HOW TO DECIDE WHAT TO DO

WILLIAM M. BRIGGS

Abstract. This article will provide you with an introduction and a step-by-step guide in how to make good decisions in particular situations (I'll have a list of these situations at my web site wmbriggs.com). These techniques are invaluable whether you are an individual or a business. Please email me if you have any questions: william@wmbriggs.com.

The results that you'll read about hold for all manner of examples – from lie detector usefulness, to finding a good stock broker or movie reviewer, to intense statistical modelling, to financial forecasts. But a particularly large area is medical tests, and it is these that I'll use as examples.

Many people opt for precautionary medical tests – frequently because a television commercial or magazine article scares them into it. What people don't realize is that these tests have hidden costs. These costs are there because tests are never 100% accurate. So how can you tell when you should take a test?

1. When is worth it?

Under what circumstances is it best for you to receive a medical test? When you "Just want to be safe"? When you feel, "Why not? What's the harm?"

In fact, none of these are good reasons to undergo a medical test. You should only take a test if you know that it's going to give accurate results. You want to know that it performs well, that is, that it makes few mistakes, mistakes which could end up costing you emotionally and financially.

Let's illustrate this by taking the example of a healthy woman deciding whether or not to have a mammogram to screen for breast cancer. She read in a magazine that all women over 40 should have this test "Just to be sure." She has heard lots of stories about breast cancer lately. Testing almost seems like a duty. She doesn't have any symptoms of breast cancer and is in good health. What should she do?

What can happen when she takes this (or any) medical test? One of four things: The test could correctly indicate that no cancer is present. This is good. The patient is assured. The test could correctly indicate that a true cancer is present. This is good in the sense that treatment options can be investigated immediately. The test could falsely indicate the no cancer is present when it truly is. This error is called a false negative. This is bad because it could lead to false hope and could cause the patient to ignore symptoms because, "The test said I was fine." The test could falsely indicate that cancer is present when it truly is not. This error is called a false positive. This is bad because it is distressing and could lead to unnecessary and even harmful treatment. The test itself, because it uses radiation, even increases the risk of true cancer because of the unnecessary exposure to x-rays.

Date: 8 August 2002.

	Presence	Absence
Test +	Good: True Positive	Bad: False Positive
Test -	Bad: False Negative	Good: True Negative

Table 1. Test Graph

	Presence	Absence
Test +	0	Cost +
Test -	Cost -	0

Table 2. Cost Matrix

2. A graphical view

Here is a graph that labels all the possibilities in a test for the presence of absence of a thing (like breast cancer, a lie, AIDS, and so on). For mammograms, "Presence" means that cancer is actually there, and "Absence" means that no cancer is there. For a lie detector, "Presence" means a lie is actually there, and "Absence" means that truth is there (more on this later).

"Test +" says that the test indicates the presence of the thing (cancer), before who or what is doing the testing knows whether or not the thing is truly there or not. "Test -" says that the test indicates the absence of the thing.

There are two cells in this graph that are labelled "Good," meaning the test has performed correctly. The other two cells are labelled "Bad," meaning the test has erred, or made a mistake. Take a moment to study this graph to be sure you understand how to read it because it will be used throughout this article.

3. Error everywhere

The main point to take away is this: all tests and all measurements have some error. There is no such thing as a perfect test or perfect measurement. Mistakes always happen. This is an immutable law of the universe. Some tests are better than others, and tables like this are necessary to understand how to rate how well a particular test performs.

4. The costs

The same graph can be used to examine the costs of the test's performance.

When the test performs correctly (true positives and true negatives) there are no costs. The graph shows this by putting a 0 in these cells. There may be, of course, costs that must be paid to have the test; however, this cost is identical for every cell so it is not included. There may also be, in the case of true positives, subsequent treatment (or other) costs – but this is not because of the test. It is assumed that you would want to pay these costs as the test was correct – there is no error cost of the test.

But when the test gives a False Positive or False Negative there is a definite cost: these costs are labelled "Cost +" and "Cost -". They do not have to have strict dollars figures attached to them – for example, they may be emotional costs. In some cases it will be possible to specify exact dollar amounts. Examples will be

	Presence	Absence
Test +	7	70
Test -	1	922

TABLE 3. Mammogram Performance

	Presence	Absence
Test +	0	0
Test -	8	992

TABLE 4. Naive-o-gram Performance

given below that will make this distinction clear. Meanwhile, costs are only part of judging the goodness of a test. Performance is another.

5. Performance

Our framework can now be used to examine actual test performance. In this graph are the performance statistics from actual mammograms (From Gerd Gigerenzer's book, Calculated Risks: How to know when numbers deceive you, published 2002. This book will be quoted often here: a review of it appears at my web site, wmbriggs.com. The data is from the screening of American women over 30.).

This is called a performance table, and the cells have the same meaning as before except that the entries are the numbers from actual tests.

The data in this table are for an average 1000 women who have had mammograms. Of these 1000 women, 922, or 92.2% did not have cancer, and the test correctly indicated this.

Seven women out of every 1000, or 0.7%, had her cancer correctly identified by the mammogram. One woman out of every 1000, or 0.1%, will have her breast cancer missed by the test. A full 70 out of 1000, or 7%, will show a false positive.

6. How accurate?

The first question to ask of any test that you are considering having is: how accurate is it? Accuracy is found by adding the True Positives and True Negatives then dividing by the total number of tests. In the mammogram example, this is (922 + 7)/1000, or 92.7%.

An accuracy of 92.7% sounds impressive, but is it the best that can be done? Obviously not. The best that can be done is 100%! We already know that this is impossible (all tests have error). But is there an even simpler test than a mammogram that is more accurate? A test that could be substituted for the mammogram for no cost? The answer, perhaps surprisingly, is yes.

Look at this performance table for what I'll call a naive-o-gram, which is an exam that I perform and which simply says that every woman who comes to me for the test does not have cancer.

It's important to understand how we get the naive-o-gram results. We know from the Mammogram Performance Table that 70 + 92 = 992 women out of every 1000 do not have breast cancer. The naive-o-gram, which says "No Cancer" each time, would identify all of these 992 women correctly.

We also know that $7 + 1 = 8$ women out of every 1000 do have cancer. The naive-o-gram will make a mistake for these 8 women (a False Negative). So we can fill in the Naive-o-gram Performance Table without having to do the test by only knowing the background rates of cancer in the population (more on this later).

The naive-o-gram will never have a false positive, nor will it have a true positive because it never labels a woman as having cancer. These top cells are always 0.

Here's the crazy thing: the accuracy of the naive-o-gram is 99.2%, which is much more accurate than the real mammogram! So, considering only accuracy, which test would you rather have? The naive-o-gram or mammogram?

Of course, you don't have to have to come to me to take the naive-o-gram, you can do it yourself. Just stand up and say, "I don't have cancer" and you're finished. And why not? It's more accurate than the best scientific test. So why aren't more doctors using the naive-o-gram?

7. The difficult part: calculating costs

Accuracy isn't everything and it could turn out that – for you – a less accurate test is better than a more accurate test. How could this be?

Describing how it can be first requires an understanding of the two costs mentioned above. The answer will depend on how the ratio of these two costs interacts with the predictive accuracy of the test. Just how will be shown below, but first let's go through an example of how to calculate the costs for a mammogram. (Similar tables can be built for any predictive test: examples will be added in time for lie detectors, stock picks, movie reviewers, and so on.)

Shown in this Cost Comparison Table are examples for a mammogram. These are only examples: I am not a physician and am only estimating what I believe are the most likely costs. Your actual costs are best filled in by you and your doctor. The examples that I list are from Gigerenzer's books.

This is very important! The only way to find and value these costs is to first imagine that the case that lead to them is true: the costs are conditional on these states being true. For example, you have to imagine yourself in the case that you know that the mammogram has made a mistake (false positive and false negative). Then fill in the table. You must imagine all the bad scenarios that can happen if the test makes a mistake.

Now the hard part. Each of these costs must be rated and assigned some sort of score. The good news is that there is no need to give a dollar figure to each cost. All that is needed is to assign a relative difference between the two. An example will clear up what I mean by that.

8. Examples

Say that you are desperately scared of breast cancer. The very thought of it fills you with a terrible dread. You don't care about false positives, you don't care if you have to take dozens of mammograms, suffer through biopsies, and possibly undergo unnecessary treatment. Anything, to you, is better than not starting treatment on true cancer should it develop. Likewise, the thought of missing the cancer in a mammogram is frightening. You want to know as soon as possible.

If you felt like this you would certainly rate Cost - higher than Cost +. Would you say Cost - was twice as high as Cost +, four times, ten? It's up to you to pick a number after going through each list. Higher numbers reflect higher costs.

Cost + (False Positive)	Score	Cost - (False Negative)	Score
1. Stress, worry, depression. These affect health and well being.		1. Cancer allowed to develop to a potentially dangerous size	
2. Follow-up tests necessary: prolongs time of worry.		2. Cancer symptoms can be ignored because "The test said I was fine."	
3. Possible biopsy required, with risk of infection.		3. Treatment is delayed (although no treatment is guaranteed, there is evidence that earlier treatment is more effective at extending life).	
4. Unnecessary and possibly harmful surgical procedures used (like mastectomies and lumpectomies).			
5. The finding and unnecessary treatment of harmless growths.			
Total:		Total :	

Table 5. Mammogram Cost Comparison Table

Of course, if you have actual dollar figures, use these. Some situations, like stock picks, will have natural interpretation (dollars won and lost, for example), others will not.

One way to do this is to go through each point of the lists and assign a score, a number which reflects your feeling. For example, you might assign the first item under False Positives a "10." The 10 is, of course, arbitrary and it only has meaning in relation to the other items in the list. The 10 could mean dollars or "stress units" or anything. It's up to you. Your only goal is to be consistent across all items. Here is one possible table.

I pretended that I was a woman and labelled the costs for each possible error. As you can see, I thought the items under Cost - (False Negatives) were much worse than the errors for False Positives. You probably feel the same. And remember: you are not judging the likelihood of any error here. You are assuming the error is true, that it actually has happened to you, and then you're scoring its cost. I'll show you how to fit it in with test performance in a minute.

I thought that the total error for False Negatives was 210, and for False Positives it was 70. It will become important to look at these numbers is through their ratio. I'll be giving you a calculator to do all this, so don't worry about the math. The ratio will always be (Total Cost False Positives) / (Total Cost False Negatives) = 70 / 210 = 1/3. Thus, I thought that False Negatives were three times worse than False Positives.

9. Your costs are different than the testers

Is that it? Not quite. These are my costs, yours may be slightly different. But your costs are not the same as for the doctor (or advocacy group) who orders the test. This means that your goals are not the same as your doctors (or stock brokers,

Cost + (False Positive)	Score	Cost - (False Negative)	Score
1. Stress, worry, depression. These affect health and well being.	10	1. Cancer allowed to develop to a potentially dangerous size	50
2. Follow-up tests necessary: prolongs time of worry.	15	2. Cancer symptoms can be ignored because "The test said I was fine."	100
3. Possible biopsy required, with risk of infection.	15	3. Treatment is delayed (although no treatment is guaranteed, there is evidence that earlier treatment is more effective at extending life).	60
4. Unnecessary and possibly harmful surgical procedures used (like mastectomies and lumpectomies).			
5. The finding and unnecessary treatment of harmless growths.	10		
Total:	70	Total :	210

TABLE 6. Mammogram Cost Comparison Table

Cost + (False Positive)	Score	Cost - (False Negative)	Score
1. Follow-up tests necessary.	10	1. Cancer allowed to develop to a potentially dangerous size	500
2. Possible biopsy required, with risk of infection.	10	2. Cancer symptoms can be ignored.	100
3. Unnecessary and possibly harmful surgical procedures used (like mastectomies and lumpectomies).	20	3. Treatment is delayed (although no treatment is guaranteed, there is evidence that earlier treatment is more effective at extending life).	200
4. The finding and unnecessary treatment of harmless growths.	10	4. Possible malpractice suit brought for missing cancer.	200
Total:	50	Total :	1000

TABLE 7. Doctor's Cost Comparison Table

or polygraph examiners, and so on). You may not be able to estimate their costs, but it is important for you to understand that these cost differentials can lead you and your doctor to reach different conclusions about whether to have test or not.

Here's an example of a doctor's cost. Again, I am making these up. These will be different for any particular physician. The important thing for you to understand is that these costs are almost always going to be different for you.

As you can see, not only are the costs different, but the items are not all the same. For example, as a doctor I have to worry about malpractice suits from missing cancers. These will cause my (already outrageously high) insurance rates to rise, and I could lose respect and business. The ratio is (Total Cost False Positives) / (Total Cost False Negatives) = 50 / 1000 = 1/20.

You cannot directly compare the scores from the doctor's table to your scores. The units are arbitrary and meaningless except for one person. A 10 for me may be a 112.8 for you. So how can you compare tables if you can't compare scores?

10. Skill

Skill is defined as the ability of an expert predictive system to perform better than a naive prediction system. In the mammogram example, a mammogram would have skill if it performed better than the naive-o-gram. We have already seen that the mammogram is worse than the naive-o-gram with regard to accuracy. The mammogram thus has no skill. But we have yet to see how the naive-o-gram compares to the mammogram with regard to cost.

You would never want to use a predictive system that does not have skill. But notice that part of the definition of skill requires us to supply a "naive" prediction system. The naive-o-gram came out because it turned out that the probability of any woman having breast cancer was so small. A natural naive prediction was to say "you do not have cancer" for each woman. Different systems, such as lie detectors and stock predictions, may have different naive systems. More on this later.

First, here's how we handle cost for the expert and naive prediction systems. The Expected Cost Comparison Table below has the results. Here's how it works.

11. Expected Cost

As we already know, all predictive systems have error. We have already learned how to rate the costs of these errors through filling in a Cost Comparison Table for the two kinds of errors, Cost - and Cost +. We also know how to look at a Performance Table (we don't yet know how to get the numbers of a Performance Table, which generally must be supplied by experiment – more on this later). These data now give us enough to let us make a decision. It's about time!

We know have to define the concept of expected cost. That is the error cost of the test that we would expect any random person to experience (given they had your values in the Cost Comparison Table). This is a statistical concept and it means the costs that the average person will experience – it does not imply that the expected cost is the cost that any given person will experience, just the average person.

Calculating this cost is easy, but it does require some work. Don't worry about the math, because I'll be giving you a web page that does the work for you. But we are going to go through an example here so that you can see how it works. We first need to modify the Performance Table into a Performance Probability Table. This is simple because all we need to do is to divide each cell by the total of all cells. The Mammogram example is given below.

This was an easy case because the total was 1000 – making division simple. Now we have to multiply each error cell's probability by your cost estimate and then calculate the total. That sounds complicated, but here's an example.

Focus on the row where it says "Woman: Mammogram." We know, from the Mammogram Cost Comparison Table, that the Cost + is 70, and we know that the probability of a False Positive is 0.07. Multiplying these numbers together gives

	Presence	Absence
Test +	0.007	0.07
Test -	0.001	0.992

Table 8. Mammogram Performance Probability Table

Test	Expected Cost +	Expected Cost -	Total
Woman			
Mammogram	0.07 * 70 = 4.9	0.001 * 210 = 0.21	5.11
Naive-o-gram	0 * 70 = 0	0.008 * 210 = 1.68	1.68
Doctor			
Mammogram	0.07 * 50 = 3.5	0.001 * 1000 = 1	4.5
Naive-o-gram	0 * 50 = 0	0.008 * 1000 = 8	8

Table 9. Mammogram Performance Probability Table

4.9. We further know that the Cost - is 210 and that the probability of a False Negative is 0.001. Multiplying is 0.21. We add these two together and get the expected error cost of the mammogram, which is 5.11.

We can now do the exact same calculations for the naive-o-gram. The costs remain the same, and all that changes are the probability estimates. There is no chance of a False Positive by definition, so the expected cost of a False Positive is 0. There is a higher chance of a False Negative, here 0.008, so the expected cost is 1.68.

And that's it. To make the best decision all you need to do is to choose the test with the lowest expected cost. For this example, that choice is the naive-o-gram, which has an expected cost three times lower than that of the mammogram.

12. Skill score

What if there are competing versions of the expert test and we want to rate them? How can we tell which is best? We do that using a skill score. This is a score that lets us compare different expert predictors even though they have different underlying base rates. Calculation of the skill score is somewhat complicated, so I won't give the details here (the calculator does it for you: you can find it at my web site). All you need to remember is that the skill score must be positive for you to choose the expert prediction. If the score is zero of negative, then you should choose the naive guess.

$$\textit{mammography skill score} = -2.042$$

The mammography skill score is negative, so we would choose the naive guess.

13. Different choices

We can also do the same calculations for the doctor, which are given in the table above. As you can see, his best bet is to opt for ordering the mammogram! Why? Because he weights the costs differently; he's far more worried about False Negatives and this worry shows up in the expected cost. Note: you cannot compare the expected costs of the doctor with your expected costs – the numbers have different meanings. These costs can only be compared against themselves, between predictive systems for one person (and that person is whoever specified the Cost Comparison Table).

The doctor's best decision is to order the mammogram, your best decision is to not take it. Who wins in the end? Probably whoever has the stronger will (usually the doctor). But remember: It is always your decision to accept any medical procedure. And you should never make these decisions lightly. And I can only hope that this guide helps to make this decision easier.

14. Other examples

Is this all? Not quite. It can be that the expected loss for the expert prediction is less than that of the naive guess, but it may be so only because of chance. There is a statistical test based on the skill score that lets us tell. If you have questions about this, please email me.

What about lie detectors, stock picks, and other decision types? I'll be posting examples of each, along with an online calculator. So, if you have information about performance statistics for any decision, please send them to me and I can help you fit them into the decision calculator. My email address is at the bottom of this page.

Naturally, I have left out a lot of details. I'll post more of these through time. Questions are always welcome.

Some predictions are not yes or no, like the examples given here. One example is a high temperature forecast, which is a number like "82 degrees." Can forecasts like this be fit into the decision calculator? The answer is yes, although the math becomes more complicated. It also becomes more useful.

If you are a manager of a brokerage, I can show you how to use these methods to rate your brokers – even if these brokers pick different stocks over different time periods. If you are making sales forecasts, or want to rate a group's decision making success, these methods are a must. Please contact me for information.

300 E. 71st, Apt. 3R, New York, NY, 10021
E-mail address: william@wmbriggs.com