Homework 2 Problem 2

William Matzko

February 11, 2021

Let event A be a person has cancer. Let event B be a positive test. We wish to find the probability that a person has cancer given that they have a positive test, $P(A|B)$. We can apply Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

From the problem statement, we are given that 80% of people with cancer will test positive, so $P(B|A) = 0.8$. We are also given that the probability that someone (in this age group) has cancer is 2%, so $P(A) = 0.02$. We just have to find $P(B)$. The probability of having a positive test involves the probability of a person having cancer and being tested positive, *and* the probability of a person testing positive while not having cancer. Formally, this can be expressed as the law of total probability:

$$P(B) = \sum_n P(B|A_n)P(A_n)$$

where A_n is a possible event of interest and n is the number of possibly events (loosely speaking, I don't think this is a completely accurate characterization). In our case, this becomes

need A_n are exhaustive and exclusive to tighten.

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

where A^c is the complement of event A, meaning a person does not have cancer. We already know what the first term in the sum is; that is given in the problem statement as described above. We are also given that the probability of someone testing positive and not having cancer is 9.6%, so $P(B|A^c) = 0.096$. We know that only 2% of people have cancer, so $P(A^c) = 0.98$. Thus, we have

$$P(B) = (0.80)(0.02) + (0.096)(0.98) = 0.11008$$

and the probability is

$$P(A|B) = \frac{(0.80)(0.02)}{0.11008} = 0.145$$

Thus, the probability that a person has cancer given they have a positive test is only 14.5%. I would say that this isn't a particularly reliable test.

There would be lots of debate in the medical literature about this. Need to balance how many patients are saved vs cost of so many patients visiting Dr. after a false positive.

The problem asks us to include figures, so please enjoy the beautiful MS paint drawings below. The first figure shows a rough sketch of the sample space and its partition into who tests positive/negative. Personally, I don't find this figure very useful, but I thought it was worthwhile to show a rough visualization of this.

The second figure is, in my opinion, a little more illuminating. It gives a visualization of $P(B)$ and offers insight into why the formula given above for it is true. The left (green) partition tells us that 2% of the people have cancer. The right (red) partition tells us that, naturally, that 98% of people don't have cancer. The probability of testing positive is weighted by the efficacy of the cancer test. In other words, each of these cancer populations has a certain probability of having a positive test. Hence, to find the probability a group tests positive, we should multiply the 'abundance' of the group (population percentage) by the probability that group tests positive. Of course, we would subsequently add up those probabilities for each group to obtain $P(B)$, which is what we see in the above formula.

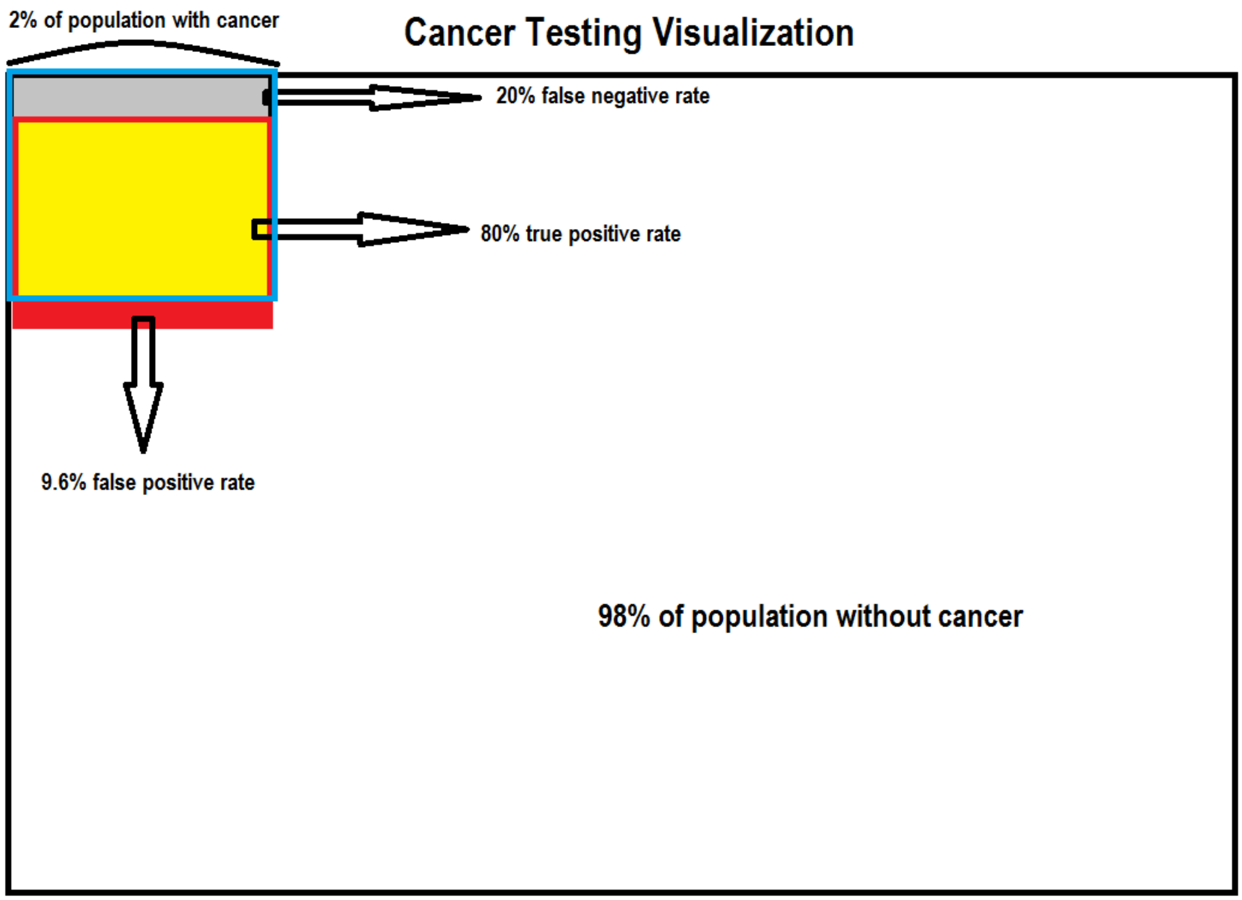

Figure 1: Diagram of cancer population (ages 50-60). The blue square represents the 2% of the population that has cancer. The shaded yellow region represents the 80% who test positive that have cancer. The grey region is those who have cancer but test negative. The shaded red area shows the false positive rate. Not to scale.

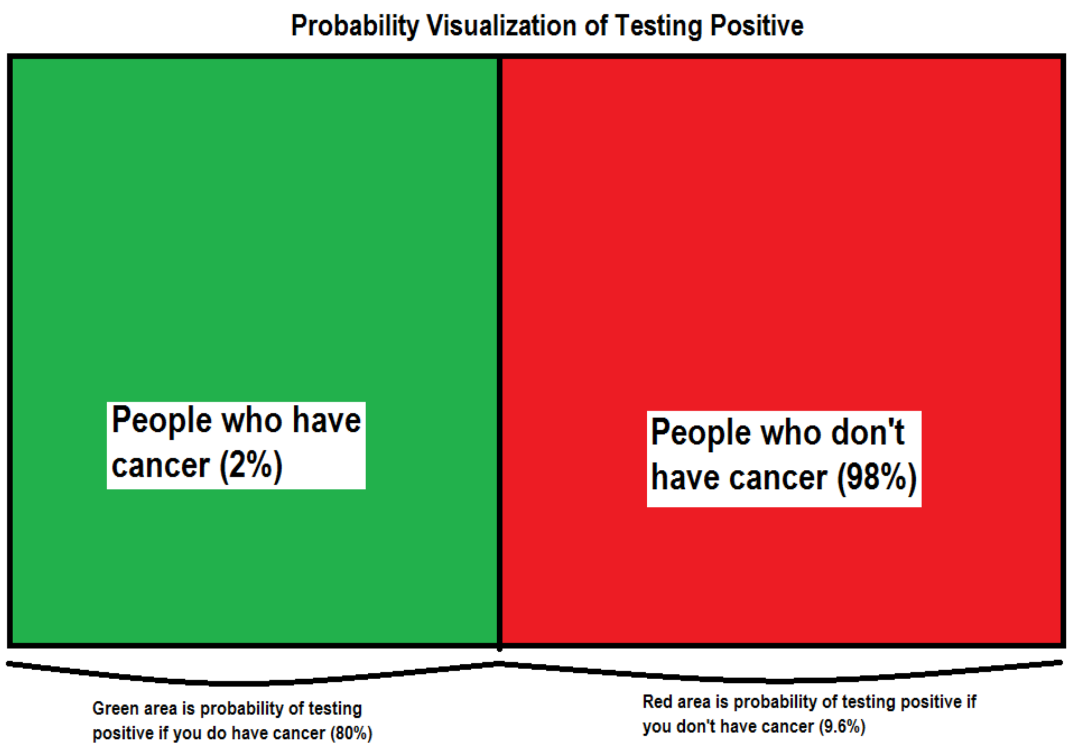

Figure 2: Visualization of the probability of testing positive. The green and red side have two 'components' each: the percentage of people who fall into that category, and the probability of that category testing positive. Not to scale.