Consider the Following Scenario:

During a routine annual physical examination, your doctor orders a number of tests for various medical ailments. Unfortunately, when the results come back, the doctor informs you that the test for Disease X came back positive. You are stricken with fear and shock and consider all the ways in which your life has now been altered as the doctor discusses treatment options. Wisely, you go home to ponder your situation before committing yourself to a particular course of action.

As the emotion subsides you remember how to do predictive analysis with base rate assumptions and Bayesian mathematics. You do a little searching on the internet and find the following information from multiple independent sources.

- The probability of any one person having Disease X in the general population is 0.014, which is a 1.4% chance.

- The chance that the diagnostic test that the doctor gave you correctly detects Disease X when you do in fact have the disease is 0.786 or 78.6%.

- The chance that the diagnostic test gives a false positive for Disease X, which is to say the test says you have the disease when you in fact do not, is 0.107 or 10.7%.

Your initial inclination when looking at this information might be to assign the chance that you have Disease X as 78.6%, since that's the chance the diagnostic test will accurately say you have the disease if you do. With such a large chance of having the disease, some very invasive and difficult treatment options seem like a good choice since there are overwhelming odds that you have the disease. However, upon further reflection, you realize that this might not be the case.

Consider that before you had the diagnostic test, your chance of having Disease X was 1.4%, the occurrence rate in the general population. Let's call that P(A), the probability that A is true, where A is having Disease X.

But now you have had the diagnostic test with a positive result. This reflects a change in information affecting the assessment. Now we have to do a little math using Bayes' equation to find out what your chances of having Disease X are in light of the additional information.

Let's call Value B the result of the diagnostic test. If B is true then the test came back positive, if B is not true then the test came back negative. Now we can construct the following two probabilities:

P(B/A) = the probability that the test comes back true (B) when A (you have the disease) is true. From our research this is 78.6%.

P(B/~A) = the probability that the test comes back true (B) when A is not true. (That's what the little ~ means, it means A isn't true; you don't have the disease.) Our research said that this is 10.7%.

Okay, now let's use a form of Bayes' equation. In this equation, the P(A/B) is the probability that you have Disease X (A) given that the result of the test (B) was true.

$$P(A/B) = \frac{P(A) \times P(B/A)}{P(A) \times P(B/A) \; + \; P(B/\sim A) \times (1 - P(A))} \qquad \text{(Equation 1)}$$

Substituting in the probabilities we found in our research we get:

$$P(A/B) = \frac{0.014 \times 0.786}{0.014 \times 0.786 \; + \; 0.107 \times (1 - 0.014)}$$

When you work that all out you find that P(A/B) = 0.094 or 9.4%

Wow. This is a very different result than almost 80%. This would very likely affect your decision to undergo invasive and difficult treatment since you know that the chance you have the disease is less than 10%. How is it that this very counter-intuitive result can be?

The answer is that the incidence rate (base rate) of having Disease X is very low, and the chance of having a false positive on the diagnostic test is significant. So what? Well, consider giving the diagnostic test to 1000 people. 14 of them actually have the disease, and the test will show positive for 11 of them. 986 do not have the disease, but with a 10.7% false positive rate, 106 of them will get a false positive test result. Therefore, a total of 120 people get a positive result, but only 11 actually have the disease. Hence, there is only a 9% chance that you actually have the disease with a positive test result.

So what should you do? Order another test! The result will further refine your probabilities of having the disease. P(B/A) and P(B/~A) remain the same since they are properties of the diagnostic test, but P(A) has now changed to 9.4% since that is the current probability that you have Disease X based on the results of the first test. Plugging in the numbers into Equation (1) above now yields:

P(A/B) = 0.434 or 43.4%

With two independent diagnostic tests for Disease X giving a positive result, your chances now are 43.4% that you actually have the disease. Repeating the test again and getting a third positive result yields an 84.9% chance you have the disease. At this point, it would be reasonable to proceed with invasive and difficult treatment since your chances are quite good that you do in fact have the disease.

But it would not be reasonable to proceed with aggressive treatment based on the results of only one diagnostic test with these probabilities. Yet, this happens all the time, since doctors are not immune from not understanding counter-intuitive Bayesian mathematics.

Bottom line, be good at predictive analysis and protect yourself from those who aren't.

P.S. By the way, if you plug in the numbers for your chances of not having the disease when the initial diagnostic test comes back negative, the result comes out to a 99.6% chance that you don't; so diagnostic tests like this aren't necessarily bad, if the information is interpreted correctly.

References:

Silver, N. (2012). *The Signal and the Noise.* New York: Penguin Press.