2.9.A Santa Clara County COVID-19 Seroprevalence study (FDR removed)

The Santa Clara County COVID-19 Seroprevalence study (https://doi.org/10.1101/2020.04.14.20062463) was highly controversial due to possible bias in the sampling and miscalculation of confidence intervals. We will not be discussing those issues here.

In a single day 3330 county residents were tested for antibodies to SARS-Cov-2 using a point-of-care test kit. Of the 3330, 50 tested positive.

We will call the proportion with a positive test, 50/3330 = 1.5% P(T+).

If an antibody test is imperfect, the proportion of a population with a positive test P(T+) does not accurately represent the proportion of the population that was previously infected P(D+).

The POC test kit is a lateral flow assay distributed by Premier Biotech (Minneapolis, MN) and manufactured by Hangzhou Biotest Biotech (Hangzhou, China). It tests for IgG and IgM antibodies to SARS-Cov-2. The authors reported that, in a ***previous validation study***[1], out of 157 specimens from individuals known to have had COVID-19 (we will refer to them as D+), 130 had a positive test. Out of 3324* specimens from (D-) individuals known not to have been infected, 16 had a positive test.

* Coincidence! The similarity of 3324 D- patients in the previous validation study to the 3330 in the Santa Clara County sample is purely a coincidence. These were two separate studies, a validation study to determine the accuracy of the test, and a sero-prevalence study to determine the prevalence of prior infection.

a) What are the sensitivity and specificity of the test? *(2 points: 1 pt for each Sens & Spec)*

Because the test is imperfect, the proportion with a positive test P(T+) is not necessarily the same as the proportion of the sample that has had COVID-19, which we will call the true prevalence of prior infection or P(D+). We want to adjust P(T+) to get P(D+).

b) First, ignore the study's actual P(T+) of 1.5% and assume that nobody had been previously infected, i.e., P(D+ = 0), how many positive tests would you expect to see out of 3330? *(2 points: 1 pt for each part)*

c) You knew the proportion of positives P(T+) you would see if nobody was D+ (P(D+) = 0). What proportion of positive tests would you see if 20% were D+, i.e. P(D+) = 0.2? Again, sensitivity and specificity as per part (a). *(1 point)*

d) If you did (c), you realize that you can go from P(D+) to P(T+). In the actual study P(T+) was 50/3330 = 1.5%. What's your estimate of P(D+)? *(Extra Credit 2 points)*

Now that you have done all this work, see the calculator at

https://sample-size.net/prevalence-calculator/

This calculator also gives you confidence intervals, which is not trivial.

[1] It was actually several different previous validation studies. Their results were combined together as if there were only one validation study. For this problem, you may assume this is valid.