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)*

Sensitivity: 130/157 = 82.8%
Specificity: (3324 – 16)/3324 = 99.5%

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)*

P(T+) = 0.5%. You would expect to see 0.5% positive, 0.5%*3330 ≈ 16 – 17 (Note: the wording of this question was changed. Full credit for 0.005 × 3330 = 16.65 or 17 or 16.)

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)*

You can do this with the actual counts or with the probabilities.
With counts:
20% * 3330 = 666.
True positives: 666 * 82.8% = 551
False positives: 2664*0.5% = 13
Total positives: 551 + 13 = 564
P(T+) = 564/3330 = 16.9%

[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.

With probabilities (preferred):
P(T+) = P(D+)*P(T+|D+) + (1-P(D+))*P(T+|D-)
P(T+) = 0.2 * 0.828 + 0.8 * 0.005 = 0.169

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)*

P(T+) = P(D+)*P(T+|D+) + (1-P(D+))*P(T+|D-)

P(T+) = Q
P(D+) = P
P(T+|D+) = Se
P(T+|D-) = 1 – Sp = F

Q = P*Se + (1-P)*F
= P*Se – P*F + F
= P[Se – F] + F
Q – F = P[Se – F]
[Q - F]/[Se – F)] = P

Substituting 1- Sp = F

P = [Q – (1 – Sp)]/ [Se – (1 – Sp)]
= [Q + Sp -1]/[Se + Sp – 1]

P(D+) = [P(T+) – (1 – Sp)]/ [Se – (1 – Sp)]
= [P(T+) + Sp -1]/[Se + Sp – 1]
This is called the Rogan-Gladen formula.

The answer here is (1.5% - 0.5%)/(82.8% - 0.5%) = 1.2%

The true prevalence P(D+) of 1.2% is slightly lower than the apparent prevalence P(T+) of 1.5%, because the false positives outnumber the false negatives. In this case, there are about 40 D+ out of 3330. 33 will have a (true) positive test, which means 17 false positives, and 7 false negatives.

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.