the paper's prediction that the infection fatality rate is only 0.12%–0.2% is a statement that even *with* mass infection we will see far less deaths than expected.

Here's the issue: we are *already seeing* the virus boost all-cause mortality in NYC and multiple countries, as the graphs below show:





From April 10: <https://www.nytimes.com/interactive/2020/04/10/upshot/coronavirus-deaths-new-york-city.html>



The Economist now includes graphs of excess weekly mortality.

In order to generate these thousands of excess deaths in just a few weeks with the very low infection fatality rate of 0.12–0.2% claimed in the paper, the virus would have to be wildly contagious. It would mean all the deaths are coming in the last few weeks as the virus goes vertical, churns out millions of infections per week to get thousands of deaths, and then suddenly disappears as it runs out of bodies.

I guess this isn't strictly impossible, but I was skeptical when this theory was first mooted because it's different from the way past influenza-like pandemics have played out, including H1N1 in 2009. That took 12+ months to infect 11–21% of the world over multiple waves.

4. Conclusion

To summarize, there are three broad reasons why I am skeptical of this study's claims.

1. First, the false positive rate may be high enough to generate many of the reported 50 positives out of 3330 samples. Or put another way, we don't have high confidence in a very low false positive rate, as the 95% confidence interval for the false positive rate is roughly $[0\%, > 1.2\%]$ and the reported positive rate is $\sim 1.5\%$.

2. Second, the study may have enriched for COVID-19 cases by (a) serving as a test-of-last-resort for symptomatic or exposed people who couldn't get tests elsewhere in the Bay Area and/or (b) allowing said people to recruit other COVID-19 cases to the study in private groups. These mechanisms could also account for a significant chunk of the 50 positives in 3330 samples.
3. Third, in order to produce the visible excess mortality numbers that COVID-19 is already piling up in Europe and NYC, the study would imply that COVID-19 is spreading significantly faster than past pandemics like H1N1, many of which had multiple waves and took more than a year to run their course.

These points may be mistaken. If so, I welcome corrections. And it would be wonderful news as it would imply we were much closer to herd immunity at a lower cost than people thought.

Alternatively, if these points are correct, we should try to do a second round of serosurveys that (a) aggressively reduces the false positive rate with many controls and possibly multiple independent tests and (b) that uses some form of unbiased recruitment for the serosurvey, potentially similar to jury duty.

While I disagree with the conclusion of the paper, I want to thank the authors for their hard work and hope that these comments prove useful in future serosurveys.

APPENDIX

Morning of April 18, 2020: made some updates to the document, including a link to this comment and edits/typo fixes for clarity. No substantive changes to any of the three points, namely (a) wide confidence interval for false positives, (b) possible significant enrichment for COVID-19 cases, and (c) speed vs past pandemics.

Below is just for the social media preview. You can ignore it!





[Coronavirus](#)

[Covid 19](#)

[Research](#)

[Statistics](#)

[Stanford](#)

Discover Medium

Welcome to a place where words matter. On Medium, smart voices and original ideas take center stage - with no ads in sight. Watch

Make Medium yours

Follow all the topics you care about, and we'll deliver the best stories for you to your homepage and inbox. Explore

Become a member

Get unlimited access to the best stories on Medium — and support writers while you're at it. Just \$5/month. Upgrade

[About](#)

[Help](#)

[Legal](#)