On determining the post-test probability of covid-19

Summary:

1. It is crucial that we not just test for covid-19, but also correctly interpret the test result, especially with rapid antigen tests
2. For example, even with a negative rapid antigen test result, one may still have covid
3. One needs to estimate the probability of covid given the test result, $p(dz|test, x, z)$
4. Often, this is done with Bayes rule $p(dz|test, x) = p(test|dz, x)p(dz|x)/p(test|x)$
5. Commonly, one tries to fill in the pieces of Bayes rule with estimates of $p(test|dz, x)$ from a study that measures $test, dz$, and x. However, to do so, one's pretest probability $p(test|x)$ must depend on no more and no less than x from the study (and, of course, the estimate $p(test|x)$ must be valid, which depends on the study design).
6. Often, we estimate $p(dz|test, x)$, but we really think we are estimating (and really want to estimate) $p(dz|test, x, z)$, where z contains information not in x, such as time-dependent and geography-dependent variables like prevalence, time of year, symptoms, known exposures, etc. . .
7. Sometimes this is important, sometimes it's not. Eg, based on (Prince-Guerra et al. 2021), a study that investigated covid rapid tests, the presence of symptoms can change the post-test probability estimate significantly
8. It is possible sometimes that for specific variables in z, for which we understand the correlation with disease, we can make statements about bounds on $p(dz|test, x, z)$ using $p(dz|test, z)$, if we are ok with a marginal estimate
9. Overall, we should maintain location-specific, continually updating databases of $dz, test, x$, and z, and we should also perform more studies to try to see what should be included in z

Screening:

Some months ago, I wanted to, given what I had available, try to determine the post-test probability of covid-19 in a person who had a negative Binax NOW rapid antigen test after a sustained, unmasked exposure. Originally, I hoped we could apply Bayes' rule to some estimates from the literature. Note, however, that the estimates in the literature do not take into account the extra variable that we call "exposure." However, assuming that we have all necessary variables besides exposure (a strong assumption), we still have a sort of bound at least, which might still be useful, even if only to show us how little we actually know.

We would ignore this extra variable if we were "screening," or randomly checking whether we have covid-19. In this case, we are not taking the test "because of something that happened," but just "because." However, "screening" is elusive — e.g., even if one believes that one is "screening," if one is doing so in order to determine whether it is safe to return home for the holiday, the fact that it is holiday season seems to actually make it so that they are no longer "screening". In that case, one is testing because it is near the holiday, and the holiday season affects the pre-test probability. Hence, one is rarely "screening," and there is a need to take other variables into account (e.g., see (Moons and Harrell 2003) and (Harrell 2021)).

The event "sustained exposure" is essentially one of those "other variables." I have therefore been trying to deal with the fact that we have this other information that is not taken into account in (Prince-Guerra et al. 2021) (I am not even sure whether they could have taken it into account). I have mostly just been writing out probability statements, and not made a ton of headway, but at least it seems there is a bound, which is common sense in retrospect.

The standard approach:

Define sx=symptoms, ad = adult, and dz=disease (covid-19). Note:

$$p(dz + |test-, sx) = \frac{p(test - |dz+, sx)}{p(test - |sx)}p(dz + |sx)$$

$$= \frac{1 - p(test + |dz+, sx)}{p(test - |dz-, sx)p(dz - |sx) + p(test - |dz+, sx)p(dz + |sx)}p(dz + |sx).$$

Study design:

Originally, I jumped to looking for the pieces of the equation above. We might look to a study for this purpose, but the validity of the estimates depends on the study design. Accoring to (Harrell 2020), a retrospective case-control is most appropriate. I am assuming in this case that we let the disease status be the outcome. I however need to think about this more; perhaps if (Prince-Guerra et al. 2021) does not fit this design, there may be issues with the approach I take below. Why do we want a retrospective observational instead of prospective or controlled study? Maybe the latter is obvious, and I am missing something. Below, I talk a bit about what would be "ideal," and my idea was just to record information on people as they test. I think this amounts to a retrospective case-control.

Note that, just focusing on $P(dz + |test, sx)$, we are really interested in the test effect. Let us therefore consider how this would be estimated in general. First, if the data is observational and there is an unmeasured variable that affects whether someone takes a test and also whether they have the disease (confounding), the effect estimate may be incorrect. Generally, we would condition on the confounder to remove this issue. I am trying to think more about how this relates to the discussion below on conditioning on the necessary set of variables. I think it is different, since we can defend a marginal (non-conditioned) post-test probability, if it is tailored enough to us to inform decision-making, but we can't really defend an unadjusted estimate of, e.g., a treatment effect. Both estimates however are similarly biased.

Estimates from (Prince-Guerra et al. 2021):

Assuming that the study design of (Prince-Guerra et al. 2021) allows for valid estimation, we can use their point estimates and 95% confidence intervals for

$$p(test + |dz+, sx+, ad+) \approx 0.64 \ (0.57, 0.71)$$

and

$$p(test + |dz+, sx-, ad+) \approx 0.36 \ (0.27, 0.45).$$

Also (todo: maybe I should have left this to two decimals),

$$p(dz - |test-, sx-, ad+) \approx 1 (1, 1)$$

and

$$p(dz - |test-, sx+, ad+) \approx 1 (1, 1).$$

We can then plug these into the equation above, and obtain a post-test probability, right?

Issues with the standard calculation:

Not so fast. If we define $x = (sx, ad)$, we see that it is a somewhat limited set. For example, we were originally interested in the post-test probability given a sustained exposure. Note,

$$P(dz + |test-, x) = P(dz + |test-, x, exp+)P(exp + |test-, x) + P(dz + |test-, x, exp-)P(exp - |test-, x)$$
$$= \alpha P(dz + |test-, x, exp+) + (1 - \alpha)P(dz + |test-, x, exp-)$$

where $\alpha \in [0, 1]$.

Hence, the probability of $P(dz + |test-, x)$ that we computed above is somewhere in between what we really want to know, $P(dz + |test-, x, exp+)$, and the $P(dz + |test-, x, exp-)$, which is the probability under no sustained exposure. This may appear to not be a big deal. However, consider now what happens in a similar situation, where we imagine the hypothetical scenario in which we did not measure symptom status,

$$p(dz + |test-) = (dz + |test-, sx+)P(sx + |test-) + P(dz + |test-, sx+)P(sx + |test-).$$

We see similarly that the marginal estimate is between the conditional estimates, based on (Prince-Guerra et al. 2021), and as I will show below graphically. Hence, the presence of symptoms matters. Likewise, we have no reason to believe that the presence of exposure does not matter.

This might go the other way; we may say, "well, symptoms seems to make a difference, but having the marginal estimate $p(dz + |test-)$ (somewhere in between) would not really change my behavior." Further, we may say, "Exposure matters, but it is unlikely that it is as important as e.g. symptoms, so we can expect the information loss we get by marginalizing over exposure to be even less than the information loss we get from marginalizing over symptoms, and therefore we might as well proceed with $P(dz + |test-)$ rather than $P(dz + |test-, exposure+)$."

Bound:

We can even make this a bit more precise by understanding that $P(dz+|test-) \leq P(dz+|test-, exposure+)$, which is based on the common sense fact that a sustained exposure increases probability of disease.

We can also somewhat more circuitously write

$$p(dz + |test-, sx, ad, exposure+) = \frac{p(exposure + |test-, dz+, sx, ad)}{p(exposure + |test-, sx, ad)}p(dz + |sx, ad, test-)$$

We see that if

$$\frac{p(exposure + |test-, dz+, sx, ad)}{p(exposure + |test-, sx, ad)} > 1,$$

which is the case when the event "exposure" increases the probability of the disease (I assume this is true), then

$$p(dz + |sx, ad, test-),$$

which we can obtain from (Prince-Guerra et al. 2021), is a lower bound on

$$p(dz + |exposure+, sx, ad, test-).$$

So, even though the estimates in (Prince-Guerra et al. 2021) do not apply to our situation directly, we can use them to obtain a lower bound on our post-test probability. It is still not clear whether this is even the true lower bound. When there are multiple factors that we are conditioning on, and when some of these factors have an unknown relationship with the disease status, then we can no longer even make definitive statements about the bound. However, if we were to assume that other variables, besides those concerning age cutoff, symptom status, test result, and exposure, are negligible, our post-test probability estimate can still be considered a lower bound, which still gives us some information, and also shows us, even with our strong assumptions, how little we actually know (I am reposting it here from above, but note now that I call the line a lower bound - also, I re-added the grid lines, which were lost before).

Other variables that might be relevant:

I am concerned with the omission of gender, which appears to be possibly correlated with viral load (Mahallawi et al. 2021), assuming viral load and antigen presence are essentially the same. I think age should be treated as a continuous variable. It seems that age should correlate with viral load; this was not supported by Mahallawi, but it might be supported by (Euser et al. 2021). We need to condition on anything that leads to different antigen levels in different people. If the antigens are excreted or metabolized, we would need to take into account liver and kidney status. Of course, this should depend on immune system function, which will vary along with comorbidities and medications. Antigen level also probably depends on the covid strain. I am also unsure how one should estimate $p(dz + |sx, ad)$. We can get this for the (Prince-Guerra et al. 2021) cohort, but this depends on things like location, time of year, and lockdown status. Some of these have changed a lot since the study was conducted.

Weakening the assumption of negligibility:

Perhaps the assumption that other variables in u are "negligible" is too strong. If we write $u = u^i \cup u^d$, where u^i is the set of variables that increases post-test probability and u^d is the set of variables that decreases post-test probability, then we might be able to get away with the assumption that the two sets sort of "cancel" eachother out. So, of the variables that we did not measure, some things will increase the likelihood of disease and others vice versa, but their total effect will be near enough to zero.

Implementation, assuming negligibility and proper estimates, under varying pre-test probabilities:

Likelihood ratio

```
like.rat=function(sens,spec,pretest){(1-sens)/((1-sens)*pretest+(spec)*(1-pretest))}
```

posterior probability of disease given negative test is likelihood ratio * prior

```
post.test.prob = function(sens,spec,pretest){
  like.rat(sens=sens,spec=spec,pretest=pretest)*pretest
}
```

compute over a grid of priors

```
pre.tests = seq(0,1,0.01)
get.post.tests = function(sens,spec,pretests){
post.tests = c()
for(i in 1:length(pre.tests)){
  post.tests[i]=post.test.prob(sens=sens,spec=spec,pretest=pre.tests[i])
}
post.tests
}
```

compute using, from (Prince-Guerra et al. 2021),

$$p(test + |dz+, sx+, ad+) \approx 0.64(0.57, 0.71)")$$

and

$$p(test + |dz+, sz-, ad+) \approx 0.36(0.27, 0.45)")$$

```
plot.p = function(post.tests.6,post.tests.6l,post.tests.6u,
                  post.tests.36,post.tests.36u,post.tests.36l,post.label){
lwd.point = 2
plot(pre.tests,post.tests.6,type='l',
```

```
      xlab="Pre-test: p(dz+|sx,ad)",ylab=post.label,
      lty=1,lwd=lwd.point,
      main=('Lower bound of post-test prob given exp,
            neg BinaxNOW test\n (sx=symptoms, ad=adult)'))
lines(pre.tests,post.tests.6u,lty=2,col=1)
lines(pre.tests,post.tests.6l,lty=2,col=1)
xseq = seq(0,1,0.1)
yseq = seq(0,1,0.1)
for(x in xseq){
  abline(v=x,col='gray')
}

for(y in yseq){
  abline(h=y,col='gray')
}

lines(pre.tests,post.tests.36,lty=1,lwd=lwd.point,col=2)
lines(pre.tests,post.tests.36u,lty=2,col=2)
lines(pre.tests,post.tests.36l,lty=2,col=2)
legend("topleft",
       c(TeX("sx+,ad+"),TeX("sx-,ad+")),
       lty=c(1,1),col=c(1,2),lwd=c(lwd.point,lwd.point))
}
```

```
post.tests.6=get.post.tests(sens=0.64,spec=1,pretests)
post.tests.6l=get.post.tests(sens=0.57,spec=1,pretests)
post.tests.6u=get.post.tests(sens=0.71,spec=1,pretests)

post.tests.36=get.post.tests(sens=0.36,spec=1,pretests)
post.tests.36u = get.post.tests(sens=.45,spec=1,pretests)
post.tests.36l = get.post.tests(sens=.27, spec=1,pretests)

png("post.test.png")
plot.p(post.tests.6,post.tests.6l,post.tests.6u,
                 post.tests.36,post.tests.36u,post.tests.36l,
       "Lower Bound Post-test: p(dz+|test-,sx,ad)")
dev.off()
```
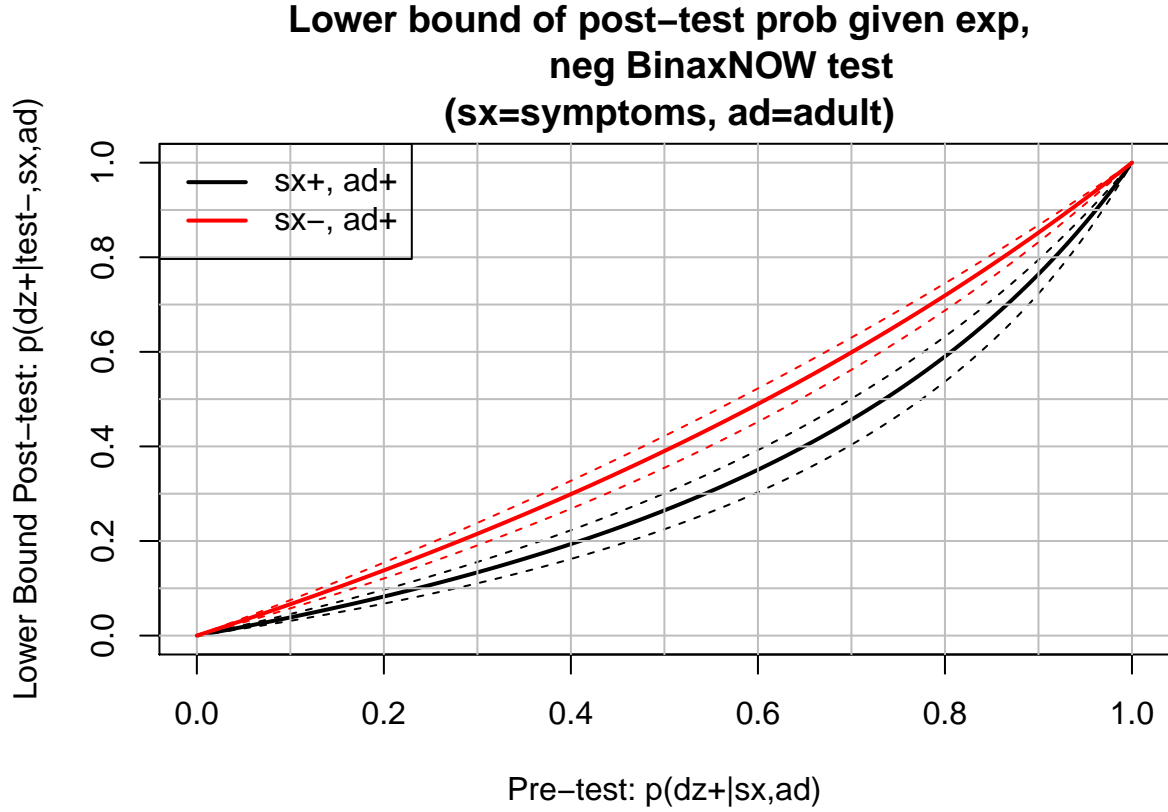
```
## pdf
##    2
```

```
plot.p(post.tests.6,post.tests.6l,post.tests.6u,
                 post.tests.36,post.tests.36u,post.tests.36l,
       "Lower Bound Post-test: p(dz+|test-,sx,ad)")
```

Lower bound of post–test prob given exp, neg BinaxNOW test (sx=symptoms, ad=adult)

Ideally:

Note that age is given to us as "adult" - a large loss of info, symptom status is binarized, test result is binarized, and exposure was binarized by me, unfortunately (although we could just treat it as continuous).

Note that to obtain the graph above, we are working with what we have available. However, the following would be ideal: every time one takes a test, one goes to a website and enters information such as age, sex, zip code, and test result. The website then calculates post-test probability based on a model for, explicitily,

$$p(dz + |test = r, age = a, sex = s, symptoms = sx, zip = z, ..)$$

The model would be updated in real-time based on geographically and temporally relevant statistics. Note that this post-test probability would depend very much on time, and therefore the model would have to be updated probably each day. It seems though that, barring sampling issues, we have this data. There is observational data collected when people report their test results (in other words, we collect information such as age, zip code, etc). Often, also, we have a PCR confirmation. It is unclear however whether we will have enough of the variables mentioned above, which are still necessary.

Also, ideally, the test result would be a continuous variable (eg, amount of viral load). This may be difficult due to at-home testing kit constraints. However, it seems currently that there is some cutoff above which a test is called positive. If it is possible for the tests to convey more information, such as through color or some type of numerical scale, it would lead to better estimates of the post-test probability (assuming there is no real hard cutoff - I am not sure).

On the test cutoff (if there is a cutoff):

It is not clear this is how it works, but, in general, test cutoffs have highly significant implications. If indeed there is a cutoff, what is the reward function that is being optimized? It appears that these tests were designed to minimize false positive results. However, that is not, in general, always a good idea. Decreasing false positives (e.g., by setting a high cutoff) also increases false negatives. In general, for someone who works in a highly populated area, or with vulnerable populations, a false negative is worse than a false positive.

Serial testing:

I have also done some more thinking on serial testing - my current thinking (maybe this is not correct, I need to still write it out here) is that **if the tests are independent**, you can essentially treat the post-test probability from the first test as the pre-test probability for the second. If this is the case, then two tests taken, premeditated, in sequence, will perform like a better test. I hope to eventually provide a post-test graph, as above, for independent sequential tests. Assuming we have conditioned on everything we need, it will still give a lower bound for the "exposure" case.

Generally, though, it is also advised to take the two tests e.g. 24 hours apart to see if the viral load increases during that time. I am not sure that two tests that are taken like this are still independent.

Positive tests:

Note that I am focusing only on the negative test case, although you could do the same for positive tests (I said originally that this was a non-issue, but I should not have — you can, of course, have a positive test with no disease, and I should repeat the analysis above for that case).

Code: https://github.com/samuelweisenthal/BinaxNOWpostTestProb

I appreciate conversations with the datamethods forum participants, Anna Park, and my brother on this topic.

References

Euser, Sjoerd, Sem Aronson, Irene Manders, Steven van Lelyveld, Bjorn Herpers, Jan Sinnige, Jayant Kalpoe, et al. 2021. "SARS-Cov-2 Viral Load Distribution Reveals That Viral Loads Increase with Age: A Retrospective Cross-Sectional Cohort Study." *medRxiv*. Cold Spring Harbor Laboratory Press.

Harrell, Frank E. 2020. https://www.fharrell.com/post/backwards-probs/.

———. 2021. http://hbiostat.org/doc/bbr.pdf.

Mahallawi, Waleed H, Ali Dakhilallah Alsamiri, Alaa Faisal Dabbour, Hamdah Alsaeedi, and Abdulmohsen H Al-Zalabani. 2021. "Association of Viral Load in Sars-Cov-2 Patients with Age and Gender." *Frontiers in Medicine* 8. Frontiers: 39.

Moons, Karel GM, and Frank E Harrell. 2003. "Sensitivity and Specificity Should Be de-Emphasized in Diagnostic Accuracy Studies." *Academic Radiology* 10 (6). Elsevier: 670–72.

Prince-Guerra, Jessica L, Olivia Almendares, Leisha D Nolen, Jayleen KL Gunn, Ariella P Dale, Sean A Buono, Molly Deutsch-Feldman, et al. 2021. "Evaluation of Abbott Binaxnow Rapid Antigen Test for Sars-Cov-2 Infection at Two Community-Based Testing Sites—Pima County, Arizona, November 3–17, 2020." *Morbidity and Mortality Weekly Report* 70 (3). Centers for Disease Control; Prevention: 100.