Modeling: Exploring Medical Test Results

Lesson Overview

In this activity, students investigate ideas that have appeared in many sources related to the recent pandemic. They will become familiar with terms such as false positives and false negatives, prevalence, sensitivity and specificity. The activity has two sections. In the first part of the activity, they will investigate the prevalence of flu and the predictive rate of screening tests based on flu data from the last several years. They will analyze screening test results for a typical rapid test for the flu to estimate the probability that someone who tests positive for the flu actually has the flu and that someone who tests negative actually does not have the flu. In the second part of the activity, students will summarize screening test information in two-way tables and use the data to estimate different conditional probabilities.

About the Lesson and Possible Course Connections:

The activity can be used whenever students have a background in elementary probability and reasoning with percentages. Students are introduced to conditional probability through simulations to develop understanding of the concepts and of the variability inherent in measuring behaviors in the real world. The activity culminates in working with theoretical probability models but does not directly address mathematical formulas such as Bayes Theorem. Students with some familiarity with either margin of error or confidence intervals can use these ideas to develop models for the impact of the disease under different conditions and to relate the work to Type I and II errors in statistical inference.

Learning Goals

Students will be able to:

1. Identify common terms used in reporting screening test results
2. Identify whether the number of false positives or false negatives is more important for a given situation
3. Create and interpret two-way tables involving conditional probabilities
4. Use both empirical and theoretical approaches to investigating probabilities.

CCSS Standards

Statistics and Probability Standards:

- 7.SP.A.1
- 7.SP.A.2
- 7.SP.C.6
- 7.SP.C.7
- 7.SP.C.8
- HSS.IC.A.1
- HSS.CP.A.4
- HSS.CP.A.5
- HSS.MD.B.7

Mathematical Practice Standards

- SMP.4

Modeling: Exploring Medical Test Results

TEACHER NOTES

Lesson Materials

- Compatible TI Technologies:

  TI-Nspire CX Graphing Calculators, TI-Nspire Apps for iPad®, TI-Nspire Software

- Exploring_Test_Results_Teacher Notes.doc
- Exploring_Test_Results_Teacher Notes.pdf

Background

Tests to check whether an individual has a disease are very common. (Note this is very different from investigating how fast a disease spreads or the effectiveness of a treatment or vaccine.) However, the results of these tests can vary considerably, and very few if any tests are 100% accurate. The language used to describe the results can also be confusing. For a certain test, suppose that 14% of all the negative results were "false negatives". What does this mean? For the same test the probability of a "false positive" is 25%. What does this mean? Should a person who tests positive for a disease be really worried? Which is more critical – the false positive rate or the false negative rate? Depending on the prevalence and severity of the disease the answers will vary greatly.

The following activity describes how simulation can be used to investigate answers to some of these questions.

Facilitating the Lesson

Part I. In the winter, the prevalence of flu is approximately 20%, i.e., depending on age and location, 20% of the population has the flu (Smith, 2018). One test for flu is 75% accurate - 75% of those tested get a correct result, i.e., 75% of people who have the flu will test positive, and 75% of those without the flu will test negative. A question that occurs to many people is about the test's impact on them: If my test is positive, what is the probability I actually have the flu?

> **Teacher Tip:** A typical student answer might be 75%. The question is not "if you have the flu, what is the probability that the test is positive?" but rather the converse: "if you test positive, what is the probability that you have the flu?". The answer to the first question is 75%. The answer to the second question depends on knowing how many of the tests overall were positive, which in turn is dependent on the percentage of people who actually have the flu. The simulation activity below explores this thinking.

1) Open-Ended Approach:

Students can be given the information and asked to think about how they might use a simulation to approach the problem. After individual think time, students should share their thoughts in groups of two or three. To prevent the task from being overwhelming or to deter students from just putting the two numbers together without much thought, they might be encouraged to try some simulations to investigate the situation. Students should be careful to think about the meaning of percentages and what cautions should be considered in working with them.

2) More-Structured Approach to Finding a Model:

The teacher might lead the class through the first part of the investigation as described below where each student or small group generates their own simulations, with frequent pauses to check that students understand what they are simulating, what the numbers and lists they generate mean, and how their results compare. It is important to recognize that samples drawn from the same population will vary, that the variability will have a certain regularity depending on the sample size, and this variability will show up when students compare their simulated results.

What to Expect: Example Student Approaches

Exploration 1 What is the probability if a person tests positive, that person actually has the flu?

No Technology:

Give students decks of cards and let them spend a few minutes talking and planning how they might simulate the possible results from the test. For example, some students could propose making two small decks, one with the Ace, King, Queen, Jack and 10 (or 5 other cards) of a certain suit, and another with 1 card from each suit. Using the first deck, students could draw a single card and if the 10 is drawn, the individual has the flu (representing one out of five or 20% having the flu.) Then they draw from the second deck, and if they draw a Spade, their test is an incorrect test, otherwise, their test is accurate (three suits out of four represent the 75% accuracy). For example, a 10 from the first deck and a Diamond from the second deck means the individual does have the flu and tests positive. Have students repeat the process to simulate a small population, say 50 or 100 people, and use the class results to estimate the probability that a person with a positive test result actually has the flu.

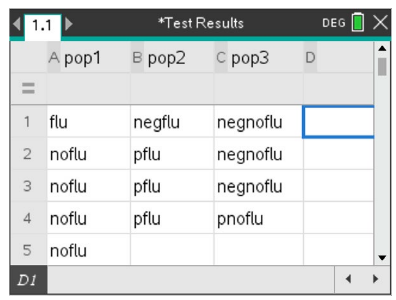
A more sophisticated approach might combine the above process and take the 10-A of all four suits and draw a single card. Each 10 is a person who has the flu, and each Spade is an incorrect test. Let students play with this simulation individually and perhaps as a class, prior to exploring ways that are less time consuming by using random number generators and writing steps to facilitate the process. Ensuring that students actually understand what the technology is doing often means taking small steps to lay the groundwork. Some students might be content with primarily using technology to generate the random numbers and work from there to make statements about the simulation. Several possible pathways are described below.

A first task is to investigate the number of people in a random sample who have the flu assuming it is likely that 20% of the population has the flu. A second task is to investigate the probability of a positive test for someone who has the disease knowing that the test accurately shows a positive result in 75% of the flu cases. And a third task is to investigate the number of those without the flu knowing that the test accurately predicts negative outcomes (those who test negative do not have the flu) in 75% of the cases.

With technology:

To simulate these situations, students might define three representative populations: Note: populations must be entered as strings. Use "flu" for flu and "noflu" for noflu etc.

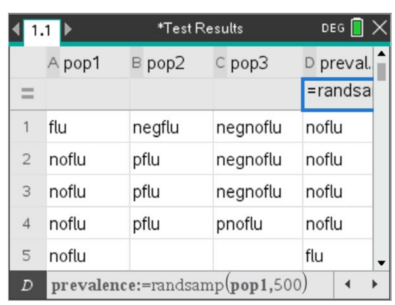
- first, where one person out of five has the flu (pop1 in Figure 1); (Note: any representative population with 20% having the flu and 80% not having the flu would work,)
- second, looking only at those with the flu, three out of four test positive (pop2 in Figure 1),
- and third, looking only at those without the flu, where three people in four test negative (pop3 in Figure 1).



	A pop1	B pop2	C pop3	D
=				
1	flu	negflu	negnoflu	
2	noflu	pflu	negnoflu	
3	noflu	pflu	negnoflu	
4	noflu	pflu	pnoflu	
5	noflu			

Figure 1: Representative populations for the simulation

Based on task one, the next step is to figure out how many in a random sample (say of size 500 people) are likely to have the flu when the prevalence is 20%. This will vary from sample to sample. Simulate the situaton by taking a random sample with replacement of 500 people from pop1, labeled prevalence in Figure 2. (Note that sampling <u>with replacement</u> from pop1 is the same as sampling from an infinitely large population with these characteristics.)



	A pop1	B pop2	C pop3	D preval.
=				=randsa
1	flu	negflu	negnoflu	noflu
2	noflu	pflu	negnoflu	noflu
3	noflu	pflu	negnoflu	noflu
4	noflu	pflu	pnoflu	noflu
5	noflu			flu

D prevalence:=randsamp(pop1,500)

Figure 2: Generating a random sample of 500 people where flu prevalence is 20%

To investigate how a screening test is related to those with the flu, go to a notes page and use the command countif(prevalence,?=flu) to display the number of those with the flu in the random sample labeled prevalence (Figure 3).

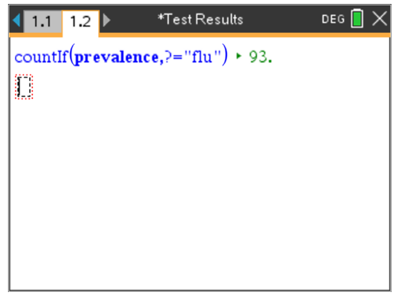


Figure 3: Counting the number with the flu

What were the test results for those who had the flu, 93 people in the example from Figure 3 (task 2)? In column E, create a random sample of size 93 using pop2, which represents the distribution of positive and negative tests among people who have the flu, and observe the results of the test (Figure 4). Name the column test_flu.

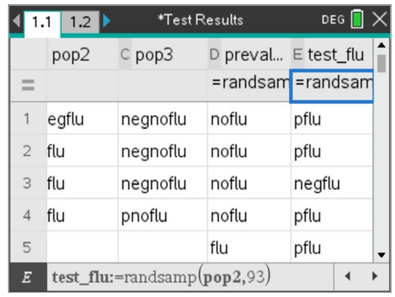


Figure 4: Examining test results for those with the flu

Back on the notes page, use the countif(command to find the number of those with the flu whose test results were positive (Figure 5).

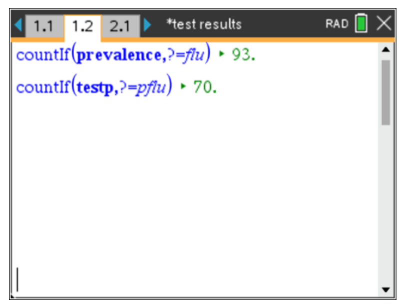


Figure 5: Counting positive test results for those with the flu

To answer the question "given a positive test, what is the likelihood of having the flu?" involves knowing how many altogether (both with and without the flu) tested positive. To examine those in the sample of 500 who did not have the flu knowing that 93 people in the random sample had the flu, then 500-93 = 407 did not. Repeating the process used for those with the flu (task 3), Figure 6 displays a random sample of size 407 drawn from pop3 (assuming 75% of the people without the flu are expected to have an accurate test result that was negative).

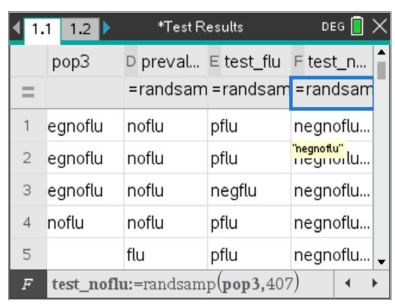


Figure 6: Examining test results for those without the flu

Figure 7 shows the count of those without the flu who tested positive. Use the counts to answer the question, “if someone tests positive, what is the probability they have the flu?” From the simulation, 69 people tested positive and had the flu, and 107+69=176 people tested positive all together. Thus, the probability that someone with a positive test had the flu is 69/176= 0.392. Approximately 39.2% of the people who tested positive actually had the flu.

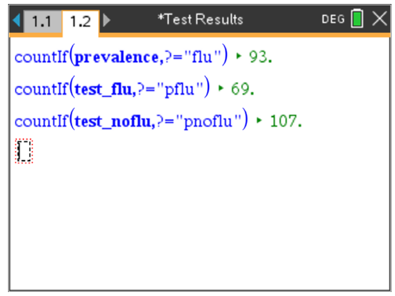


Figure 7: Counting positive test results for those without the flu

The question “if someone tests negative, what is the probability they do not have the flu?” can also be answered from the work above either by using the total of 500 people and subtracting or by repeating the three countif(commands for those who tested negative (Figure 8). Thus, the number of people who tested negative and had no flu was 300, and the total number of negative tests was 324 giving 300/324=0.926 or about 92.6%.

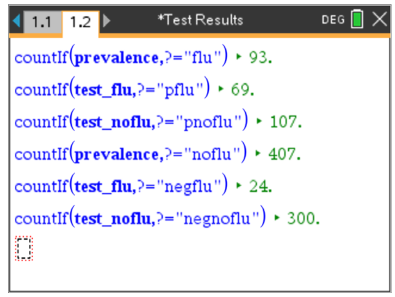


Figure 8: Counting negative test results for those without the flu

Question 2 What are false positives and negatives?

The media often talk about “false positives”- those who test positive for the flu when they did not have the flu- and “false negatives” – those who test negative but actually have the flu. For the simulation above, the false positives would be given someone without the flu, who also tested positive. Thus, the rate of false positives for people who were healthy using the simulated outcomes would be 107/407 or about 26.3%. The rate of false negatives, for people who were indeed sick, would be the proportion of the people who had the flu but tested negative or 24/93 or about 25.8%.

Students should discuss what the 26.3% and the 25.8% mean and which outcome would be more serious,a false positive or a false negative,and why. They should also discuss the difference between the question, "given a positive test, what is the chance I have the flu?" and "the chance of having a false positive result given I have the flu."

> **Teacher Tip:** Note that given the original information that 75% of the tests were accurate actually translates to 75% of the people who had the flu will have a true positive result. And 75% of the people who do not have the flu will have a true negative result.

An efficient way to organize the results of the simulation is to use a two-way table. Table 1, called a contingency table, displays the results of the simulation above.

Table 1 Sample simulation results for 20% flu prevalence and 75% accuracy for positive and negative predictions.

	Flu-infected	No flu-uninfected	Total
Test positive	69	107	176
Test negative	24	300	324
Total	93	407	500

Teacher Tip: Rather than having each student work through the example below, some might change just the prevalence rate, others might change the positive accuracy rate, others the negative accuracy rate and some change all three inputs. Comparing the results should give students insights into the effect of the prevalence on the outcomes. Some might investigate whether the sample size makes a difference, always remembering that results will vary from sample to sample.

<u>Question 3 How does changing the prevalence rate and accuracy rates affect the outcome?</u>

Overall flu prevalence in the United States is about 8% (Tokars et al., 2018). Assume that the test being used correctly detects the presence of the flu 70% of the time, and the test correctly detects the absence of the flu 90% of the time. How will this change the answers to the questions above? The only real difference from the original simulation will be in the representative populations needed to simulate the situation. For pop1, an 8% prevalence is 8 out of 100 or 2 out of 25. To set this up, in Column A, enter flu in rows one and two, noflu in row three and fill down to row 25 (Remember to use " "). For pop2, 70% can be modeled by 7 out of 10, so enter negflu in the first three rows and pflu in the next row and fill down to row 10. For pop3, 90% will be 9 out of 10, so enter pnoflu in row one, negnoflu in row two and fill down to row 10. (Figure 9). Define prevalence in Column D as a sample of 500 people as in the first example.

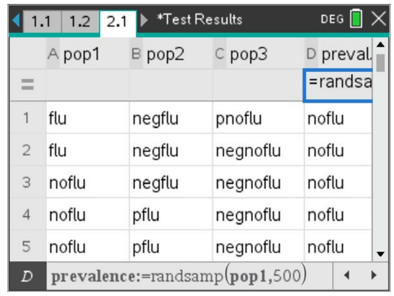


Figure 9: Changing representative populations and generating a random sample

Count the number of those in the random sample with the flu (Figure 10).

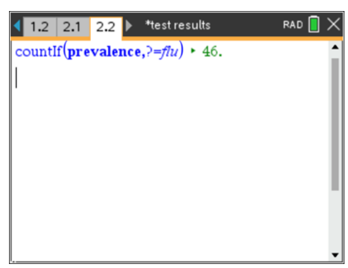


Figure 10: Counting the number of those in the sample with the flu

To check the test results for those who tested positive, in Column E generate a random sample of size 46 (the number of those with the flu out of the sample of 500) and label the column as test_flu. In Column F, generate a random sample of size 454 (the number of those without the flu in the sample of 500) and label the column as test_noflu. (Figure 11)

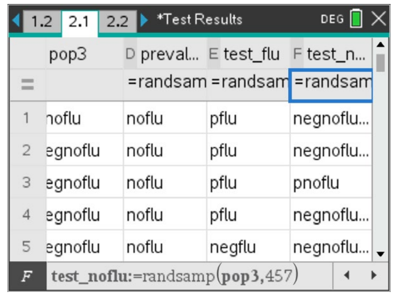


Figure 11: Examining the results of the tests on the sample

Figure 12 displays the counts of those with positive results.

Figure 12: Counting the positive results

The results can be organized in a two-way table (Table 2). Students can use the information to answer the questions investigated above.

Table 2 Sample simulation results for 8% flu prevalence, with probabilities of 70% positive correct and 90% negative correct

	Flu-infected	*No flu-uninfected*	*Total*
Test positive	33	47	80
Test negative	13	407	420
Total	**46**	**454**	**500**

In their groups, students should consolidate their results in a table such as Table 3. They can use this to consider how their answers compare to the probabilities when the flu prevalence was 20% and the accuracy rates were both 75%. Students should simulate other prevalence rates and accuracy rates and add the results to the table. They should look for patterns in the table and take time to verbalize what the numbers tell them; e.g., the false negatives decreased, which means that fewer people will be diagnosed with the flu when they really did not have it.

Table 3 Comparing simulated outcomes for different prevalence rates (to two decimal places)

	Given positive, prob(have flu)	Given negative, prob(no flu)	True Positive	True Negative	False Positive	False Negative
20% prevalence; accuracy 75% positive and negative	39%	93%	74%	74%	26%	26%
8% prevalence; accuracy 70% positive, 90% negative	39% 41%	97% 97%	65% 72%	91% 90%	9% 10%	35% 27%

Formalizing the Vocabulary

Several formal terms are used by scientists in discussing screening tests, and studies show that these terms are often confused. Students might refer back to the results of their simulations and identify the numerical values for each of the terms below:

- *Sensitivity*: the probability of a positive result correctly identifying detection of the condition: the proportion of people with the disease who will have a positive result (true positive)
- *Specificity*: the probability of a negative result correctly identifying absence of the condition (true negative)
- *Positive Predictive Value* (PPV): the probability that people with a positive screening test result indeed do have the condition of interest. To estimate the probability that someone with a positive screening test result does have the flu, divide the number correctly identified by the test by the total number who tested positive for the test.
- *Negative Predictive Value* (NPV): the probability that people with a negative screening test result truly don't have the disease. To estimate the probability that someone with a negative screening test result does not have the flu, divide the number correctly identified by the test by the total number who tested negative for the test.
- *False positive rate*: Number of healthy people incorrectly identified as infected divided by the number of healthy people.
- *False negative rate*: Number of infected people incorrectly identified as healthy divided by the number of infected people.
- *Accuracy*: The sum of the number of healthy people identified as healthy and the number of infected people identified as infected divided by the total number of people.

Teacher Tip: Note that sensitivity and specificity are concerned with the accuracy of a screening test, where the screening test is being assessed. For PPV and NPV, people are being assessed. If a person's screening test yields a positive result, what is the probability that that person has the relevant condition (PPV) and if the screening test yields a negative result, what is the probability that the person does not have the condition (NPV)? (Trevethan, 2017).

Part II. Exploring a pre-made simulation

The variables that were important in setting up the simulations to investigate questions related to screening tests were the prevalence, sensitivity and specificity. The tns file Test Simulation can be used to engage students in quickly and easily exploring the effect of sampling on the results and how different values for the variables affect the results. Students might work through problems such as those below. If students seem to struggle with the terms, the diagram on the Resource Sheet might be useful.

1. Flu testing
   a. In the summer the flu prevalence is about 3%. Assume 90% true positives (sensitivity) and 98% true negatives (specificity). Use the sliders on page 2.2 for a sample of 1000 to estimate the positive predictive value for the given conditions. Interpret this number in the context of an individual being tested for the flu.

   b. Repeat the process in a) to collect about 50 estimates for the positive predictive value for the given conditions. Make a dot plot of the simulated results. Describe the distribution, using measures of center and variability.

   c. Using the distribution from b), explain the difference to someone being tested if the PPV were close to the minimum value of the distribution of the PPV rates. Close to the maximum value of the distribution of PPV rates.

   d. Change the sample size and use the file to create a sampling distribution of possible PPVs as you did in b). How would your answer to part c) change for the new distribution of PPVs?

2. Probabilities can be estimated from experimental results such as those from simulations. In some cases, they can also be found using theoretical probabilities. Use the sliders on page 3.2 for a sample of size 1000 to find the theoretical positive predictive value for the conditions in problem 1.
   a) How do the results compare to the simulated results using page 2.2 of the tns file?

   b) Use the sliders on page 3.2 for a sample of size 1000 to find the theoretical positive predictive value when the prevalence is 20% assuming 90% sensitivity and 98% specificity. Interpret this number in terms of an individual being tested for the flu. How does this outcome compare to those you found in the simulation in part 1?

3. Use the tns file to investigate the statement that sensitivity and specificity are inversely related.

4. Does the size of the population make a difference in finding the predictive values? Explain why or why not.

5. Use the tns file to check the conjectures you made when you analyzed the patterns in Table 3 above.

Validating the Models

Students should validate their models either by asking whether the models make sense in different scenarios related to the context or by finding other information to reflect against the model. The suggestions below might be useful in helping students think about whether their model was reasonable:

1. Students should compare their results to those others found. Note that the values might vary by several percentages. If the results are quite different, they should reexamine what they did.

2. The table below contains the results of a study of measles investigating whether a positive result on a certain test (IgM) is sufficient to confirm the presence of measles (Bolotin, et al., 2017). Use the information in the table to verify the positive and negative predictive values, sensitivity and specificity.

Performance of measles serology test Canada, 2009-2014

IgM serology result	classification confirmed	classification Not confirmed	Positive predictive Value	Negative predictive value	Sensitivity	Specificity
Positive	42	199	17.4	97.2	79.2	65.7
Negative	11	381				

3. Create a two-way table using the theoretical probabilities for the examples in Part 1 and compare the results to the those from the class simulations done above.

Teacher Tip: This is another opportunity to point out the variability inherent in any prediction; here it is the difference between the theoretical probability and empirical probability, which involves variability. Students should recognize that an outcome for a given situation will vary depending on slight changes in the situation but that the variability can be quantified by observing the overall patterns in many replications of the simulations.

Extension

1. Decide whether the following are true or false and explain the reasons for your decisions in each case.
   a) If a test for a disease is 99% accurate and you receive a positive result, the chance that you actually have the disease is 99%.
   b) if you test positive for a rare disease (one that affects, say, 1 in 100,000 people), your chance of having the disease might be less than the percent that actually have the disease.
   c) If an antibody test has a specificity of 98%, i.e. 98% of people without antibodies correctly test negative, 2% of all people without antibodies will test false-positive.
   d) The impact of false-positives is larger when most people who are being tested don't have the antibodies being tested for.

2. An article from the Center for Disease Control and Protection, Rapid Diagnostic Testing for Influenza: Information for Clinical Laboratory Directors. (https://www.cdc.gov/flu/professionals/diagnosis/rapidlab.htm) states
   - "When influenza prevalence is relatively low, the positive predictive value (PPV) is low and false-positive test results are more likely. By contrast, when influenza prevalence is low[and] the negative predictive value (NPV) is high, and negative results are more likely to be true."
   - "When influenza prevalence is relatively high, the NPV is low and false-negative test results are more likely. When influenza prevalence is high[and] the PPV is high, and positive results are more likely to be true."

   The article provides tables to support the claims. Use the tables and the mathematical meaning of each of the terms to explain why the two statements are true.

3. According to a New York Times report on August 20, 2020, a study of 120 people aboard a ship found six that, when given an Abbott test before the boat's departure, had antibodies to the virus indicating prior exposure. But when the researchers reanalyzed those samples using more sophisticated tests, only three of the six were confirmed to have antibodies, suggesting that three test results were false positives. The Abbott test is advertised as returning fewer than one false positive for every 100 samples. Why did a researcher say: "That's a little concerning that the Abbott may be a little less specific than we thought,"?

4. Find two examples of real contexts where a) a false positive is of more concern than a false negative and b) where a false negative is of more concern than a false positive.

5. Other diseases and screening tests
   Choose two or three of the diseases below or another disease you would like to investigate and fill in the table for a population of 10,000 people. (Note in some cases you might need to solve an equation to find the solution, using the solve functionality of Nspire.) What do you notice? Wonder?

Disease	***Positive Predictive value***	***Negative Predictive value***	***Prevalence***	***Sensitivity (True positive)***	***Specificity (True negative)***	***False positive***	***False negative***
Peanut allergy	22%		2%	28%			
Covid-19			9%**	90%*	99%*		
Rubella	3.6%		1.2%	100%			
Mammograms under 40 Over 75			2% 2%	76.5%	87.1% 93.5%		12%
HIV Brazil US	33%		5% 2%	99% 99%	99%*		
Prostate cancer	30%	85%	17% 12%	21%			9%

HIV

https://www.aidsmap.com/about-hiv/false-positive-results-hiv-tests *
https://www.who.int/hiv/mediacentre/news/hiv-misdiagnosis-qa/en/index5.html **

Prostate cancer

https://www.uptodate.com/contents/screening-for-prostate-cancer

Peanut allergy

https://www.sciencedaily.com/releases/2018/05/180503085604.htm

Mammogram

https://emedicine.medscape.com/article/1945498-overview#a5

COVID-19

https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/testing-in-us.html**
https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antibody-tests-guidelines.html#table1

Rubella

The utility of measles and rubella IgM serology in an elimination setting, Ontario, Canada, 2009–2014

6. Explain the connection between Type I and II errors in hypothesis testing and false positives and false negatives.

7. If the prevalence of a disease is estimated to be 20% given a sample of 1000, find a margin of error for the true prevalence of the disease. Find the percentage of false negatives and false positives for the upper and lower bounds of the interval using a sensitivity of 90% and specificity of 95%. Explain what the difference in the results tells you about the screening test.

8. Retesting:

   In many instances, people are encouraged to have a second test to confirm the result of the first test.

   a) With prevalence 8%, sensitivity 70% and specificity 90% use the tns file to simulate the probability you actually have the flu, given a positive test.
   b) Use the total number of positive tests, refigure the prevalence within all those who had positive test, and use that prevalence and the total number of positive tests to simulate the probability of actually having the flu after two positive tests. What is that probability?
   c) What do your results suggest? What do you conjecture will happen to the results if you did a third test?

Resource:

		The Truth:		
		Has the disease	Does not have the disease	
Test Result:	Positive	TP True Positives	FP False Positives	Positive Predictive Value $\frac{TP}{TP+FP}$
	Negative	FN False Negatives	TN True Negatives	Negative Predictive Value $\frac{TN}{TN+FN}$
		Sensitivity $\frac{TP}{TP+FN}$	Specificity $\frac{TN}{TN+FP}$	

Trevethan, R. (2017). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5701930/

References:

Bailer, J. (August 25, 2020). My COVID-19 test is positive… do I really have it? Statisticans React to the News. International Statistical Institute. https://blog.isi-web.org/react/2020/08/my-test-is-positive/

Bolotin, S., Lim, G., Dang, V., Crowcroft, N., Gubbay, J., Mazzulli, T., & Schabas, R. (2017) The utility of measles and rubella IgM serology in an elimination setting, Ontario, Canada, 2009–2014 PLoS One. 2017; 12(8): e0181172.

Glen, S. "False Positive and False Negative: Definition and Examples" From StatisticsHowTo.com: Elementary Statistics for the rest of us! https://www.statisticshowto.com/false-positive-definition-and-examples/

Smith, C. (August 18, 2018). Flu Tests in the summer and other bad ideas https://journalfeed.org/article-a-day/2018/flu-tests-in-summer-and-other-bad-ideas

Tokars, J., Olsen, S., & Reed, C. (2019). Seasonal Incidence of Symptomatic Influenza in the United States. PMC. 66(10), pp.1511-1518. US National Library of Medicine. National Institutes of Health. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5934309/

Trevethan, R. (2017). Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. 5(307).Frontiers in Public Health. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5701930/