Program Documentation: ScreeningTest.py: (Medical Screening Tests)

First Please Watch A Video Introducing The Screening Test Model

https://youtu.be/EuKDZNXmOU8

Try Downloading This Document For Best Viewing

You may want to DOWNLOAD THIS PDF file to read it in a full featured PDF viewer. (E.g. Chrome). This document has a convenient PDF features such as an navigation window, searching and active links which will only be available in a full featured pdf viewer.

How To Run The Screening Test Program

This program is run in a web page. Navigate to the following link in any major web browser.
The program will then instruct you on how to proceed.
The "About" menu on the GUI guides the user to all files and documentation and a documentary video.

https://share.streamlit.io/profbrockway/screeningtest/main/screeningtest.py

Program And Project Purpose

This program provides graphing and statistical reporting of a typical medical screening test according the test's parameters input by the user. The program reports the efficacy statistics for the specified test over a range of disease prevalences.

The project uses this tool to explore the effect of disease prevalence on the false positives rate. (FPR)

As a specific example this project models one of the Covid 19 antigen screening tests.

Project Documentation

This document and all the project files are hosted at GitHub: Project ScreeningTest.py At Github.

The program is self-documenting. See the program's web page "About" menu for all information.

The GUI also contains contacts, error and feedback reporting etc.

Contacts

Course: CCSU Stat 476. Spring 2022.
Author: Tim Brockway. Student ID: 30259316 Email: BrockwayTim@My.CCSU.edu
Professor: Roger L. Bilisoly. bilisolyr@ccsu.edu https://www2.ccsu.edu/faculty/bilisolyr

Program Input / Outputs

Inputs

- A GUI in the form of a web page:
  - The user will enter the following medical test statistics:
    - Test sensitivity.
    - Test specificity.
    - A slider that defines any interval of prevalences to the plotted.
      - The graph initially displays the plots over the [0%, 100%] prevalence.
      - A slider on the graph allows the user to zoom in on any smaller interval.
      - The plot is set up for more detail if you zoom in on realistic prevalence levels: [0%, 5%]
    - Disease "Prevalence Of Interest". A report will be made on this prevalence.
    - A text annotation to label the program report.
    - Population.

Outputs

- A printable report on the input and calculated results.

- An interactive graph with:
  x axis. The specified range of disease prevalences.
  y axis: Any or all of the following:
    - False positive percentage.
    - False negative percentage.
    - Positive predictive value.
    - Negative predictive value.
    - False Positives (NPV)
    - False Negatives (NPV)

- The program permits the user to show or hide the plotted variables to avoid clutter and examine different aspects of screening test performance.

- A display of a data grid showing the data generated from the users input and used for the plots.
  - The most interesting entries will are highlighted. The grid is fully scrollable.

- Error messages reporting errors in user's input.

- Standard webpage links to all documentation, code, error reporting, contacts etc.

- Optional downloads to users' computer:
  - A report on the screen test statistics as a text file.
  - The data grid as a CSV file with the plot data.
  - The data grid as an Excel file with the plot data.
  - The plots as a static images.
  - The plots as an *interactive* html web pages.

COVID 19 Screening Tests: Background

Basics Of Medical Screening Tests Assumed

It is assumed the reader is familiar with the basics of medical screening tests, their statistical variables and the difficulties of estimating the test's efficacy. Those basics are well documented by the CDC, FDA and other websites. The statistical variables of screening tests are nicely summarized here: Wiki. Additionally the program code itself explains all variables and calculations.

Evaluating Screening Tests Is Very Complicated

Judging the efficacy of screening tests is subtle and complicated. We can't go into this vast subject too deeply but it's important to keep in mind that simplistic summaries of screening tests using just sensitivity and specificity are extremely misleading. This fact should always kept in mind when considering the Covid 19 screening tests and how effective they are in their stated role. The linked article gives an idea of the many pitfalls in research and practice of screening tests and their statistics. FIS

Medical Screening Tests And The Low Prevalence Problem

One of the many problems with medical screening tests is that the prevalence of the subject disease in the tested population effects the proportion of true positives and true negatives.

> *Uneven test accuracy and statistical challenges, especially in areas of low disease prevalence, further complicate use of antibody tests for individual decision making.*
> *For example, a test with 98% specificity at 0.1% prevalence, the PPV would only be 4%, meaning that 96 out of 100 positive results would be false positives*. FDA
>
> *Unfortunately, the false positive rate can be shockingly high. Based on the prevalence estimated throughout the US and serology studies in California, New York and Boston, the FPR (False Positive Rate) of antibody test results* [for Covid 19] *range from 2% to 88%.* Nature (Refers to antibody tests but applies to all testing logic including the usually less accurate antigen screening tests.).

Definitions

The FDA definitions recommendations for creating and testing medical tests and all calculations performed by the program are here: FDA.

The following are convenient simplifications of those definitions uses in this report.

"Screening test" or Test hereafter means the "antigen" type of quick result, at home, Covid 19 screening test**.** A typical example of a "test" would be the Abbott BinaxNow Rapid Antigen Test for SARS-CoV-2.

"Antigen test": Antigen tests look for fragments of proteins that make up the SARSCoV-2 virus to determine if the person has an active infection. Types Of Covid Test. A nice summary

"**Prevalence"**: A COVID-19 case is counted as active and part of the disease population prevalence during the 14 days after it is confirmed. Covid Tracker

"**Asymptomatic"**: This word is used to mean "asymptomatic and pre-symptomatic". These are not the same thing, but for prevalence and false positive modelling of antigen tests they can be combined.

Conclusion

I modeled a variety of what I believe to be realistic scenarios.

The reports of the scenarios modeled are listed below and are self-documenting.

One scenario using the variables most favorable to the BinaxNow test and a prevalence of 1% the model predicts 69% of positives are false. Running the same scenario at an unrealistic 5% prevalence rate we still get about 1 in 3 false positives. This seems so appalling I cannot convince myself that I am not missing something. Doubtless I am.

These and other runs of the model persuade me that:

(1) Covid antigen screening tests can be rendered worse than useless by adverse (low but realistic) prevalence. At high asymptomatic rates (using the CDC level of sensitivity for that condition) and a prevalence of 1% essentially all positives will be false. (96%)

(2) The model confirms that if I am right about the uncertainty of the values of the main test parameters and other variables then we have absolutely no idea whether the current Covid 19 mass screening program using theses "quick" screening tests is helpful or counterproductive.

(3) The results of the model at prevalences up to 5% are not consistent with some of the manufactures performance claims, particularly the following statement:

> *A positive test result means it is very likely you have COVID-19 and it is important to be under the care of your healthcare provider.* BinaxNow Insert
>
> According to the model exactly the reverse is true.

The Utility Of The Screening Tests Is Dubious

Screening tests are subject to many shortcomings.

Lack of knowledge about prevalence, asymptomatic cases, pre symptomatic cases, constantly evolving Covid variants, skills of the tester, testee disease exposure, testee vaccination status, viral load, sensitivity, specificity, similarity of symptoms to many other diseases and many other variables affecting accuracy, seriously undermine the credibility of covid screening tests, especially when used for diagnosis or isolation.   ASM FDA.

The screening tests are NOT adequate for diagnosis.

As the antigen testing algorithms indicate, confirmatory testing may be needed regardless of the symptom or exposure status of the person being tested. Confirmatory testing should take place as soon as possible after the antigen test. [using a serological test] CDC

The claimed test accuracy assumes *multiple* applications of the test.

These tests are widely supposed to be effective in a single "*quick*" test. In fact the manufacturer's documentation states that it requires two tests separated by 36 hours, which is not consistent with the claim that these are "rapid" result. In any case I suggest that many, if not most, self-testers will alter their behavior based on the first test and not repeat the tests.

> *The BinaxNOW™ COVID-19 Antigen Self-Test is a lateral flow immunoassay intended for the qualitative detection of nucleocapsid protein antigen from SARS-CoV-2 from individuals with or without symptoms or other epidemiological reasons to suspect COVID-19 infection when tested twice over three days with at least 36 hours between tests*. Abbot

The prevalence of Covid is unknown.

There is an absence of good estimates of the prevalence of Covid 19.

> *At the time of this writing there are few published, population-representative COVID-19 prevalence studies.* PNAS

This lacuna and prevalence variability across time and space, alone put a large question mark over the results of the tests since the efficacy of the tests depend upon population prevalence.

> *… reliable prevalence estimates are limited. Prevalence, which affects predictive value estimates, can be considered unknown, and varies over time.* Nature

Prevalence of Covid 19 appears to be typically less than 1%

Where measurement has been attempted the whole population prevalence of Covid 19 seems to be typically less than 1%. This is a very low prevalence and if accurate will play havoc with screening tests. Covid Tracker

Prevalence varies tremendously (a problem in itself) but such measurements as we have seem to put a typical range of prevalence between 0% and 3%. Prevalence is usually **less than 1%.** One brief peak of 3% is alleged but this occurred at a time of record low deaths, so make of that what you will.

- 1.25% to 3.09%. Highest prevalence as of Nov 2020 in UK: Imperial College London
- 0.05 of 1 percent. Case Prevalence in USA in June 29 2020. (1/1978) Covid Tracker
- 0.09 of 1 percent (1/107) Case Prevalence in USA in June 2020 peak. Covid Tracker
- 0.3 of 1 percent. The "Utah" Study. Prevalence very low: Utah Study
  Our own randomized viral testing was conducted in Utah between May 4th and July 1st, 2020 and estimates that the prevalence of COVID-19 in Utah was 0.27%. [0.027]. At the same time, our method predicts a median viral prevalence of 0.3% [0.003]

Manufacturer's claims of accuracy have not been independently verified.

The tests are released under "emergency" FDA licenses and few of the manufacturer's claims of accuracy have been independently verified by the FDA or CDC.

> *Limited data have been published for these home tests given that they are available through EUAs* [Emergency Use Authorization] *that do not require clinical trials.* NCBI Covid Tests Untested PBS

The claimed sensitivity and specificity are not "real world".

Sensitivity and specificity estimates shown may not be indicative of the real world performance of the tests. The number of samples in the panel is a minimally viable sample size that still provides reasonable estimates and confidence intervals for test performance, and the samples used may not be representative of the antibody profile observed in patient populations. FDA

One of the few evaluations by the CDC of an emergency approved test discovered that a test claiming to have a sensitivity of 99% proved in fact to have a sensitivity of 35.8% in uninfected and asymptomatic groups. CDC

WHO expects typical sensitivity to be 34% to 80%

Based on experience with influenza *"the sensitivity of these* [Covid 19 screening] *tests might be expected to vary from 34% to 80%"* WHO. Clearly this is very different from the manufactures typical claims for sensitivity in the high 90's% which run counter to long experience and thus are suspect.

Asymptomatic and presymptomatic unknowns have a serious effect on test accuracy.

Rapid antigen tests have received Food and Drug Administration (FDA) Emergency Use Authorization (EUA) for use in symptomatic persons, but ***data are lacking on test performance in asymptomatic persons to inform expanded screening testing*** *to rapidly identify and isolate infected persons.* CDC CDC2

> *Among asymptomatic participants, antigen test sensitivity was 41.2%, specificity was 98.4%, and PPV in this population was 33.3%. This low PPV was observed despite a relatively high prevalence of SARS-CoV-2 in this population (5.2% prevalence overall; 2.0% among asymptomatic persons), suggesting that PPV could be even lower when using this antigen test among populations with lower expected SARS-CoV-2 prevalence.* CDC

***The rate of asymptomatic or presymptomatic covid cases is unknown*.**

The rates of asymptomatic or presymptomatic cases in the general population are uncertain. The few studies report such a wide range of percentages it would be more honest to say we have no idea what the general prevalence is.

Even the lowest estimate undermine the claimed reliability of the tests.

17% JAMMI
35% : PNAS
25% : Nature
50%: PNAS
56% IJBS
91% Lancet
98% Shanghai

Interpreting the tests is subjective & requires expert intrepretation

Interpreting the results of an antigen test for SARS-CoV-2 depends primarily on the clinical and epidemiological context of the person who has been tested (e.g., symptoms, close contact to others with COVID-19, setting in which they live, likelihood of alternative diagnoses, or disease prevalence in their geographic location). CDC Guidance

In other words the result of the test in meaningless without the *"Antigen Test Algorithm"* requiring judgement of a skilled analyst and full and honest information about contact of the testee with the virus and vaccination status. This of course is impossible to perform properly with scientific consistency and accuracy in the sub-clinical settings for which the rapid tests are vaunted for use, such as at home, or mass screening in schools. Thus the CDC is allowing that these tests are not suited for the main purpose for which the FDA is "*flooding*" the country them.

> *I actually have been saying that for months and months and months – that we should be literally* [stet] *flooding the system with easily accessible, cheap, not needing a prescription, point of care, highly sensitive and highly specific, . … you are going to be seeing more of that soon,"* Fauci

See also *" there is an element of subjectivity in scoring the results "* below.

***Some tests are less effective or useless against new variants of Covid 19*.**

- *Genetic variants of SARS-CoV-2 arise regularly, and false negative test results can occur.*
- *Test performance may be impacted by certain variants.*
- *Tests with single targets are more susceptible to changes in performance due to viral mutations, meaning they are more likely to fail to detect new variants.* FDA2

Bias In estimating the test's efficacy "often" cannot be ruled out.

This point is not specific to the Covid 19 tests but it highlights one of the many difficulties of screening tests in general and reminds us to be a little sceptical, in particular with tests produced and released without usual evaluation in an "emergency".

> *Sensitivity and specificity estimates (and other estimates of diagnostic performance) can be subject to bias. Biased estimates are systematically too high or too low. Biased sensitivity and specificity estimates will not equal the true sensitivity and specificity, on average.* ***Often the existence, size (magnitude), and direction of the bias cannot be determined.*** FDA

“Risk illiteracy” and lack of evidence based medicine

The linked article reveals widespread misconceptions about screening tests in the medical profession and general public. The article also underlines the dangers of misunderstanding the "accuracy" of screening tests. *WP* The article provides links to some worrying studies on medical testing in general.

In short, the fact that these tests are approved by the CDC and recommended by doctors does *not* mean they are useful. In a time of panic and hysteria it would be easy to promulgate useless or even counterproductive measures. JAMA NCBI NCBI2 Overdiagnosis Similar article. Guardian

> *“We are unsure whether combined screenings, repeated symptom assessment, or rapid laboratory tests are useful*". Cochrane Evidence Based Medicine

Models of mass screening benefits have ignored common sense.

> *Studies in which models of test parameters have been used to evaluate effects on transmission of SARS-CoV-2* [by mass screening] *have paid little or no attention to the specificity of tests or the practical impacts of imperfect specificity. According to these models, both turnaround time and test frequency are more important than test sensitivity for preventing transmission (6, 7). These models rely on theoretical test performance parameters and assume ideal test utilization and human behavior…. In reality, if specificity is even modestly compromised, it will strike at the core of an important parameter for real-world impact of testing: the reliability of the results.* ASM

So there is an attitude that any test will do so long as it's quick. Well I can see the financial and political benefits of such thinking. It would certainly make a lot of money and create a great impression of the public health system "doing something".

CDC infers that screening tests are not determinative.

> *Confirmatory testing with an FDA-authorized nucleic acid amplification test (NAAT), such as RT-PCR, should be considered after negative antigen test results in symptomatic persons, and after positive antigen test results in asymptomatic persons*. CDC

So if we have to subjectively decide if a testee is asymptomatic or not and then depending on the screening result use a laboratory test to get a meaningful result, are these tests really helping ?

The Abbott BinaxNOW Rapid Antigen Test for SARS-CoV-2.

As a test case we will use the screening test program to model a typical over the counter "Covid Instant Test" that might be used by (say) students at a university or at home: *the Abbott BinaxNOW Rapid Antigen Test for SARS-CoV-2.* BinaxNow The US Federal government is spending billions of dollars on this particular test for mass screening.

Manufacture's claimed "accuracy" is implausible

BinaxNow Claims.

On its "*BinaxNOW Performance*" web page the manufacturer vaunts a sensitivity of 93.3% on high viral load cases. Abbot BinaxNOW Performance. This a selective use of data, and an improper context for the statistic and highly misleading. It also seems to be just plain wrong. (Below)

Shortcomings not mentioned.

Except for the statement " *The BinaxNOW COVID-19 Ag Card has not been FDA cleared or approved"* the web page does not mention any of the qualifications and shortcomings listed below. (For example the CDC found the real world sensitivity could be as low as 35.8%).

BinaxNow literature cites misleading and inappropriate statistics.

Abbot's BinaxNOW COVID-19 antigen self-test has an accuracy rate of 84.6% for detecting Covid-19 infections, and 98.5% for correctly identifying Covid-19 negatives. BinaxNow Sales Documentation.

"*Accuracy rate*" is not a term mentioned by the CDC definitions (above) and in a document targeting consumers is the phrase could not be better composed to mislead. Other documents show that by "accuracy rate' Abbot is referring to the sensitivity and specificity. Using sensitivity and specificity in this context is wrong. The accuracy for screen testing individuals *should* be cited using the PPV and NPV.

> *Of particular importance, although it is desirable to have tests with high sensitivity and specificity, the values for those two metrics should not be relied on when making decisions about individual people in screening situations. In that second context, use of PPVs and NPVs is more appropriate.* FIPH

Why would BinaxNow do this ? Could it be because the sensitivity and specificity are (or can easily be made to be) impressively high numbers like 99.9 % which creates a false sense of confidence in the test compared with a PPV of (say) 60% ? Misleading statistics like these are readily picked up and passed on by an uncritical media. Time

CDC says manufactures BinaxNow accuracy is greatly exaggerated.

BinaxNow is one of the few "emergency" approved tests that have been tested by the government. CDC Evaluation of Abbott BinaxNOW Rapid Antigen Test for SARS-CoV-2. CDC

.. the BinaxNOW antigen test had a sensitivity of 64.2% for specimens from symptomatic persons and ***35.8%*** *for specimens from asymptomatic persons.*

Contrast this to the manufacturer's flat claim of sensitivity of 98.5%.

Poor studies ?

The quality of studies cited by the manufacturer (Abbot) as underlying their claims, seem less than ideal. A quick look at just one of them reveals some rather glaring shortcomings. (MedxRiv) :

- The study was deliberately conducted where prevalence was abnormally high. Why ?

- The study excluded people under 10 years of age.
  When tested on children under two the test sensitivity was a ridiculous 7.6%. JID.
  This and the extremely small percentage of Covid deaths in children make it reasonable to assume younger people are less affected by covid 19 and may be presumed to be more asymptomatic. BinaxNow is "indicated" for anyone over 2. Abbot So this leaves a group of disproportionately asymptomatic people ages 2 through 9 out of the sample, leading to an exaggeration of sensitivity.

- The study used self-selecting subjects in a public place where - absent proof to the contrary - it may be assumed subjects who sought testing were more likely to believe they were infected. (That is more symptomatic or having a high expectation of having been exposed). This is like calibrating a screening test for Aids in an needle exchange clinic. The results would hardly be representative.

- " *there is an element of subjectivity in scoring the results* ".
  The result stripe on the test paper is often ambiguous and subjective in interpretation. So much so that the test requires "*supplementary technician training* ". This is not consistent with the vaunted "in home" use for which the US Government is using this test in staggering numbers. The complexity of interpreting the test is underscored by the complicated decision diagram for the test. BinaxNow-training. Studies confirm lower accuracy when not administered by a professional.

  > *Our results, however, indicate that home use of these rapid COVID-19 diagnostics as self/caregiver-operated tests may decrease this already lowered sensitivity even further.*
  > Nature

- The sensitivity claimed by BinaxNow depended on a disputed variable "*the range* [viral load] *thought to be the most transmissible*"**.** A "*conservative*" estimate of this variable estimate reduced the (claimed) specificity by 5% to 93.8% even in an unrepresentative high prevalence level. Why was a non-conservative value used in the BinaxNow conclusions ? (MedxRiv)

- Initially the study found that, *using the manufacturer's proposed criteria, 9/14 Binax-CoV2 (+) tests (64%) in this population were likely false positives….. Clearly, these initial criteria were problematic. Therefore, on subsequent test days, we evaluated additional criteria for classifying a band as positive, in consultation with experts from the manufacturer's research staff.*

  Prima facie this seems to be special pleading, using highly subjective criteria to alter the performance of the test and a lack of independence between evaluator and manufacturer. A CDC review found the maximum sensitivity of 64% in symptomatic persons. CDC

A scenario for the simulation on our computer model:

Unfortunately nobody gets paid for saying "*I don't know*"

Here's a fact of life: nobody gets paid for saying "*I don't know*".
This is a problem for science based medicine.
In the case of Covid 19 "*I don't know*" and it's even less popular partner "*It's not possible to know*" *should be* the most frequent answer to many of the most vital questions concerning screening tests.

The efficacy of a screening test depends on variables that are hard to quantify and haven't be quantified. Prevalence, asymptomatic cases, pre symptomatic cases, constantly evolving Covid variants, skills of tester, testee disease exposure, viral load, sensitivity, specificity, similarity of symptoms to many other diseases are all variables whose value is highly debatable and variable.

Small changes in any of these variables can make a huge difference to the practical usefulness of a screening test.

These fundamental uncertainties are confounded further by the sometimes counterintuitive interaction between the uncertain variables. For example a small decrease in specificity (a measure of false negatives) can (surprisingly) create a large increase in false positives. So having the exact specificity is extremely important but - like almost all the other variables - is highly uncertain.

In the face of this complexity and lack of information, any definitive statement about performance and role of Covid screening tests (and therefore of Covid screening) is not justified by evidence based medicine.

So we will use the screening test model to simulate a variety of what I believe to be realistic scenarios. If nothing else this should demonstrate how these indeterminate variables create uncertainty about the BinaxNow and other Covid screening tests.

Focus on false positives

To limit the scope of the project to a practical size we will focus our simulations on the relationship between prevalence and false positives.

The Scenario. Testing students at a school

Imagine all the students at a university being screened using the BinaxNow kit on the first day of term.

Parameters we will use.

There are an infinite combinations of the screening test parameters, so I will pick a few scenarios to model with (in my opinion) realistic values. I justify the choice of values of the variables below, but of course nobody knows what the values are. The model is flexible and the user can play with other scenarios.

Asymptomatic Cases 60%

Asymptomatic cases in the general population is anywhere between 25% and 91%. (See above). Given the arbitrary and unknown value for this important parameter I felt justified in placing the value at a fairly high rate. Young people are more likely to experience low and absent symptoms. Also students who have symptoms or feel unwell will tend to not attend and thus not be tested.

Although higher rates of asymptomatic testees decreases accuracy of the tests the model has no code to account for this directly. Instead the model uses the lower or higher sensitivity rates for asymptomatic and symptomatic testees provided by the CDC to account for asymptomatic rates.

Sensitivity:

The sensitivity the main prevalence comparison will be 64.2%.

I feel justified using 64.2%. The 64% sensitivity for symptomatic persons is the highest sensitivity established by (putatively) independent CDC testing for 100% symptomatic testee samples. This of course is an absurd presumption in a screening scenario but we will favor the BinaxNow test by using it.

The WHO expects, based on experience with influenza, *"the sensitivity of these* [Covid 19 screening] *tests might be expected to vary from 34% to 80%".* Notice that the CDC value is almost in the center of the WHO interval and is consistent with long historic experience. The much higher manufacturer's numbers are well outside the WHO interval. WHO.

Bear in mind that even this appallingly low level of specificity (64.2%.) assumes high prevalence, and unrepresentatively skilled testers. Neither of these things obtain (probably) in either real world screening scenarios or those assumed for this evaluation.

We'll also test some combination of 35.8%. 50% and 64.2% per CDC and 93.3% per BinaxNow as time permits.

Specificity: 90% to 98.5%.

The manufacturer claims a specificity of *98.5* %. I have not had time to research the specificity. However the false claims about sensitivity have made me very dubious about all claims for this test. Like those dictatorships where 99.9% are supposed to have voted for the glorious leader. One becomes skeptical. Although I will model the manufactures claim, I feel more than justified in trying modestly lower values for specificity.

Prevalence range:

The comparative prevalence will be 1%.
See " *Prevalence of Covid 19 is typically less than 1%*" above.

But we will vary the prevalence from 1 % to 5%

Conclusion

I modeled a variety of what I believe to be realistic scenarios.

The reports of the scenarios modeled are listed below and are self-documenting.

One scenario using the variables most favorable to the BinaxNow test and a prevalence of 1% the model predicts 69% of positives are false. Running the same scenario at an unrealistic 5% prevalence rate we still get about 1 in 3 false positives. This seems so appalling I cannot convince myself that I am not missing something. Doubtless I am.

These and other runs of the model persuade me that:

(1) Covid antigen screening tests can be rendered worse than useless by adverse (low but realistic) prevalence. At high asymptomatic rates (using the CDC level of sensitivity for that condition) and a prevalence of 1% essentially all positives will be false. (96%)

(2) The model confirms that if I am right about the uncertainty of the values of the main test parameters and other variables then we have absolutely no idea whether the current Covid 19 mass screening program using BinaxNow is helpful or counterproductive.

(3) The results of the model at prevalences up to 5% are not consistent with some of the manufactures performance claims, particularly the following statement:

> *A positive test result means it is very likely you have COVID-19 and it is important to be under the care of your healthcare provider.* BinaxNow Insert
>
> According to the model exactly the reverse is true at lower prevalences.

High Asymptomatic rates. CDC Sensitivity. Slightly lower Specification.

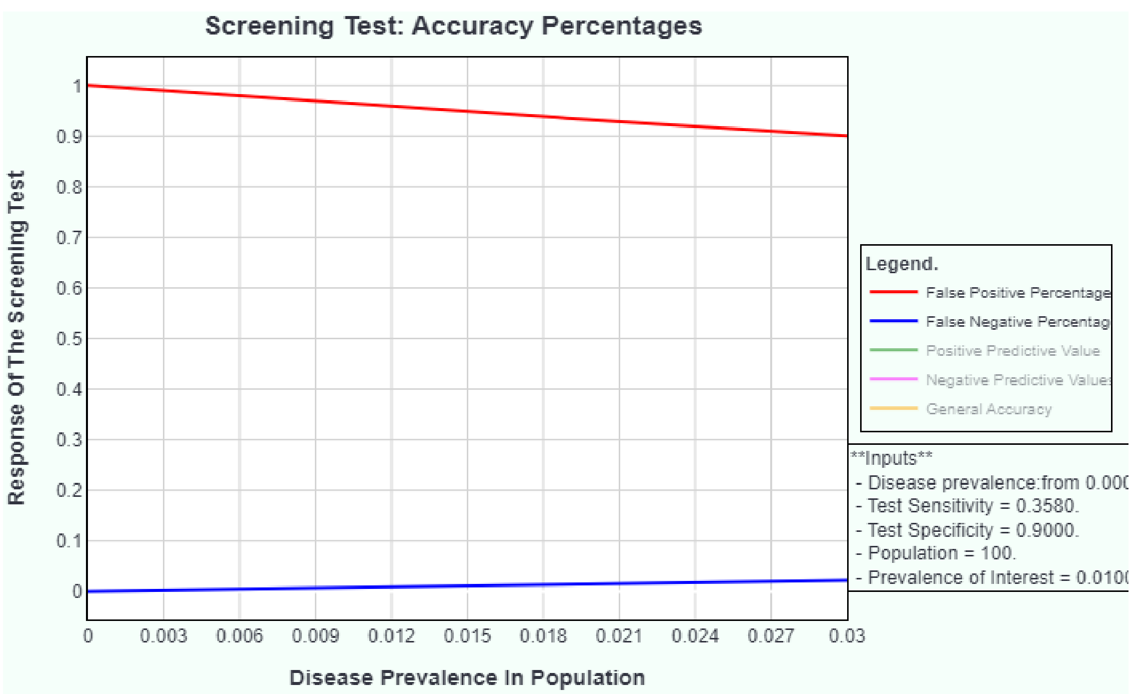
REPORT ON YOUR SCREENING TEST

1% prevalence.
CDC Sensitivity for High Asymptomatic rates
Slightly lower than manufacturers Specification.

** Inputs specifying the simulation. **
-- In this simulation the disease prevalence varies from 0.00000 to 0.03000.
-- Test Sensitivity = 0.3580.
-- Test Specificity = 0.9000. Binax says 98%.
-- Plot Prevalence Start = 0.00000.
-- Plot Prevalence End = 0.03000.
-- Population = 100.

** The range of false results. **
-- The false positive rate varies from 1.00000 to 0.90032.
-- The false negative rate varies from 0.00000 to 0.02159.

** At The Prevalence Of Interest = 0.010000. **
-- About 0.96 of all positives are false.
-- About 0.01 of all negatives are false.
-- Positive Predictive Value (PPV) = 0.0356.
-- Negative Predictive Value (NPV) = 0.9927.
-- Claimed True Positives = 0.37.
-- Claimed False Positives = 9.90.
-- Claimed True Negatives = 89.08.
-- Claimed False Negatives = 0.65.



High Asymptomatic rates. CDC Sensitivity. BinaxNow High Specification

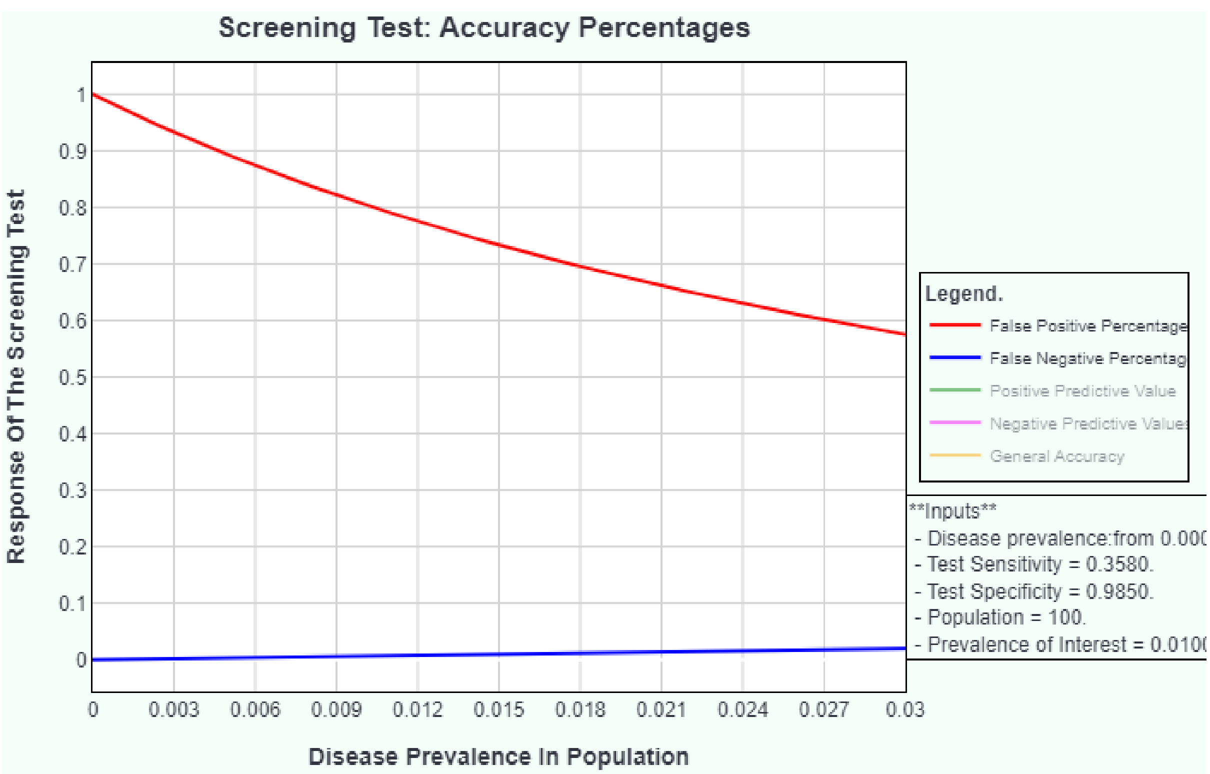
REPORT ON YOUR SCREENING TEST

1% prevalence.
CDC Sensitivity for High Asymptomatic rates.
BinaxNow highest Specification.

** Inputs specifying the simulation. **
-- In this simulation the disease prevalence varies from 0.00000 to 0.03000.
-- Test Sensitivity = 0.3580.
-- Test Specificity = 0.9850.
-- Plot Prevalence Start = 0.00000.
-- Plot Prevalence End = 0.03000.
-- Population = 100.

** The range of false results. **
-- The false positive rate varies from 1.00000 to 0.57533.
-- The false negative rate varies from 0.00000 to 0.01976.

** At The Prevalence Of Interest = 0.010000. **
-- About 0.80 of all positives are false.
-- About 0.01 of all negatives are false.
-- Positive Predictive Value (PPV) = 0.1974.
-- Negative Predictive Value (NPV) = 0.9933.
-- Claimed True Positives = 0.37.
-- Claimed False Positives = 1.48.
-- Claimed True Negatives = 97.50.
-- Claimed False Negatives = 0.65.



Best Case For BinaxNow - 1% prevalence

REPORT ON YOUR SCREENING TEST

Best Case For Binax*Now* - 1% prevalence.
Sensitivity. CDC Symptomatic level.
Specificity highest as claimed by manufacturer.

** Inputs specifying the simulation. **
-- In this simulation the disease prevalence varies from 0.00000 to 0.06000.
-- Test Sensitivity = 0.6420.
-- Test Specificity = 0.9850.
-- Plot Prevalence Start = 0.00000.
-- Plot Prevalence End = 0.06000.
-- Population = 100.

** The range of false results. **
-- The false positive rate varies from 1.00000 to 0.26796.
-- The false negative rate varies from 0.00000 to 0.02267.

** At The Prevalence Of Interest = 0.010000. **
-- About 0.69 of all positives are false.
-- About 0.00 of all negatives are false.
-- Positive Predictive Value (PPV) = 0.3061.
-- Negative Predictive Value (NPV) = 0.9963.
-- Claimed True Positives = 0.65.
-- Claimed False Positives = 1.48.
-- Claimed True Negatives = 97.50.
-- Claimed False Negatives = 0.37.

Best Case For BinaxNow - 2% prevalence

REPORT ON YOUR SCREENING TEST

Best Case For BinaxNow - 2% prevalence.
Sensitivity. CDC Symptomatic level.
Specificity highest as claimed by manufacturer.

** Inputs specifying the simulation. **
-- In this simulation the disease prevalence varies from 0.00000 to 0.06000.
-- Test Sensitivity = 0.6420.
-- Test Specificity = 0.9850.
-- Plot Prevalence Start = 0.00000.
-- Plot Prevalence End = 0.06000.
-- Population = 100.

** The range of false results. **
-- The false positive rate varies from 1.00000 to 0.26796.
-- The false negative rate varies from 0.00000 to 0.02267.

** At The Prevalence Of Interest = 0.020000. **
-- About 0.53 of all positives are false.
-- About 0.01 of all negatives are false.
-- Positive Predictive Value (PPV) = 0.4713.
-- Negative Predictive Value (NPV) = 0.9925.
-- Claimed True Positives = 1.31.
-- Claimed False Positives = 1.47.
-- Claimed True Negatives = 96.49.
-- Claimed False Negatives = 0.73.

<u>Best Case For BinaxNow - 3% prevalence</u>

REPORT ON YOUR SCREENING TEST

Best Case For Binax***Now -*** 3% prevalence.
Sensitivity. CDC Symptomatic level.
Specificity highest as claimed by manufacturer.

** Inputs specifying the simulation. **
-- In this simulation the disease prevalence varies from 0.00000 to 0.06000.
-- Test Sensitivity = 0.6420.
-- Test Specificity = 0.9850.
-- Plot Prevalence Start = 0.00000.
-- Plot Prevalence End = 0.06000.
-- Population = 100.

** The range of false results. **
-- The false positive rate varies from 1.00000 to 0.26796.
-- The false negative rate varies from 0.00000 to 0.02267.

** At The Prevalence Of Interest = 0.030000. **
-- About 0.43 of all positives are false.
-- About 0.01 of all negatives are false.
-- Positive Predictive Value (PPV) = 0.5747.
-- Negative Predictive Value (NPV) = 0.9887.
-- Claimed True Positives = 1.96.
-- Claimed False Positives = 1.45.
-- Claimed True Negatives = 95.49.
-- Claimed False Negatives = 1.10.

Best Case For BinaxNow - 5% prevalence

REPORT ON YOUR SCREENING TEST

Best Case For BinaxNow - 5% prevalence.
Sensitivity highest. Symptomatic level.
Specificity highest as claimed by manufacturer.

** Inputs specifying the simulation. **
-- In this simulation the disease prevalence varies from 0.00000 to 0.06000.
-- Test Sensitivity = 0.6420.
-- Test Specificity = 0.9850.
-- Plot Prevalence Start = 0.00000.
-- Plot Prevalence End = 0.06000.
-- Population = 100.

** The range of false results. **
-- The false positive rate varies from 1.00000 to 0.26796.
-- The false negative rate varies from 0.00000 to 0.02267.

** At The Prevalence Of Interest = 0.050000. **
-- About 0.31 of all positives are false.
-- About 0.02 of all negatives are false.
-- Positive Predictive Value (PPV) = 0.6943.
-- Negative Predictive Value (NPV) = 0.9811.
-- Claimed True Positives = 3.24.
-- Claimed False Positives = 1.42.
-- Claimed True Negatives = 93.54.
-- Claimed False Negatives = 1.80.

Final Project Proposal

Course

CCSU Stat 476. Spring 2022.

Student

Tim Brockway. Student ID: 30259316 Email: BrockwayTim@My.CCSU.edu

Professor

Roger L. Bilisoly. bilisolyr@ccsu.edu https://www2.ccsu.edu/faculty/bilisolyr

Basic Idea

It seems that medical screening tests vaunting very high "general accuracy" can give staggering levels of false results when the prevalence of a disease is low. I'd like to understand this by writing a program that explores the effect the prevalence of an infection in a population on the usefulness of screening tests.

Python Exploration

I'd like to use the project to explore Python's capabilities and weaknesses.
In particular I wish to learn about.

- Program modularization using functions driven by a main line.
- Handling global variables.
  - Creating global variables using the 'global' statement is deprecated.
  - These statements are in any case counterintuitive and complex to use.
  - I want to try python classes to create global, self-initializing variables to allow program modularization.
- Panda based data handling. Pandas seem to be the standard data table for python.
- Some sort of GUI to permit the user to enter the medical tests parameters.
- A somewhat interactive plot to allow the user to choose which variables are displayed.
- Use Github to get a feel for how most folks are storing and publishing their work.
  - Storing the code, project documentation and all project files in one place will be useful in any case.
- Some other Python features I've played around with this and would like to work it into the project.
    - Nested Python dictionaries and pretty printing Python dictionaries.
    - Enumeration of non iterable objects (Enum).
    - Etc.

Inputs

- A GUI of some kind, will invite the user to enter the following medical test statistics.
  - Inputs will have to be validated by the code.

  - Population.
  - Test sensitivity.
  - Test specificity.
  - Start of population disease prevalence range of interest.
  - End of population disease prevalence range of interest.
  - Prevalence of particular interest.
  - A way of indicating which graphs are required. All at once would be a bit cluttered.
  - Some gratuitous file handing to explore Pythons i/o methods. (E.g., Download the pandas dataframe generated).

Outputs

- A report on the statistics resulting from the users test specification.

- An interactive graph with:
  x axis. The required range of disease prevalences.
  y axis: Any or all of the following:
    - False positive percentage.
    - False negative percentage.
    - Positive predictive value.
    - Negative predictive value.
    - False Positives (NPV)
    - False Negatives (NPV)
    - The general accuracy (The suspect statistic).
    - Prevalence of interest (A vertical line).

- Error messages indicating errors in user's input.

- Code to allow.
  - Download of a CSV file with the plot data. (Optional).
  - Download the plot itself. (Optional).
  - Some hyperlinks to (say) documentation and code at github.
  - Anything else that the usual input/calculate/plot statistics program might use.

Statistics

I will have to explore the epidemiological issues and terminology to be sure I generate the statistics correctly. I do not know if the various python libraries have the statistical functions needed to do screening test calculations per se. But I want to do the calculations myself in raw python. Prima facia once the variables' purpose and derivation are understood calculating them is easy. I should disclose that at the time of writing I am not fully on top of the epidemiology but there is a ton of help online.