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

To Do

```
# TO DO
# - Create documentation including a documentation vidio and deploy it
#
# For final submission.
# - Do hand calculations for verification.
# -spell check comments to get rid of gross errors.
# - check results carefully against another calculator.
# - create an actual test case of a specific covid screening test.
```

How To View This Document.

You may want to DOWNLOAD THIS PDF file to read it in a proper full PDF viewer.
This document has a convenient index, searching, active links and other pdf features which will only be available in a full featured pdf viewer.

Contacts

Course: CCSU Stat 476. Spring 2021.
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 Purpose

This program evaluates the effect of disease prevalence on the accuracy of medical screening tests.

How To Run This Program

This program is run in a web page. Enter the link below in any major web browser.
The program will then instruct you on how to proceed.
   https://share.streamlit.io/profbrockway/screeningtest/main/screeningtest.py

Project Files: Location

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

Program Verfication

- The outputs of this program have been verified by:
   (a) Hand calculations of critical numbers.
   (b) Comparison with the output of another Screening test calculator.
         https://epidemiology.sruc.ac.uk/shiny/apps/predictive_values/

Conclusion:

- The program successfully models the theoretical accuracies of a screening test.

- After reading around extensively in my humble opinion:
    - Medical screening tests should be treated with a high degree of scepticism since:
        - The manufacturers claims of accuracy are often not verifed by official sources.
        - The test population prevalence is often unknown and may be rendering the tests ludicrously inaccurate.

Program Input / Outputs

Inputs

- A GUI in the form of a web page:
    - The user will enter the following medical test statistics.
        - Population.
        - Test sensitivity.
        - Test specificity.
        - Start of population disease prevalence range of interest.
        - End of population disease prevalence range of interest.
        - Disease "Prevalence" of particular interest. AKA "Prior Probability Of Infection".
        - A text annotation to label the plots.

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 "general" accuracy (The suspect statistic).
        The program permits the user to show or hide the variables to avoid clutter.

- A display of a data grid showing the data generated from the users input and used for the plots.
    - The most interesting entries will be highlighted. The grid will be 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:
    - The data grid as a CSV file with the plot data.
    - The data grid as am Excel file with the plot data.
    - The plot as an image.
    - The plot as an *interactive* html web page.

Screening Tests And The "Low Prevalence" Problem

Overview:

This program explores the effect the prevalence of an infection in a population on the usefulness of medical screening tests.

Medical screening tests vaunting a very high "general accuracy" can nevertheless give very high levels of false results when the prevalence of a disease is low.

This phenomenon, definition of terms, explanations, calculations and premises for this program are given in the following FDA document describing the way to test screening tests: FDA Guidance.

Observations:

Disease prevalence if very difficult to determine and often is not known.
    In the case of covid 19 disease prevalence is very low
        https://open.fda.gov/apis/device/covid19serology/

      Important caveats
      Sensitivity and specificity estimates shown may not be indicative of the real-world performance of the tests.

In the case of covid 19
    Screening tests are being made largely without any independent verification of their accuracy.

    https://open.fda.gov/apis/device/covid19serology/
    Important caveats
    Sensitivity and specificity estimates shown may not be indicative of the real-world performance of the tests.

definition of prevalence.
    statistically "lacking base data: problem.

This phenomenon and the calculation.

FDA description of how to calculate the tests stats
https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting-results-studies-evaluating-diagnostic-tests-guidance-industry-and-fda

- Sreening Tests For Covid are "Rapid antigen tests".
Health care providers typically rely on molecular tests, particularly when people have COVID-19 symptoms, whereas antigen testing is often used when quick results are needed or for general screening and surveillance.

https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antigen-tests-guidelines.html#:~:text=The%20U.S.%20Food%20and%20Drug,which%20indicates%20current%20viral%20infection.
- antigen tests perform best when subject is positive.

FDS BIAS in sens and spec
https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting-results-studies-evaluating-diagnostic-tests-guidance-industry-and-fda
Simply increasing the overall number of subjects in the study will do nothing to reduce bias.

approved at home tests
https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/home-otc-covid-19-diagnostic-tests?utm_medium=email&utm_source=govdelivery#list

https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/covid-19-test-uses-faqs-testing-sars-cov-2

Screening for COVID-19 is looking for occurrence at the individual level even if there is no individual reason to suspect infection such as a known exposure. This includes broad screening of asymptomatic individuals without known exposure with the intent of making individual decisions based on the test results. Screening tests are intended to identify infected individuals prior to development of symptoms or those infected individuals without signs or symptoms who may be contagious, so that measures can be taken to prevent those individuals from infecting others. FDA regulates screening tests as in vitro diagnostic devices and has provided recommendations and information regarding EUA requests for COVID-19 screening tests in the EUA templates referenced in the Policy for Coronavirus Disease-2019 Tests. Examples of screening include testing plans developed by a workplace to test all employees returning to the workplace regardless of exposure or signs and symptoms and testing plans developed by a school to test all students and faculty returning to the school regardless of exposure or signs and symptoms, with the intent of using those results to determine who may return or what protective measures to take on an individual basis.

Should SARS-CoV-2 antibody test results be used to assess whether or not a person is protected from COVID-19? (02/24/22)
A: No, antibody testing should not be used to assess immunity to COVID-19. More research is needed to understand what antibody test results can tell us, both in people who have been infected with SARS-CoV-2 and in people who have received a COVID-19 vaccination. While a positive antibody test result can be used to identify SARS-CoV-2 antibodies in a person's body, it should not be used to evaluate a person's level of immunity or protection from COVID-19 at this time.

cdc guidence on screening tests for covid
https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antibody-tests-guidelines.html
Antibody testing is not currently recommended to assess for immunity to SARS-CoV-2 following COVID-19 vaccination, to assess the need for vaccination in an unvaccinated person, or to determine the need to quarantine after a close contact with someone who has COVID-19.

cdc about screening tests ALSO LIST OF TESTS WITH THEIR SENS AND SPEC
   they state the ppv and npv at prev = 5% unrealisticly high.
    https://open.fda.gov/apis/device/covid19serology/
      Important caveats
      Sensitivity and specificity estimates shown may not be indicative of the real world performance of the tests.

POP PREV NOT KNOWN !!!
      https://www.pnas.org/doi/10.1073/pnas.2026412118

     In epidemiology, "prevalence" is the fraction of a population *currently* infected, and "incidence" is the fraction of susceptible people infected in a unit of time. Prevalence tells us the size of the infected group and, in some circumstances, gives us information about the size of the susceptible group. Incidence describes the rate of spread.

InteliSwab COVID-19 Rapid Test - Instructions for Use Healthcare Provider
approved by fda
https://www.fda.gov/media/149911/download
• The product has not been FDA cleared or approved; but has been authorized by FDA under EUA.

list of quick at home tests
https://www.nytimes.com/wirecutter/reviews/at-home-covid-test-kits/

BINAX rapid test THIS IS OUR TEST CASE
  It's a big selling rapid at home test. I'ts one of the few that the cdc has tested(Emergency fda licicenses have mostly not been evaluated by the government). It's claimed stats are massively more than the cdc found. CDC found spec ranging between .642 an .358. really low. Even this was with a very high prev= 0.087 !!!!
  Claimed: ??
evaluation:
     https://www.cdc.gov/mmwr/volumes/70/wr/mm7003e3.htm

BinaxNOW antigen test had a sensitivity 52.5% https://www.cdc.gov/mmwr/volumes/70/wr/mm7003e3.htm#T2_down

The results of the current evaluation differ from those of an evaluation of the BinaxNOW antigen test in a community screening setting in San Francisco (7), which found a BinaxNOW antigen test overall sensitivity of 89.0% among specimens from all 3,302 participants, regardless of the Ct value of the real-time RT-PCR–positive specimens.

The prevalence of having SARS-CoV-2 real-time RT-PCR positive test results in this population was moderate (8.7% overall; 4.7% for asymptomatic participants); administering the test in a lower prevalence setting will likely result in a lower PPV.

SCREENING TESTS NOT WORKING
https://www.sciencedirect.com/science/article/pii/S0966842X20302808

     Antibody tests should not be used for individual decision making at this time. This paper describes why this is the case: (i) antibody test results may be inaccurate without multiple sequential testing due to low disease prevalence and statistical limitations; (ii) many tests on the market are not of good quality, leading to inaccurate results; (iii) it is unknown whether past exposure to SARS-CoV-2 leads to durable immunity or if reinfection is possible; and (iv) it is unknown whether transmission is possible even if reinfection does not lead to clinical symptoms.

     if the disease is of low prevalence. This has to do with the predictive value, also referred to as the '==base rate fallacy,==' which takes into account both test performance and the background population prevalence of disease.

list of tests with their sens/spec
https://www.centerforhealthsecurity.org/covid-19TestingToolkit/serology/Serology-based-tests-for-COVID-19.html

cdc recommended quick tests
https://www.cdc.gov/coronavirus/2019-ncov/testing/self-testing.html

Nice summary
https://medical.mit.edu/faqs/faq-testing-covid-19
https://www.verywellhealth.com/coronavirus-antibody-test-uses-4844950

A LOW DISEASE PREVALENCE
https://bfi.uchicago.edu/wp-content/uploads/2020/07/BFI_WP_202054_Revised2.pdf
We use recent Covid-19 serology studies in the US, and show that the parameter confidence set is generally wide, and cannot support definite conclusions. Specifically, recent serology studies from California suggest a prevalence anywhere in the range 0%-2% (at the time of study), and are therefore inconclusive. However, this range could be narrowed down to 0.7%-1.5% if the actual false positive rate of the antibody test was indeed near its empirical estimate (~0.5%).

FDA Approving crap tests
https://www.centerforhealthsecurity.org/our-work/pubs_archive/pubs-pdfs/2020/200618-serosurvey-strategy.pdf
     The FDA, NIH, CDC, and NCI should release the results of their antibody test validation study. Validation of serological tests is critical to ensuring that the tests perform as they are intended, and a lack of validation has led to a patchwork of false positives and false negatives across the country, interfering with estimates of seroprevalence. Currently, tests need to be internally validated for EUA submission. Upon EUA submission, the manufacturer now must also submit the test for independent validation through institutes such as the NCI. Currently approved tests must also submit their kits for independent validation. Outside studies, typically in academic settings, have found discrepancies between the accuracy claimed by the manufacturer and their independent tests.

Very very low prevalances
http://www.nathanseegert.com/papers/Yang2020a.pdf
At the same time, our method predicts a median viral prevalence of 0.3%

==VERY LOW PREVALANECES> This could be my source==
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009374
As of December 31, 2020, we estimate nation-wide a prevalence of 1.4% [Credible Interval (CrI): 1.0%-1.9%] and a seroprevalence of 13.2% [CrI: 12.3%-14.2%], with state-level prevalence ranging from 0.2% [CrI: 0.1%-0.3%] in Hawaii to 2.8% [CrI: 1.8%-4.1%] in Tennessee, and seroprevalence from 1.5% [CrI: 1.2%-2.0%] in Vermont to 23% [CrI: 20%-28%] in New York. Cumulatively, reported cases correspond to only one third of actual infections. The use of this simple and easy-to-communicate approach to estimating COVID-19 prevalence and seroprevalence will improve the ability to make public health decisions that effectively respond to the ongoing COVID-19 pandemic.

```
# (But remember that what matters is not the prevalence of a disease in the population,
# but the proportion among tested people who actually have the disease.)


# ALARMIST STATEMENTS ABOUT TESTING
# https://finance.yahoo.com/news/omicron-testing-failure-134504476.html


##################################################################################
#                    ScreeningTest: TESTING
##################################################################################
#
# -  You can verify the results of this program at
#          https://epitools.ausvet.com.au/predictivevalues

#
# - This file has a built in driver so we can test without a separate driver
#    file.
#
##################################################################################
#      ScreeningTest: SPCICIFIC TEST CASE USING REAL DATA
##################################################################################
#
# THE EFFECT OF DISEASE PREVALENCE ON THE EFFICACY OF SCREENING TESTS
#   Medical screening Tests vaunted with very high 'Overall accuracy' can give
#   staggering levels of false results when the prevalence of a disease is low.
#   This program explores the effect the prevalence of an infection in a
#   population on the usefullness of screening tests. The goal is to
#   demonstrate that screening tests are complex and less reliable
#   than commonly supposed.
#
# NIH report on Spec and Sens of screening tests.
#    https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8220942/

#   Screening tests are 2 types
#      Lateral flow antigen" test or "rapid antigen" test
#
# THE "OVERALL ACCURACY OF A SCREENING TEST
#   The "Overall Accuracy"  is the  measure sometimes (IMHO) creates a
#    impression that a screening test is useful when
#   in fact it is dangerously misleading.
#    https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1492250/


# CASE PREVALENCE
#   Case prevalence measures the number of active COVID-19 cases in a
#   state as a percentage of the state's population.
#   A COVID-19 case is counted as active during the 14 days after it
#   is confirmed.
#   Case Prevalence in USA  is 0.002 of 1 percent in March 2021 (1/489)
#   Case Prevalence in USA  in June 2020 peak  was 0.009 of a percent  (1/107)
#     https://covid-tracker.mckinsey.com/prevalence

THE PREMISE OF MY PROJECT CONFIRMED BY THE FDA.
#  "At 0.1% prevalence, the PPV would only be 4%, meaning that 96 out
#   of 100 positive results would be false positives. the following link confirms the above tracker link.
#     https://www.fda.gov/medical-devices/letters-health-care-providers/potential-false-positive-results-
antigen-tests-rapid-detection-sars-cov-2-letter-clinical-laboratory


# Google
```

fda medical devices letters health care providers potential false positive results
#

PREVALENCE NOT KNOWN
https://academic.oup.com/ectj/article/25/1/1/6325165

   Prevalence of a novel infection like SARS-CoV-2 (the virus causing COVID-19 disease) **is a quintessential missing data problem.** Only a small subset of the population has been tested, this subset is almost certainly selective; we do not even know the accuracy of tests, and our understanding of the pandemic is vague enough so that we might not want to overly rely on heavily parameterized models.

PREVALENCE 1.9 FROM SERO STUDIES.
 https://www.sciencedirect.com/science/article/pii/S1047279722000369

Prevalence Studies Rare
Measuring Prevalence of the Coronavirus
   fFebruary 24, 2021 https://www.pnas.org/doi/10.1073/pnas.2026412118#sec-2
At the time of this writing there are few published, population-representative COVID-19 prevalence studies. In a recent review Franceschi et al. list 37 (5). Two in North America are the state of Indiana (6) and the state of Connecticut (7) in the United States. In addition to these, the state of Ohio Department of Health released results from a prevalence study conducted in that state during July 2020 (8, 9). Two important challenges affected many of these studies.

Highest prevalence yet (2021 1.25% to 3.09%
https://www.imperial.ac.uk/news/231715/react-study-records-highest-coronavirus-prevalence/