On determining the post-test probability of covid-19

Summary:

Covid-19 tests are crucial. Interpreting tests, however, is not necessarily just an application of Bayes rule. A test that does not provide a means of interpretation (e.g., a website where the user can enter their result and information and receive a post-test probability of disease) might be hazardous to the user (and people around the user). Here, I will try to provide an interpretation of a negative test, discuss some of the difficulties, and mention a more systematic solution, which has been discussed previously in the literature.

In more detail:

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 $p(dz|test, x) = p(test|dz, x)p(dz|x)/p(test|x)$, where dz=covid and x is some set of covariates. 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 covariates 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).
5. 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 that depend on local prevalence, time of year, symptoms, known exposures, time since exposure, etc. . .
6. Sometimes this is important, sometimes it's not. Eg, based on (Prince-Guerra et al. 2021), a study that investigated the BINAX Now covid rapid tests, the presence of symptoms can change the post-test probability estimate significantly. Hence, if we ignored the fact that we have symptoms when estimating post-test probability, we would obtain an estimate that is too low.
7. It is possible sometimes that, for specific variables in z for which we understand the correlation with disease,assuming other variables we have not taken into account are negligible or cancel eachother out, we can make statements about bounds on $p(dz|test, x, z)$ using $p(dz|test, z)$.
8. Overall, we should maintain location-specific, continually updating databases of $dz, test$ and covariates that are relevant, and we should also perform more studies to try to see what should be included in this set of covariates.

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.

Screening:

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." It appears actually that (Prince-Guerra et al. 2021) did have this data available, but the raw data is not provided. The sample size is fairly large in this study, but there appear to be only 4 false positives.

The standard approach:

Define sx=symptoms, ad = adult, and dz=disease (covid-19). I am including symptom status and "adult" as done in (Prince-Guerra et al. 2021), but this is more for the sake of illustration than to give an accurate estimate. I will include some charts from a more recent meta-analysis below.

Note:

$$
\begin{aligned}
p(dz+|test-,sx) &= \frac{p(test-|dz+,sx)}{p(test-|sx)}p(dz+|sx) \\
&= \frac{1-p(test+|dz+,sx)}{p(test-|dz-,sx)p(dz-|sx)+p(test-|dz+,sx)p(dz+|sx)}p(dz+|sx) \\
&= \frac{1-sens}{spec(1-pretest)+(1-sens)pretest}pretest
\end{aligned}
$$

So, we have an equation that seems to allow us to find the post-test probability given commonly estimated test operating characteristics. Originally, I jumped to looking for studies to help me fill in the pieces of the second to last line of the display above [Note that in the last line, I rename things. This renaming, interestingly, can cause some difficulty in interpretation; we understand the terms in the second to last line, but the renaming in the last line renames and strips the terms of their probabilistic meaning (along with removing information about what is being conditioned upon), which can lead to misuse]. It is worth noting however that test interpretation would be much better if there were databases kept such that we could just estimate $p(dz+|test-,sx,$ other variables) directly rather than having to use this equation. This is completely possible, and would be much better and easier for the user of the test (I discuss this more below in what I call the "ideal" scenario). This type of modeling does not occur usually because each model would be specific to time and location, at least. This sounds like a daunting task, therefore, to fit all of these models, but the reality is that if this is not done that way, it has to be done another way — it just shifts the responsibility of interpretation from the test manufacturer or some independent organization onto the person using the test. So, we have no choice - let's try to do this.

Study design:

We first need to fill in the pieces of the right hand side 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. Below, when I talk a bit about what would be "ideal," my idea is 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 still 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. This actually seems to be the tension here. In some cases, marginal, $p(dz+|test-)$, is good enough, even though ideally we would have conditional $p(dz+|test-, age, sex, location, time, incubation, \dots)$. Especially, marginal is good enough if somehow the variables we do not condition on in the marginal estimate sort of cancel each other out.

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),

$$p(dz-|test-, sx-, ad+) \approx 1(.99, 1)$$

and

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

We can then plug these estimates 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 the situation where instead of exposure, we did not measure symptom status:

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

According to (Prince-Guerra et al. 2021), however, the presence of symptoms matters. Hence, if we blindly compute the left hand side above, we will be estimating our post test probability ignoring symptoms, and therefore underestimate our risk. With the equation above, this is very clear that this should be the case. However, the mechanical use of Bayes rule to estimate post-test probability often leads us to ignore this fact, surprisingly.

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 $p(dz+|test-, exposure+)$." Hence, in the beginning when I said that a test without an interpretation mechanism is like a car without a windshield, I was somewhat cautious with this statement. Only gathering the appropriate data can tell us whether it is like a car without a windshield, a car without an air refreshener, or a car without brakes.

Bound:

The following section is just some musings on possibly being able to make a statement like "disease probabilty is greater than this value, even though we know it is not this value precisely." We can say, for example, $p(dz + |test-) \leq p(dz + |test-, exposure+)$, which is based on the 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:

As I said above, we can only know how good or bad our current estimates are by collecting more data. However, one can also use previous studies to detect variables that might be relevant, but are ignored. 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. It appears that Table 2 in (Prince-Guerra et al. 2021) provides some interesting information that can help us here (there are counts of eg false positive and negative rates conditioned on sex, ethnicity, etc; unfortunately not jointly - I will have to look for the raw data). I am just eyeballing it, and it appears that the percentage of FN differs by sex, ethnicity, race, age, symptoms, days from symptom onset, and known exposure (!!).

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. One can avoid having to think about the latter if things are done with updating, location-specific databases, as mentioned above. It really should not be the responsibility of the test taker to think about the latter, especially because it depends on the covariates chosen in the study, and it is unreasonable to expect that someone can guess, without data, the value of their conditional pre-test probability.

Weakening the assumption of negligibility:

I said before that possibly if the variables we do not condition on have equally opposing effect, we may be able to ignore them. This could allow us to relax the assumption that these variables are negligible. We really though have to gather data to see if this is the case.

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

Here, I am going to calculate the post-test probability given a negative test for the BINAX Now rapid antigen test under the assumption that the variables not conditioned upon in the (Prince-Guerra et al. 2021) are negligible, either in effect or in total effect, and that the study had a reasonable design.

We start with the likelihood ratio; recall from above that

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

Note that this implies that higher sensitivity will increase post-test probability of the disease being present and higher specificity will decrease post-test probability. Hence, our lower bound will be the upper confidence bound of sensitivity and the lower confidence bound of specificity and vice versa for the upper bound.

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

We then get that the posterior probability of disease given negative test is likelihood ratio times the pretest probability.

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

Let us compute the post-test probability over a grid of pre.tests.

```
pre.tests = seq(0,1,0.01)
get.post.tests = function(sens,spec,pre.tests){
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
}
```

We have, recall, from (Prince-Guerra et al. 2021),

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

and

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

Now just plot

```r
plot.p = function(lns,post.label,xlab,
                  leg.labels,main){

lwd.point = 2
plot(c(0,0),pch=0,xlim=c(0,1),ylim=c(0,1),
     xlab=xlab,ylab=post.label,
     lty=1,lwd=lwd.point,main=main,col='white')

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')
}
colorBlindBlack8  <- c("#000000", "#E69F00", "#56B4E9", "#009E73",
                       "#F0E442", "#0072B2", "#D55E00", "#CC79A7")
nl = length(lns)
for (i in 1:nl){
lines(pre.tests,lns[[i]]$m,lty=1,lwd=lwd.point,col=colorBlindBlack8[i])
lines(pre.tests,lns[[i]]$u,lty=2,col=colorBlindBlack8[i])
lines(pre.tests,lns[[i]]$l,lty=2,col=colorBlindBlack8[i])
}


legend("topleft",legend=leg.labels,bg='white',
       lty=rep(1,nl),col=colorBlindBlack8[c(1:nl)],lwd=c(lwd.point,lwd.point))
}
```

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

l.6 = list(m=post.tests.6,l=post.tests.6l,u=post.tests.6u)

post.tests.36=get.post.tests(sens=0.36,spec=1,pre.tests)
post.tests.36u = get.post.tests(sens=.45,spec=.99,pre.tests)
#post.tests.36l = get.post.tests(sens=.27, spec=1,pre.tests)
post.tests.36l = get.post.tests(sens=.27, spec=1,pre.tests)

l.36 = list(m=post.tests.36,l=post.tests.36l,u=post.tests.36u)

# Prince-Guerra 2021

lns = list(l.6,l.36)


plot.p(lns,
       post.label="Lower Bound Post-test: p(dz+|test-,sx,ad)",xlab="Pre-test: p(dz+|sx,ad)",
       leg.labels=c(TeX("sx+,adult+"),TeX("sx-,adult+")),main="Prince-Guerra, 2021"
       )
```
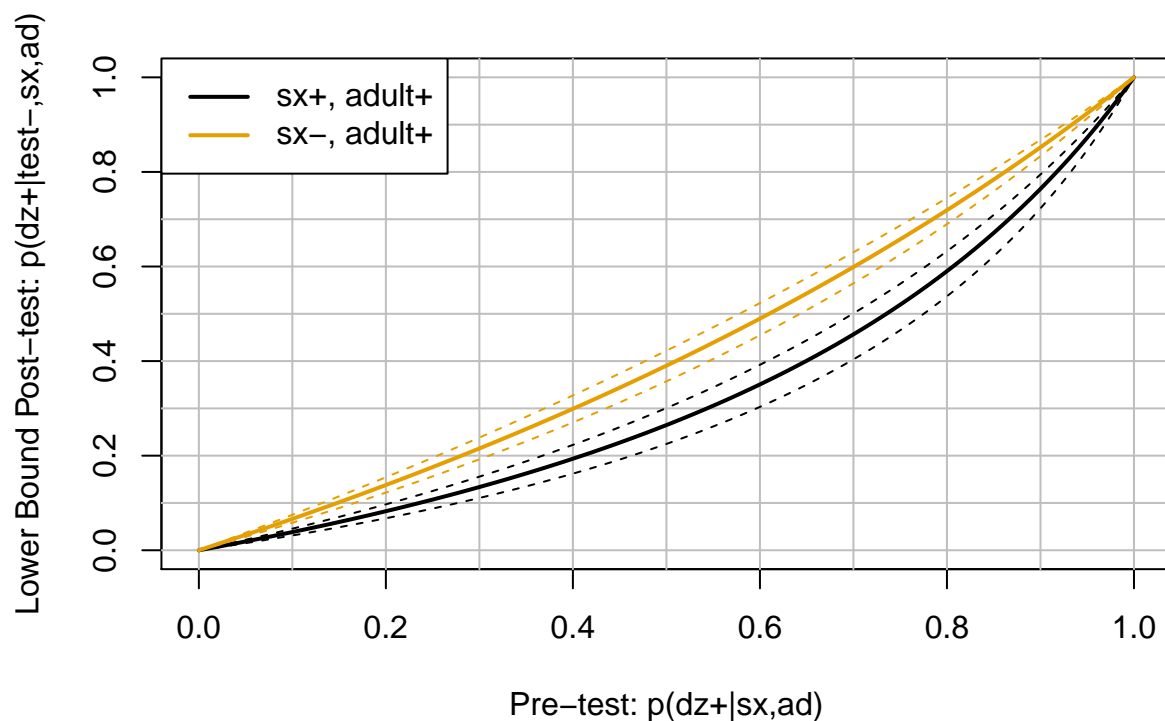
Prince–Guerra, 2021

Now consider numbers from the recent meta-analysis (Bruemmer et al. 2021). Note that these estimates, as far as I know, condition on nothing except the test result [and adult status?] (I do not know how they would otherwise pool so many estimates). Hence, we have pre-test as

$$p(dz+)$$

and post-test as

$$p(dz + |test-).$$

Here I have converted the table from (Bruemmer et al. 2021) into a list.

```
tests=list(
veritor=list(
  nm='BD Veritor',
  sens=list(m=.64,l=.49,u=.76),
  spec=list(m=1,l=.99,u=1)
),
Binax=list(
  nm='BinaxNOW',
  sens=list(m=.62,l=.48,u=.74),
  spec=list(m=1,l=1,u=1)
),
Clinitest = list(
  nm='CLINITEST',
  sens=list(m=.62,l=.47,u=.75),
  spec=list(m=.99,l=.97,u=.99)
),
Coris = list(
```

```
    nm="Coris",
    sens=list(m=.40,l=.29,u=.52),
    spec=list(m=.99,l=.95,u=1)
),
LumipulseG =  list(
    nm="Lumipulse G",
    sens=list(m=.87,l=.78,u=.93),
    spec=list(m=.97,l=.89,u=.99)
),
LumiraDx =  list(
    nm="LumiraDx",
    sens=list(m=.88,l=.59,u=.98),
    spec=list(m=.99,l=.96,u=1)
),
Panbio =  list(
    nm="Panbio",
    sens=list(m=.72,l=.65,u=.78),
    spec=list(m=.99,l=.99,u=1)
),
Rapigen =  list(
    nm="Rapigen",
    sens=list(m=.62,l=.47,u=.75),
    spec=list(m=.99,l=.94,u=1)
),
Sofia =  list(
    nm="Sofia",
    sens=list(m=.77,l=.74,u=.80),
    spec=list(m=.99,l=.98,u=1)
),
StandardF =  list(
    nm="StandardF",
    sens=list(m=.68,l=.56,u=.79),
    spec=list(m=.98,l=.97,u=.99)
),
StandardQ =  list(
    nm="StandardQ",
    sens=list(m=.75,l=.69,u=.80),
    spec=list(m=.99,l=.98,u=1)
),
StandardQnasal =  list(
    nm="StandardQnasal",
    sens=list(m=.80,l=.70,u=.87),
    spec=list(m=.99,l=.98,u=1)
)
)
```

We will plot the tests in threes, just so the plot is not too cluttered.

```
get.t.posts = function(tests){
test.posts=list()
for(t in tests){
  test.posts[[t$nm]]=list(
    m=get.post.tests(sens=t$sens$m,spec=t$spec$m,pre.tests),
    l=get.post.tests(sens=t$sens$l,spec=t$spec$u,pre.tests),
```
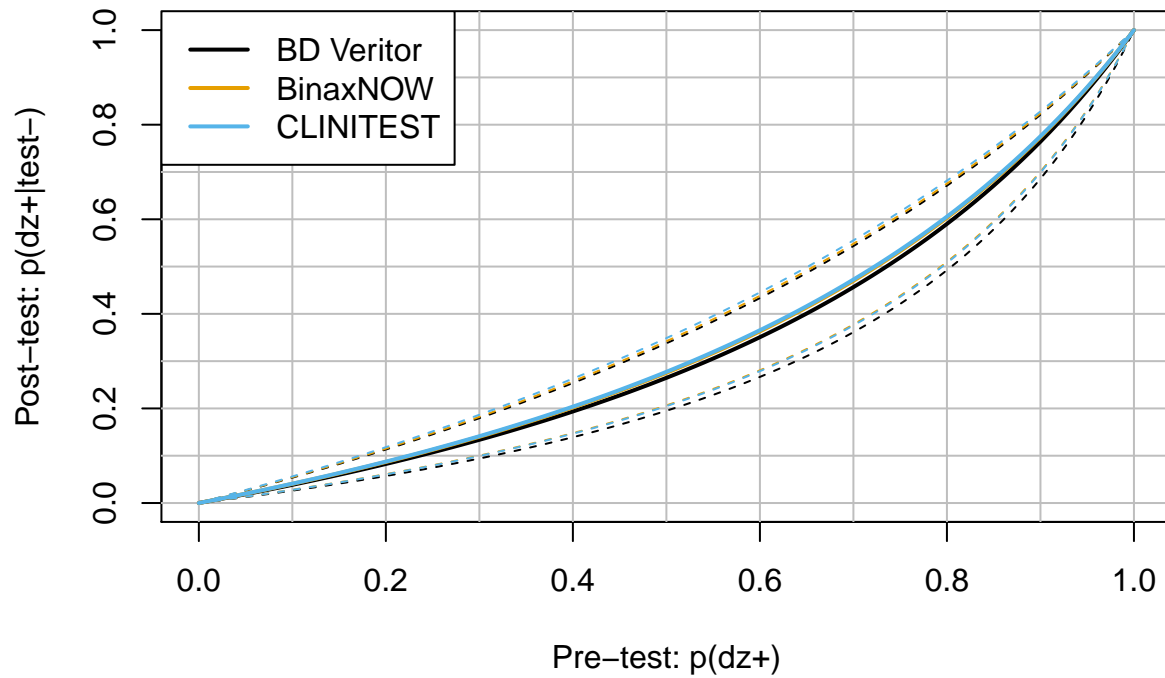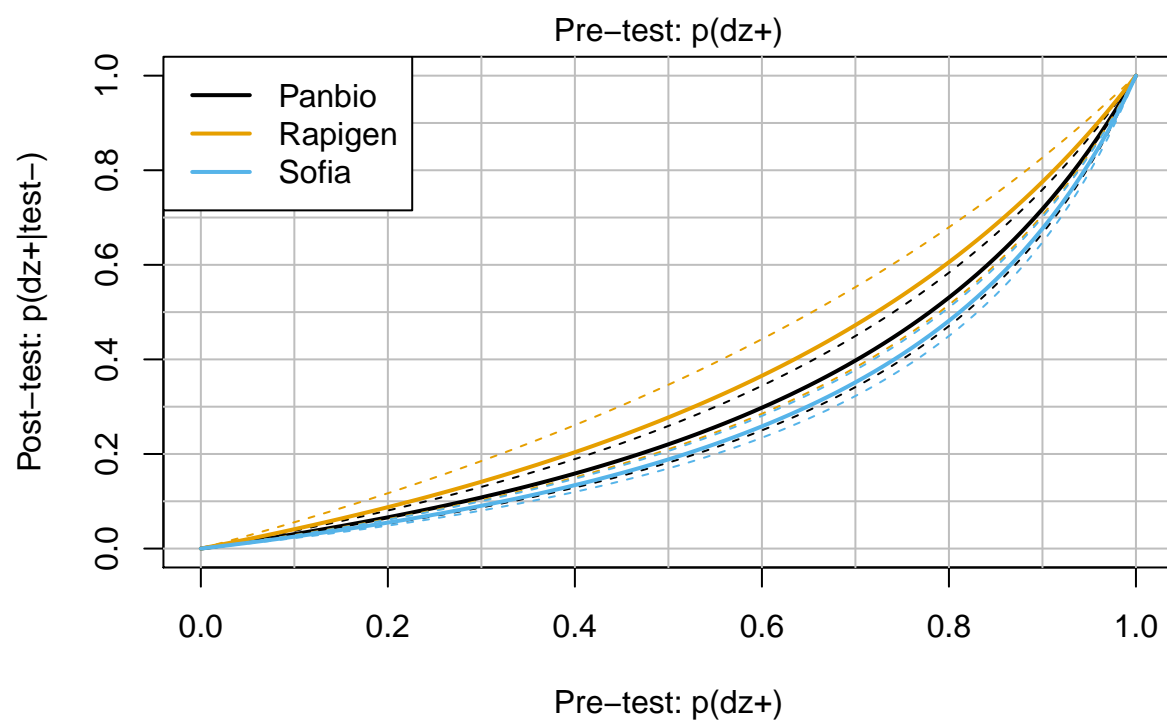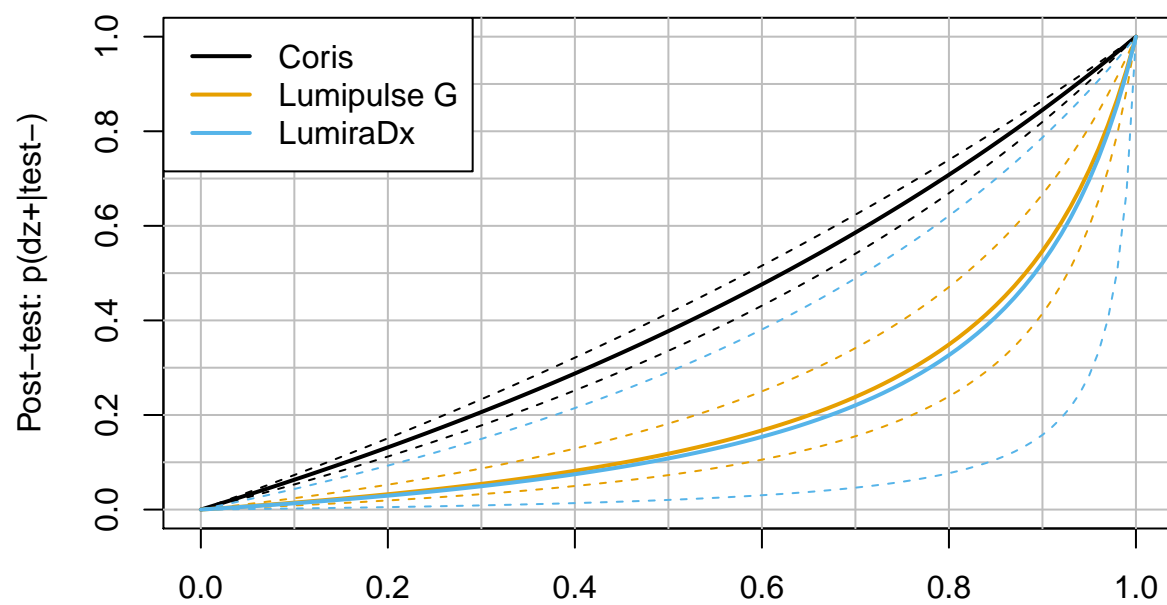
```
    u=get.post.tests(sens=t$sens$u,spec=t$spec$l,pre.tests)
    )

}
test.posts
}

ms=c(seq(1,12,3),12)
ls = length(ms)
#par(mfrow=c(2,2))
for (k in 1:(ls-1)){
en=ms[k+1]-1
if(k==4){
  en=12
}
test.posts = get.t.posts(tests[ms[k]:en])
plot.p(test.posts,
       post.label="Post-test: p(dz+|test-)",xlab="Pre-test: p(dz+)",
       leg.labels=names(test.posts),main=""
       )
}
```
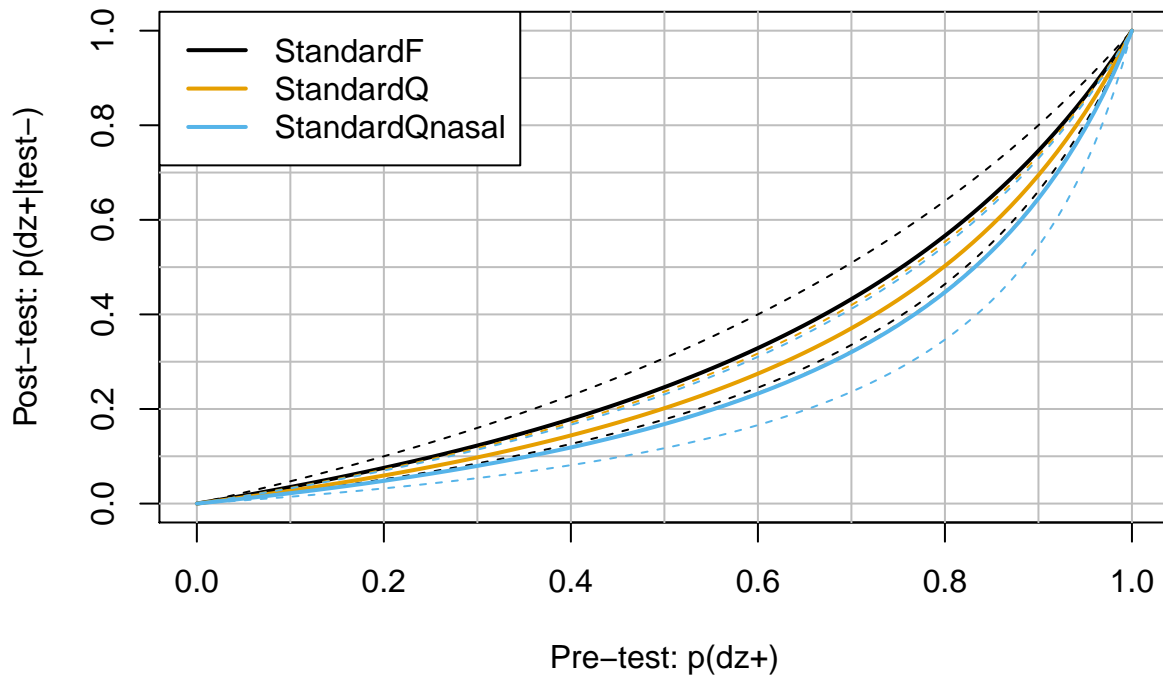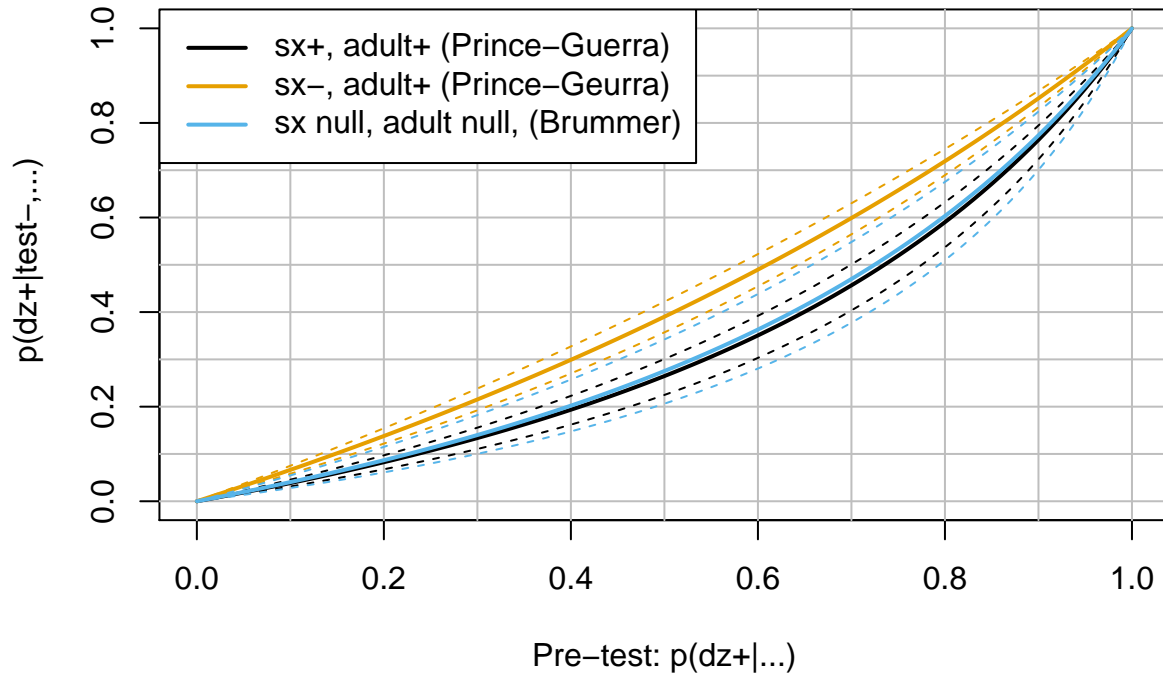
We can compare $p(dz + |test-, sx, adult)$ from (Prince-Guerra et al. 2021) with $p(dz|test-)$ from (Bruemmer et al. 2021) (although really we could just compare the numbers...).

```
pts=get.t.posts(tests)
#pts$BinaxNOW
lns = list(l.6,l.36,pts$BinaxNOW)


plot.p(lns,
        post.label=" p(dz+|test-,...)",xlab="Pre-test: p(dz+|...)",
        leg.labels=c(TeX("sx+,adult+ (Prince-Guerra)"),TeX("sx-,adult+ (Prince-Geurra)"),"sx null, adult
        )
```

Prince–Guerra vs Brummer

It seems that the asymptomatic just override the symptomatic (makes sense, fewer symptomatic.) If this is true, it is important. It means that the difference between the conditional and marginal estimates will be asymmetric. I think this is common sense, going back to α above, which combines the two conditional distributions.

$$
\begin{aligned}
p(dz + |test-) &= p(dz + |test-, sx+)p(sx + |test-) + p(dz + |test-, sx-)p(sx - |test-) \\
&= p(dz + |test-, sx+)p(sx + |test-) + p(dz + |test-, sx-)(1 - p(sx + |test-)) \\
&= p(dz + |test-, sx+)\alpha + p(dz + |test-, sx-)(1 - \alpha),
\end{aligned}
$$

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

It should (I think) be more likely that a symptom is negative given that a test is negative. There is no reason to believe, we see now, that the marginal $p(dz + |test-)$ will be halfway between $p(dz + |test-, sx+)$ and $p(dz + |test-, sx-)$. Hence, if you are asymptomatic, you will get a decent (conditional) estimate; if you are symptomatic, you will be pulled toward asymptomatic, and get a poor (conditional) estimate.

I am going to explore this a bit more. We will make our variables $Z \sim Bern(expit(\beta X))$.

```
expit=function(x){exp(x)/(1+exp(x))}
plot(-10:10,expit(-10:10))
```
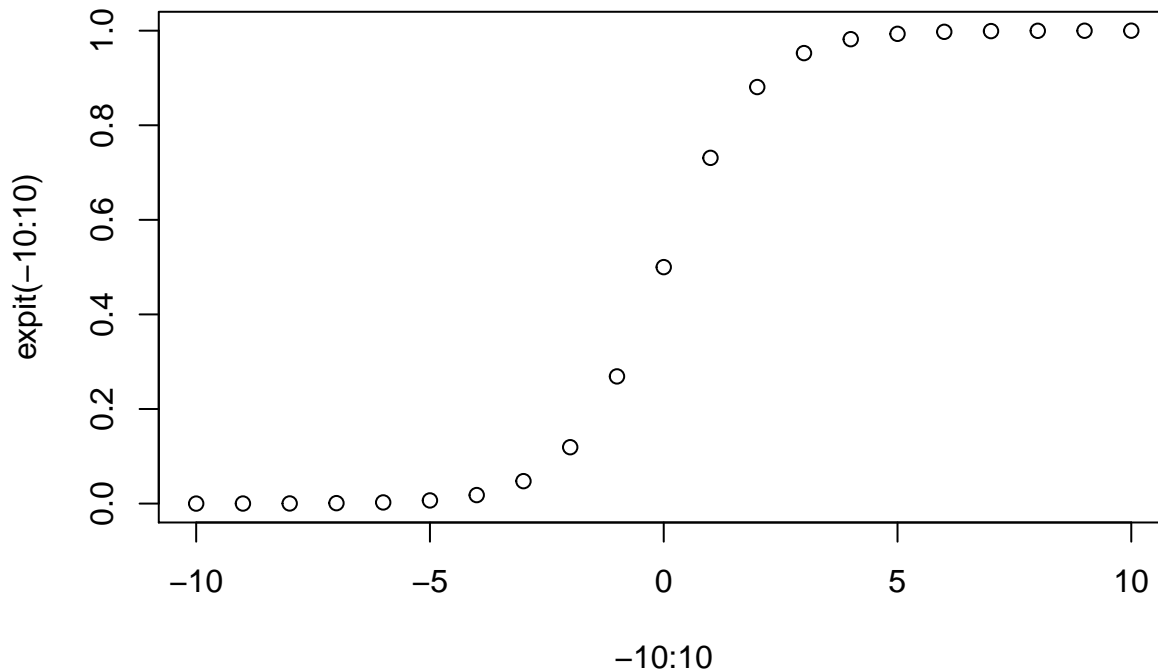
Set the sample size.

```
set.seed(1)
n=10000
```

Set prevalence, pdz.

```
pdz=.1
```

Let $dz \sim Bern(pdz), sx \sim Bern(expit(-2 + (2 + 2)dz))$, so we generally don't have symptoms unless disease is present. Let $ts \sim Bern(expit(-3 + (3 + 2.5)dz + 2.5(sx)(dz)))$, so we tend not to have positive tests unless there is disease, and the presence of symptoms along with the disease increases the probability greatly.

```
dz = rbinom(n,1,pdz)
sx=rbinom(n,1,expit(-2+(2+2)*dz))
#antigen = rnorm(n,mean=-3+sx*.1+dz*1) #todo: explore cutoff
ts = rbinom(n,1,expit(-3+(2+1.5)*dz+1.5*sx*dz))
df = data.frame(dz=dz,sx=sx,ts=ts)
```

Calculate the marginal sensitivity and specificity

```
ptspdzp=dim(df[(df$dz==1)&(df$ts==1),])[1]/dim(df[(df$dz==1),])[1]
ptsndzn=dim(df[(df$dz==0)&(df$ts==0),])[1]/dim(df[(df$dz==0),])[1]
print(c("marg sens",round(ptspdzp,2)))
```

```
## [1] "marg sens" "0.84"
```

```
print(c("marg spec",round(ptsndzn,2)))
```

```
## [1] "marg spec" "0.96"
```

Calculate the conditional sensitivities and specificities. First, given symptoms

```
ptspdzpsxp = dim(df[(df$ts==1)&(df$dz==1)&(df$sx==1),])[1]/dim(df[(df$dz==1)&(df$sx==1),])[1]
ptsndznsxp = dim(df[(df$ts==0)&(df$dz==0)&(df$sx==1),])[1]/dim(df[(df$dz==0)&(df$sx==1),])[1]
print(c("Cond sens, sx-",round(ptspdzpsxp,2)))
```

```
## [1] "Cond sens, sx-" "0.87"
```

```r
print(c("Cond sens, sx-",round(ptsndznsxp,2)))
```

```
## [1] "Cond sens, sx-" "0.97"
```

Now no symptoms

```r
ptspdzpsxn = dim(df[(df$ts==1)&(df$dz==1)&(df$sx==0),])[1]/dim(df[(df$dz==1)&(df$sx==0),])[1]
ptsndznsxn = dim(df[(df$ts==0)&(df$dz==0)&(df$sx==0),])[1]/dim(df[(df$dz==0)&(df$sx==0),])[1]
print(c("Cond spec, sx+",round(ptspdzpsxn,2)))
```

```
## [1] "Cond spec, sx+" "0.63"
```

```r
print(c("Cond spec, sx+",round(ptsndznsxn,2)))
```

```
## [1] "Cond spec, sx+" "0.95"
```

```r
psxptsn = dim(df[(df$sx==1)&(df$ts==0),])[1]/dim(df[(df$ts==0),])[1]
psxntsn = dim(df[(df$sx==0)&(df$ts==0),])[1]/dim(df[(df$ts==0),])[1]
print(c("P(sx+|test-)",round(psxptsn,2)))
```

```
## [1] "P(sx+|test-)" "0.13"
```

```r
print(c("P(sx-|test-)",round(psxntsn,2)))
```

```
## [1] "P(sx-|test-)" "0.87"
```

```r
tests=list(
condsxp=list(
  nm='Conditional sx+',
  sens=list(m=ptspdzpsxp,l=ptspdzpsxp,u=ptspdzpsxp),
  spec=list(m=ptsndznsxp,l=ptsndznsxp,u=ptsndznsxp)
  ),
marg=list(
  nm='Marginal',
  sens=list(m=ptspdzp,l=ptspdzp,u=ptspdzp),
  spec=list(m=ptsndzn,l=ptsndzn,u=ptsndzn)
),
condsxn=list(
  nm='Conditional sx-',
  sens=list(m=ptspdzpsxn,l=ptspdzpsxn,u=ptspdzpsxn),
  spec=list(m=ptsndznsxn,l=ptsndznsxn,u=ptsndznsxn)
  )
)

pts=get.t.posts(tests)

plot.p(pts,
       post.label="p(dz+|test-,...)",xlab="Pre-test: p(dz+|...)",
       leg.labels=c(TeX("Conditional: p(dz+|test-,sx+)"),"Marginal: p(dz+|test-)",TeX("Conditional: p(d
       )
```

Conditional vs. marginal (simulation)
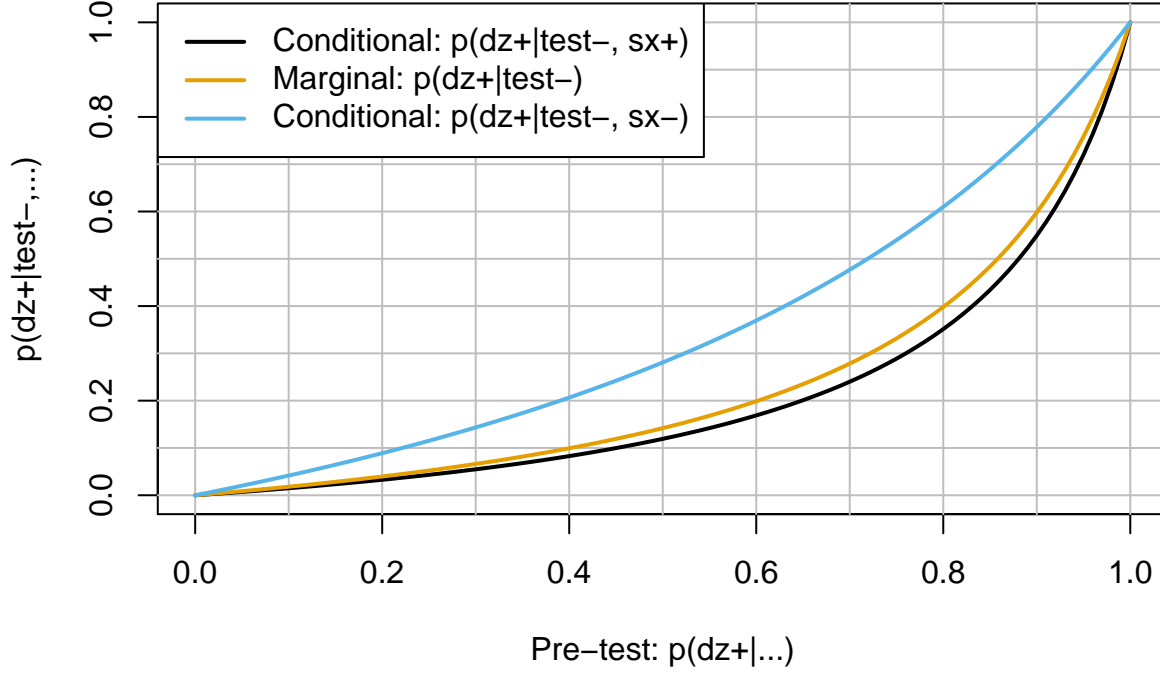
This is contrary to my expectation that we would interpolate between the two conditionals with weight placed according to $p(sx|test-)$.

$$p(dz+|test-) = p(dz+|test-,sx+)p(sx+|test-) + p(dz+|test-,sx-)p(sx-|test-)$$
$$= p(dz+|test-,sx+)p(sx+|test-) + p(dz+|test-,sx-)(1 - p(sx+|test-))$$
$$= p(dz+|test-,sx+)\alpha + p(dz+|test-,sx-)(1-\alpha),$$

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

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,

$$p(dz+|test, age, sex, symptoms, location, time, incubation, strain, ..)$$

The database and corresponding 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. Therefore, it should be a priority not to include a threshold in the test, if possible.

As a thought experiment, consider a world in which people do not interpret tests, and just act according to the test result. In other words, if I test negative, I assume I have no disease, and I act accordingly. In this case, I believe that the cutoff will set the realized reward entirely. In other words, it is important to interpret a test in order to avoid having the reward function be set externally by the cutoff. As we have seen, interpretation is complicated. Therefore, it is not unlikely that many people do act as if the test result is the true result, and therefore if there is a cutoff in the test, it has broad implications in practice.

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. If this is the case, using the plot, you can just start with a pretest and update it, and then make your next pretest your post-test from the first test.

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.

Viral load:

I mentioned viral load above. This was mentioned in a blog post https://www.lesswrong.com/posts/Sct2SNhByS8Lzenub/rapid-antigen-tests-for-covid. Essentially, viral load (obviously) affects sensitivity and specificity. It would be important to measure the time of supposed exposure in order to take this into account $p(dz + |..., \text{exposure time}, ...)$.

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

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

I appreciate comments and encouragement from the datamethods https://discourse.datamethods.org/t/determining-post-test-probability-of-covid-19/4723/12 forum participants (epiMD5,trumanfrancis), Anna Park, and my brother on this topic.

References

Bruemmer, Lukas E, Stephan Katzenschlager, Mary Gaeddert, Christian Erdmann, Stephani Schmitz, Marc Bota, Maurizio Grilli, et al. 2021. "The Accuracy of Novel Antigen Rapid Diagnostics for Sars-Cov-2: A Living Systematic Review and Meta-Analysis." *medRxiv*. Cold Spring Harbor Laboratory Press.

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.