# Guided Workshop 3: Bayes’ Theorem and Diagnostic Testing

***Instructions***: Download the file “Guided Workshop 3 – STARTER.xlsx”. I would recommend setting aside about an hour for this activity. When you are ready to start the workshop, open and begin the video “Guided Workshop 3: Bayes’ Theorem and Diagnostic Testing”.

The video will have optional in-video questions to help teach you and guide you along. You won’t submit this document, but it will be a good template/guide for the activity.

At the end, after you have completed the Excel file above, you will open the “Guided Workshop 3 Submission” quiz, where you will enter the answers to the questions at the end of this document. In short, you will change a few of the parameters in your worked Excel file and place them in the quiz. If correct, you will earn course credit for this assignment.

***Background/Terminology*** (this is also explained in the “Guided Workshop 3: Bayes’ Theorem and Diagnostic Testing” video)

In this workshop, you will analyze how to predict the rate of false positives, false negatives, the positive predictive value (PPV) and the negative predictive value (NPV) for diagnostic tests like the COVID rapid screening tests. Note that this does NOT apply to extremely accurate nucleic acid amplification tests (NAATs) like RT-PCR tests.

First, we need to define the following events:

***Inf*** = infected w/ disease

***Not Inf*** = not infected w/ disease 🡪 this is the same as ***Inf’***

***Pos*** = tested positive for disease

***Neg*** = tested negative for disease 🡪 this is the same as ***Pos’***

***Pos*** does not mean that a person *has* the disease! It simply means that they *tested positive*.

Likewise, ***Neg*** does not mean that a person *doesn’t have* the disease; it simply means that they *tested negative*.

Rapid-screening diagnostic tests typically have sensitivities of around 90% and specificities of around 70% (note that these vary depending on the test). These parameters must be evaluated or estimated based on detailed epidemiological studies. Here are definitions of these terms and how we can write them in terms of conditional probabilities:

**Sensitivity =** The probability of a positive result if patient has the disease =*P[****Pos*** *|* ***Inf****]*

* Note that this is NOT the same as *P[****Inf*** *|* ***Pos****]*

**Specificity** = The probability of a negative result if patient does NOT have the disease = *P[****Neg*** *|* ***Not Inf****]*

***Disease Prevalence***

* We need some sort of background estimate or assumption for disease prevalence, which is simply *P[****Inf****]*
* Note that this is NOT the prevalence of the disease in the population; rather, it’s the prevalence of the disease *in those who take the diagnostic test!*
* For our initial analysis, we’ll assume that only 20% of those who take the diagnostic test actually have the disease: *P[****Inf****] = 0.20*

Follow along with the “Guided Workshop 3: Bayes’ Theorem and Diagnostic Testing” video. When you are done putting together your Excel file, answer the following questions in the “Guided Workshop 3 Submission Quiz” on Coursera (the text fields below are only for your benefit – you won’t be submitting this document).

1. As disease prevalence increases, what happens to PPV? Click here to enter text.
2. As disease prevalence increases, what happens to NPV? Click here to enter text.
3. As disease prevalence increases, what happens to the probability of false positives and false negatives? Click here to enter text.
4. What is PPV when sensitivity = 0.95, specificity = 0.8, and disease prevalence = 0.4? Click here to enter text.
5. Click here to enter text.

**That’s all! 😊**