Interactomics Basics and Applications

Prof. Sanjeeva Srivastava
Dr. Joshua LaBaer
Department of Biosciences and Bioengineering
Arizona State University, USA
Indian Institute of Technology, Bombay

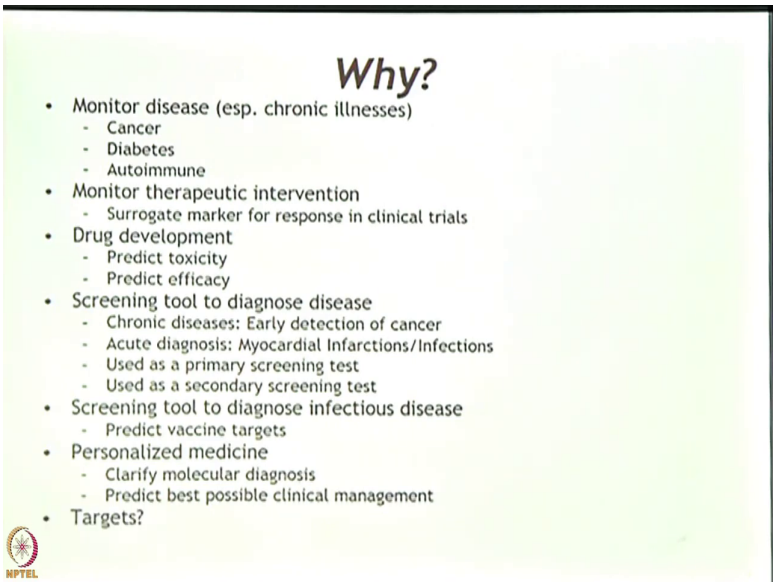
Lecture – 10

Biomarkers: Harnessing the immune system for early detection of disease - III

Today Professor Joshua LaBaer will discuss about Biomarkers in more detail and also different phases of biomarker detection. He will then continue talking about the importance of a statistics for the biomarker discovery program. Usually biomarker discovery programs are very challenging because to have a real biomarker which could work globally, one need to do large number of samples analysis or need to do many ways of data analysis to ensure that a given protein or a given candidate biomolecule, could really cater the needs of detection or the therapeutic significance in the clinics.

So, biomarker discovery programs usually depends on a big team, which involves clinicians, technologists, statisticians and many people who are together trying to make meaningful and reproducible data and a cells out of these experiments. I hope today's lecture will give you more insight a nitty gritty detail about how to do biomarker discovery based research. So, let us welcome Dr. Joshua LaBaer for today's lecture. Alright so now, we are going to talk a little bit more specifically about biomarkers alright.

(Refer Slide Time: 01:47)

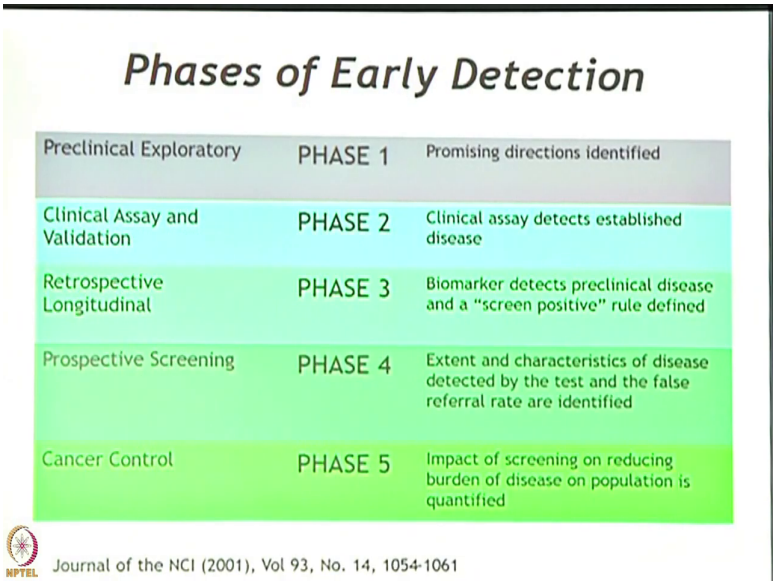


So, I will not go through this part a lot because we kind of did this. So, we talked about why you would do it you want to monitor disease, you might want to monitor whether if they are working properly, you can you might be able to use the markers to predict toxicity of drugs or efficacy of drugs. We talked about the use of at to screen for disease to the early detection or even acute diagnosis. Patient shows up in the hospital with you know crushing substernal pressure and their chest and you want to know is this patient do they just eat some bad food or do they actually have an ongoing a heart attack.

Now, in a blood test would be very useful in that setting and there are a couple of blood tests, but they are still not fast enough. You might need a test to look for a infectious disease, yesterday if you went to the symposium you heard by the need for blood test for tuberculosis. This is an illness that infects a third of the population on our planet. And, it its one of the top

10 killers of all people and yet it is very difficult to diagnose and then you know to personalize treatment of therapy again biomarkers may be helpful for that.

(Refer Slide Time: 03:05)

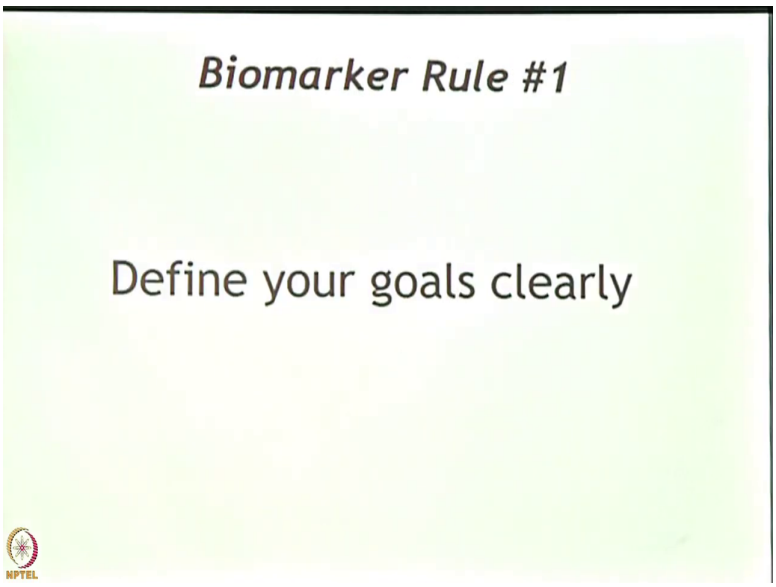


So, all of these are reasons why it want biomarkers. So, this is sort of a different way of saying what I have told you earlier. This was a this is based on a publication from the early detection research network at the National Cancer Institute in the US. This basically outlines if you are going to develop an early detection marker, the phases that you should go through first you should do exploratory studies, this is the kind of observed different study I told you about earlier, then you need to do a clinical assay and validation.

So, you need to establish that the assay can detect the disease, then they would say do a retrospective longitudinal study. So, you look these may be old samples, but you are looking at samples collected over a period of time to ask you know does the marker change when the

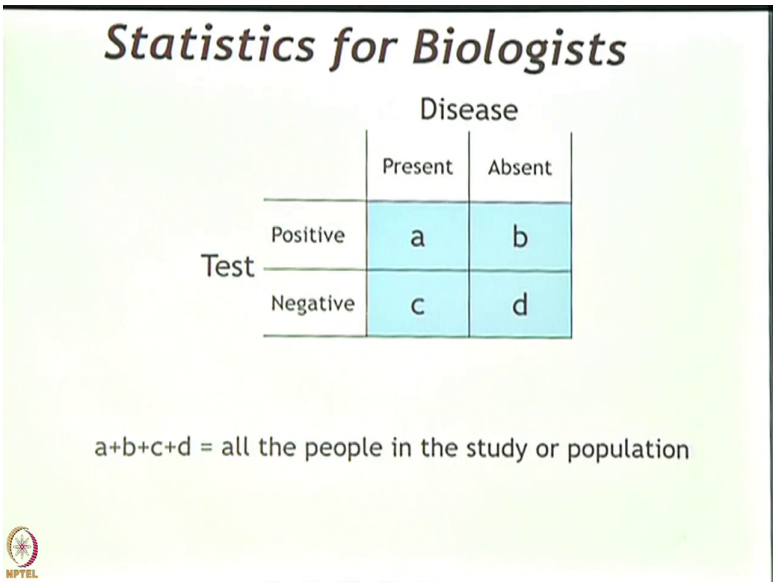
patient goes from no disease to disease. So, that is phase 3. Phase 4 would be to do a prospective study we talked about that earlier collect samples going forward starting today and asking does the marker actually identify those people who are ill. And, then cancer control would be to implement the use of that marker in a large scale screening population ok.

(Refer Slide Time: 04:19)



So, I am going to walk through about 6 or 7 rules for biomarkers and then let us see if we can understand them all. So, the first goal and I told you about this earlier is to define divine your goal clearly. So, what is it that you want to do, why are you making a marker, what do you hope that it will help you accomplish? So, let us diverge now and talk a little bit about the statistics of biomarkers ok.

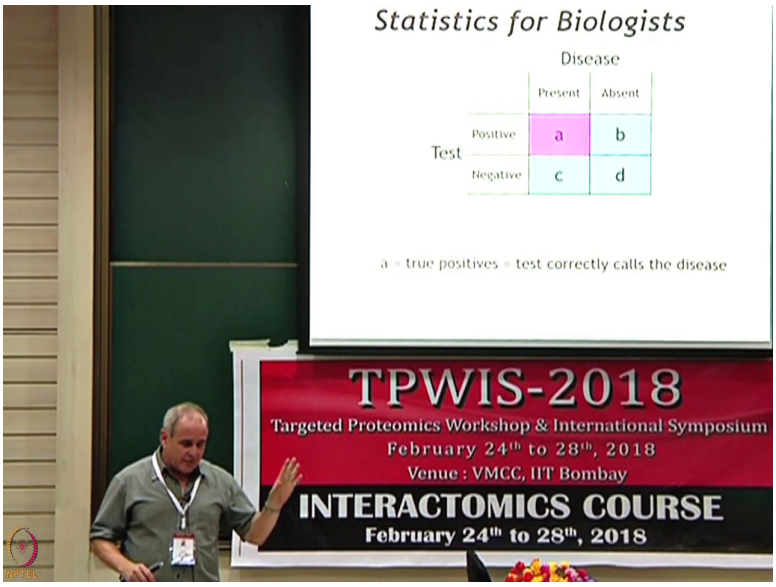
(Refer Slide Time: 04:37)



So, this is obvious right you have a population of people, some of the people have the disease and some people do not right that is true of any population anywhere. You got some people in that population that have it and some people that do not right. And we are going to for the moment now let us assume that this is absolute truth, this is you know truth with you know roman characters; you know this is this is the absolute answer. And, then now we also have a test this is our biomarker right here and our test is designed to predict these two features.

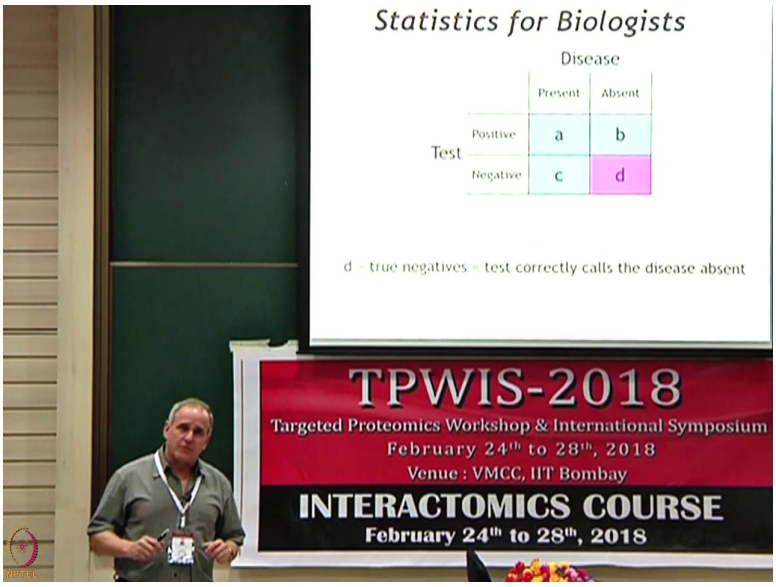
The test can either have a positive result or it can have a negative result. Ideally we want the positive result to tell us when the disease is present and the negative result to tell us when the disease is absent ok. But as you know nothing is ever perfect. So, let us look at the possible cases; the first mathematical thing we know is that a plus b plus c plus d are all the people in the study population. So, this box here is everybody in our study ok.

(Refer Slide Time: 05:51)



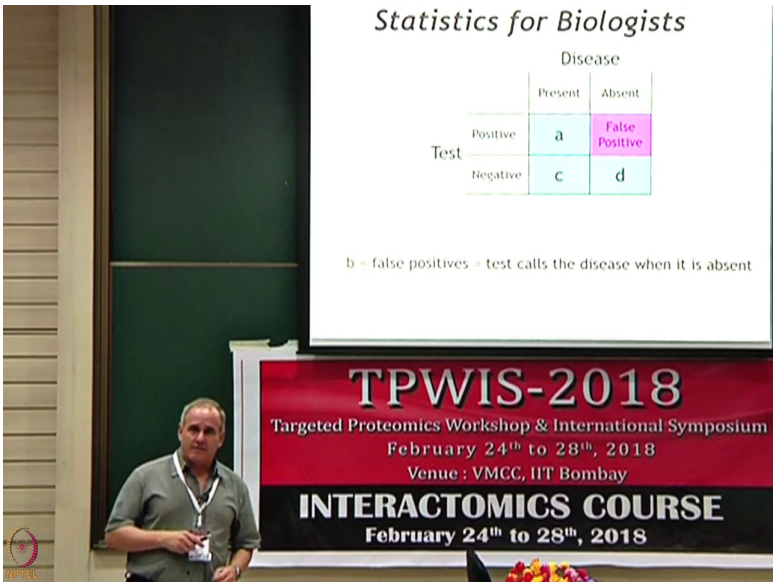
So, first we have got this group over here. So, we call those the true positives; true positives means the test was positive and they actually had a disease. So, the test got it right that is as it should be right ok.

(Refer Slide Time: 06:09)



The second group is this group down here and those are what we would call the true negatives. In this case the test was negative and these people also did not have the disease. So, once again the test was correct. So, this box here and that box there that is when the test is working well, it does what it is supposed to do right. So, that that group of people is a, this group of people is d ok.

(Refer Slide Time: 06:37)



So, what about this? That is a false positive right what is a false positive ok?

Student: The test is positive.

I got a lot of answers over here.

Student: The patients who do not have the disease, but the test result positive.

Yeah ok.

Student: They do not have the disease, but the test is positive.

But the test is positive right the test says they have it, but they do not really have it alright. So, why do we care why do we care is it is it bad to be false positive?

Student: Yes because they will be taking the treatment, if it is a false positive the healthy people will get (Refer Time: 07:18) treatment right.

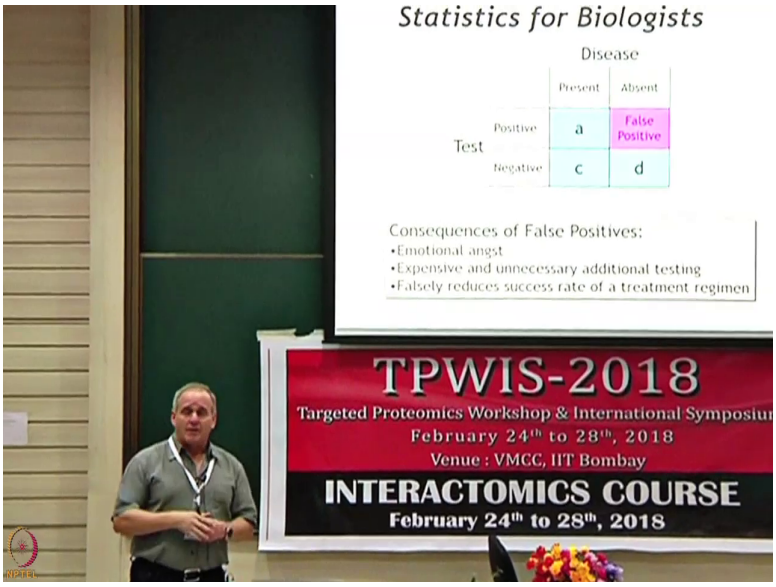
So, you might get inappropriate treatment what else?

Student: They have mental strain that they are having a disease because they think they have the diseases.

Right. So, you are going to particularly if its a disease like cancer, there is a lot of emotional anguish to thinking that you are a cancer patient when you do not really have cancer right. And, then in some cases it is also you put them through needless testing to see if they have a disease and that can be either or both expensive and tiring for patients right.

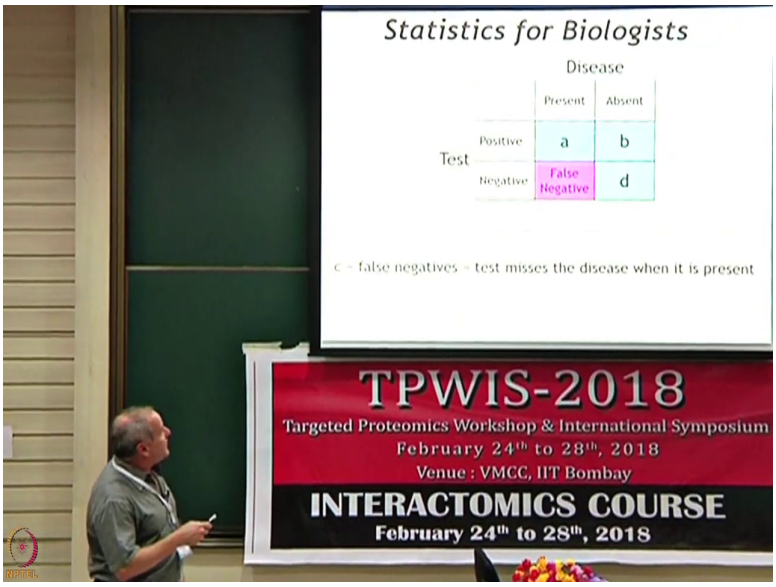
So, the consequences of false positives are as you all point out emotional angst expensive testing and it reduces the success of a treatment regimen this has to do with when you are actually testing your drugs.

(Refer Slide Time: 08:01)



If you are if the marker said that they have the disease, but they did not have the disease and your drug will not cure those people and so, you will you will get inappropriate results ok.

(Refer Slide Time: 08:15)

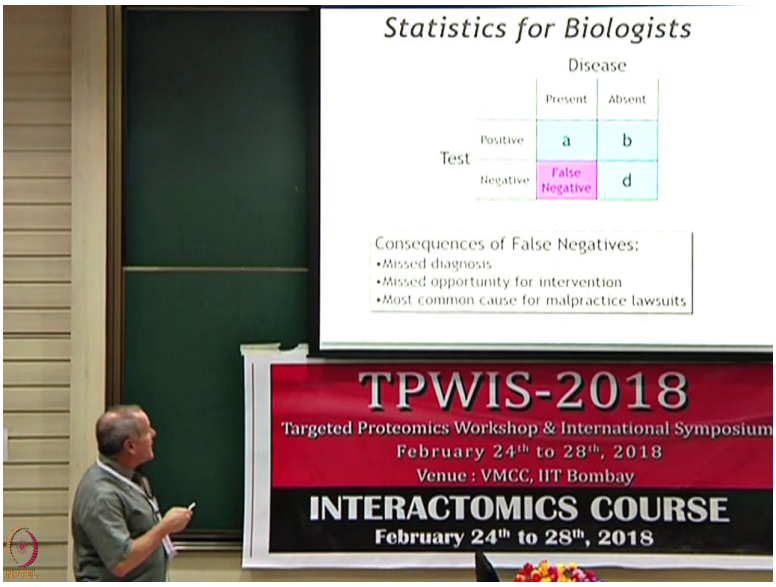


And then this group down here we call those false negatives the test was negative, but. In fact, they really have the disease. So, what is the consequence of a false negative?

Student: If patient dies or its effected (Refer Time: 08:29).

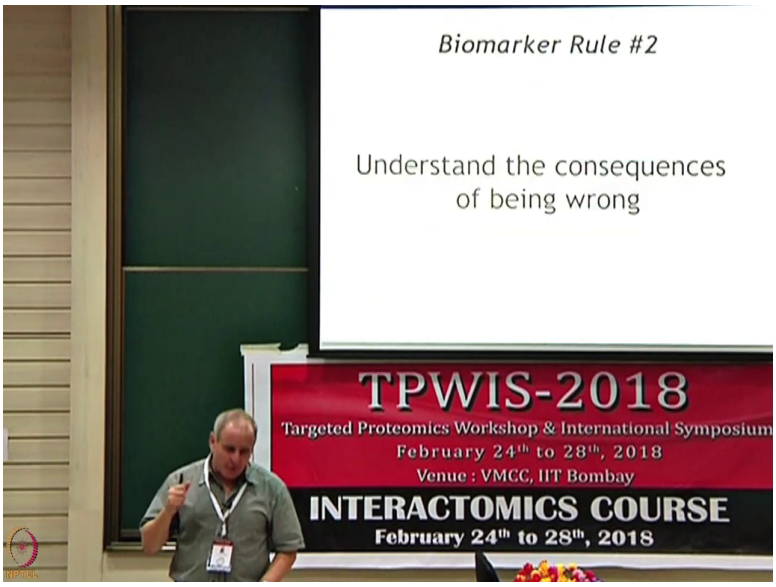
Right that right you missed the disease, the patient is ill you told them you know you know what you are perfectly healthy go about your life do not worry about it and then 6 months later they have the disease right.

(Refer Slide Time: 08:43)



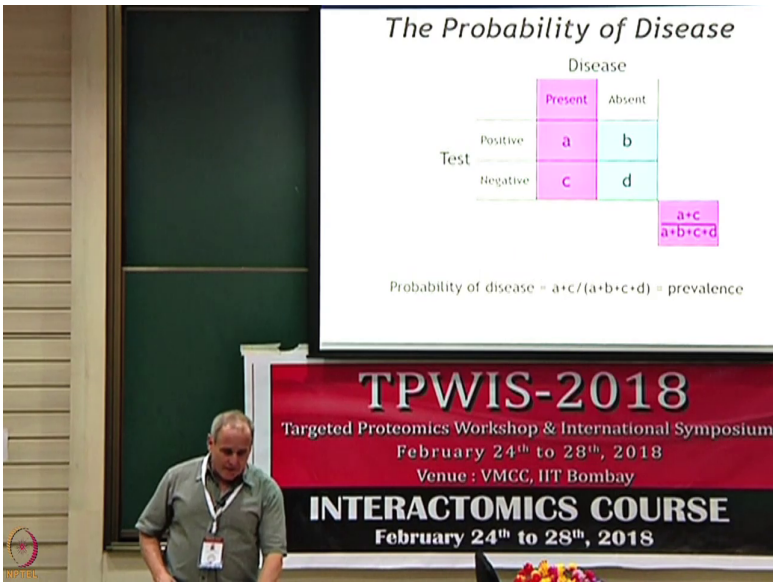
So, this is the misdiagnosis its a missed opportunity for intervention, it is by far the most common cause for malpractice lawsuits in the US. The missed diagnosis of cancer is the biggest cause of huge you know loft lawsuits in the US and so, you do not want to be wrong about this the consequences of a false negative are big.

(Refer Slide Time: 09:05)



So, rule number 2 of biomarkers is understand the consequences of being wrong, you need to know why it is important to have a good biomarker ok.

(Refer Slide Time: 09:17)

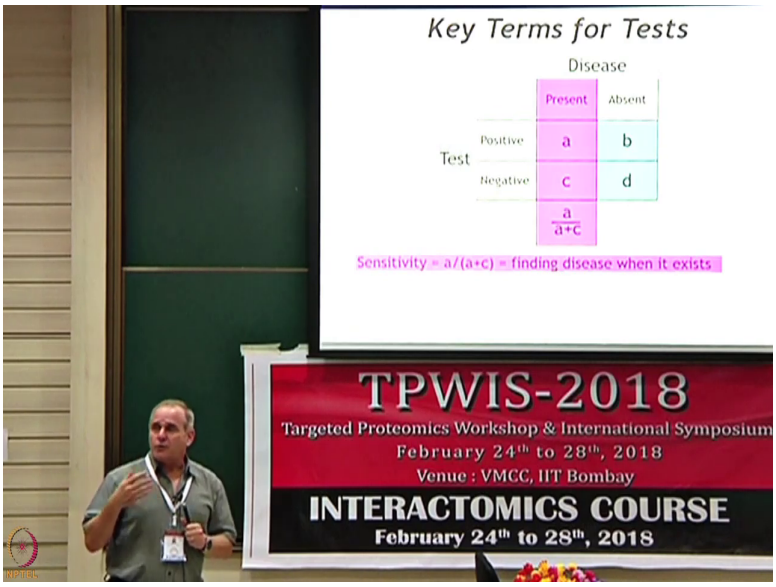


So, now how do we calculate the probability of disease well you take right? So, this is the disease and these are the people that have the disease. So, what is the probability of disease mathematically here?

Student: a plus c divided by a plus b plus c plus d.

Right. So, a and c divided by everybody right. So, that is the probability of disease. So, in your population this will tell you how often the disease occurs ok. Now the next thing we want to talk about is sensitivity ok.

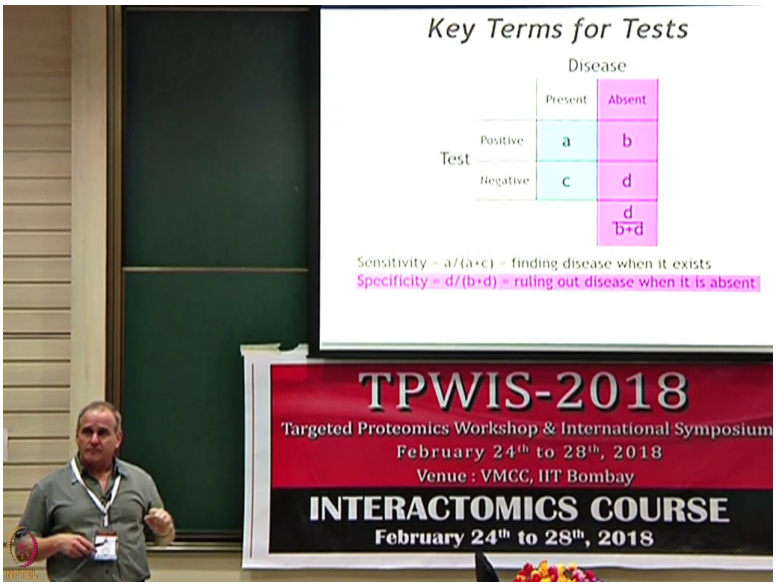
(Refer Slide Time: 09:57)



Sensitivity we define as a positive test in the presence of disease that is sensitivity and in this case mathematically it is a over a plus c. So, you are saying these the denominator is everybody with disease and a is just the people who the tests were positive for. The closer the a is to a plus c right; that means, the smaller the negative the false negative the false negatives the better the test right.

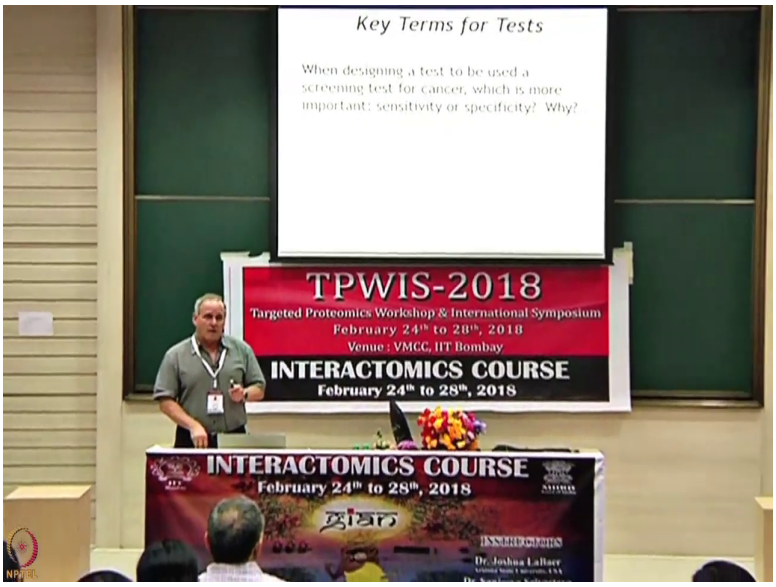
So, that is called sensitivity find disease when it is present. I make all my students memorize this because people often forget this stuff. So, this is a good measurement of how good the test is at finding it when its there ok.

(Refer Slide Time: 10:53)



Specificity is something different. Specificity is ruling out the disease when its not present ok; so, or the way that sensitivity oh yeah. So, this is specificity you are looking at the false we are taking the people who are truly negative divided by all the people who are negative. So, how well is the test how well can you count on the test to be negative when. In fact, there is no disease right in other words how low are the false positives right.

(Refer Slide Time: 11:35)



And so, we measure it by d over b plus d and that that is the equation here. So, its ruling out disease when its absent ok. Let us do a little quiz question; if you are going to design a test to be screened for cancer which is more important sensitivity or specificity.

Student: (Refer Time: 11:44).

I am hearing vaguely sensitivity right that and why is that. Well, I just told you that the biggest causes of malpractice lawsuits is the misdiagnosis of cancer. You do not want to be wrong if you tell someone that they are cancer free and they are not cancer free. So, in the case of cancer detection sensitivity is probably the most important thing.

You are willing to tolerate some false positives if you have to make sure that you do not miss anybody ok. Now, let us talk about a different circumstance. Imagine someone going to a

doctor they are coughing up blood, they have weight loss, they have night sweats right and the doctor appropriately suspects that they might have tuberculosis right that those would be common symptoms.

So, which is more important here sensitivity or specificity hey why?

Student: So, looking on some specific (Refer Time: 12:44).

Raise your hand. So, I know who told ok.

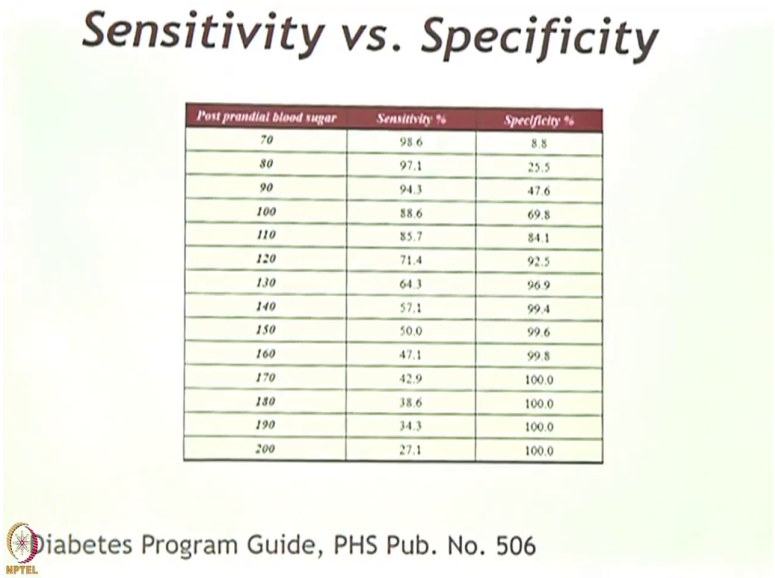
Student: Specifically you are looking always for a tb. So, basically we have three to four types to confirm that (Refer Time: 12:50) have a tuberculosis or not in fact in that manner only specifities would be nature of (Refer Time: 12:59).

Right.

Student: This is a test specific for the.

For the t b yeah; I mean the point is that sensitivity is not an issue here because the patient is right there in front of you already know this person is sick that is not the question anymore. The sickness is already a given what you want to know is it t b or not right you already suspect its t b and here what you are relying on is the test to be very specific to say yes, it really is t b and not some other you know some other illness ok.

(Refer Slide Time: 13:33)

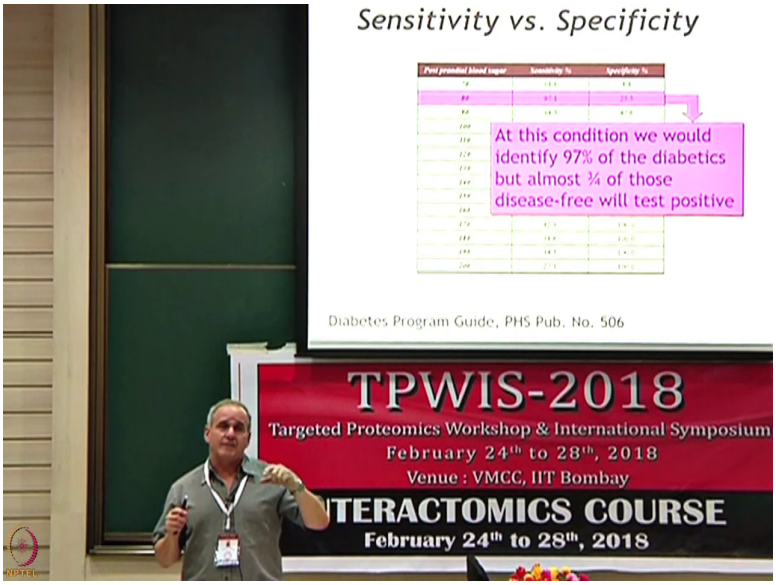


So, now, I am going to show you a little bit about it turns out sensitivity and specificity, in many cases work against each other. Because typically what happens is you have a test for a particular molecule or a typical biomarker, you set a threshold value and you say if its above this value, I am going to say it is positive if its below this value I am going to say its negative alright. And the challenge is that as you elevate or decrease that number, you will alter both the sensitivity and the specificity and oftentimes in opposing ways.

So, I will tell you right now that these are data for a test for diabetes and the idea behind this test was that they were going to measure blood sugar after a meal. It turns out this is a bad test for diabetes and no one uses it you will see why in a minute. But it is a useful test to look at this because it does illustrate the concept a little bit ok.

So, these are the blood sugars after eating a meal ranging from 70 milligrams per decimeter up to 200 milligrams per decimeter and here if you do if you use this value as the cutoff in other words if you say that if you are above a hundred you have diabetes, then this will be your sensitivity and that will be your specificity ok.

(Refer Slide Time: 15:03)

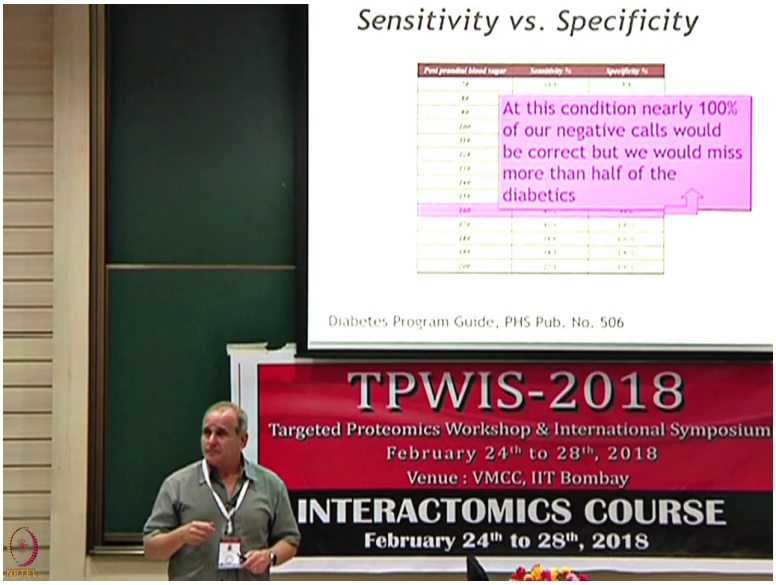


So, let us look at this example here. So, you at 80 if you if use a 80 (Refer Time: 15:07) cut off you are going to be 97 percent sensitive right, but you are going to be only 25 percent specific. So, one goes up the other goes down. So, what that means is, that that you are going to identify 97 percent of the actual diabetics.

The test will be positive in the presence of disease 97 percent of the time. But almost three quarters of the people that that test is disease free will also have diabetes. So, you will not be

very specific I mean we will also test positives. So, three forth people who have no disease will test as if they had to be a huge amount of false positives right ok.

(Refer Slide Time: 15:55)

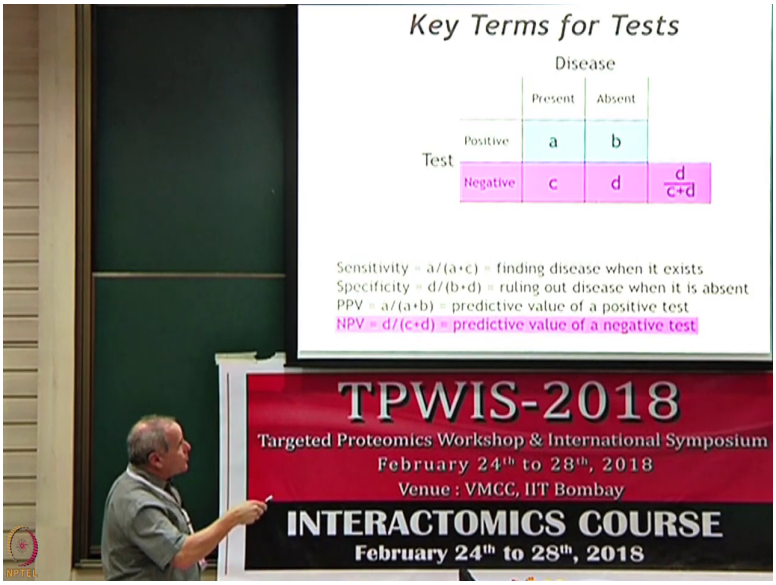


By a comparison let us say well that was too lenient, let us that you allowed too many people in let us set a more strict number let us say its 160 alright. So, now, this sensitivity is 47 percent, but the specificity is 99 percent ok. So, what that means, is that if you make a negative call, if you say that they do not have diabetes you are going to be almost always right. 99 percent of the time you are going to be correct, but you are going to miss half a diabetics you are going to miss out ok. So, you are going to have a lot of false negatives.

And so, that is just to show you that sensitivity specificity often work against each other. Of course, sensitivity and specificity are both values that specifically refer to the test itself that when you go to the doctor that is not what you care about. You do not care how good the test

is, what do you care about. What is happening to me tell me on me, I do not want to know about your test I do not know what how am I doing right. And so, so, so there are two statistical terms we use to describe what is happening to me alright. The first one is that the positive predictive value ok.

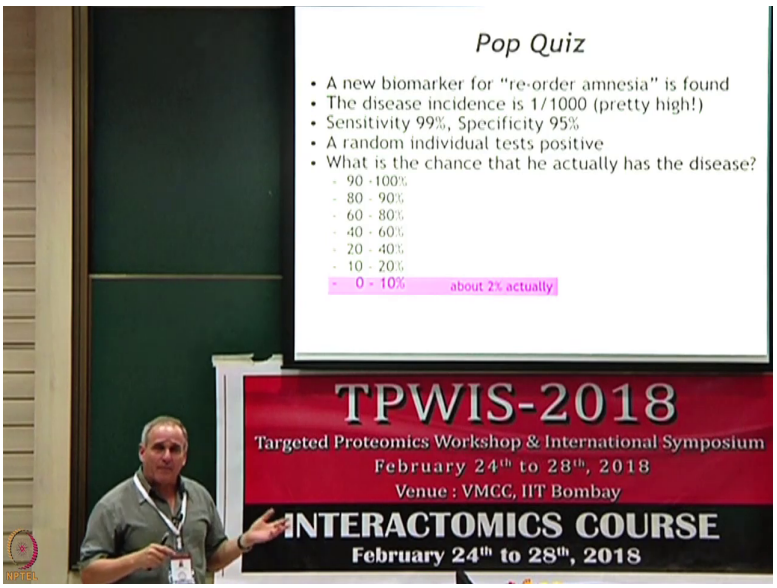
(Refer Slide Time: 17:07)



So, what do I mean by the positive predictive value? The positive predictive value is, if the test is positive what is the chance that I have the disease. So, the test says I have it do I really have it right. And so, to mathematically calculate that that is shown here, its basically taking all the people who actually have the disease divided by all the people who were tested as having the disease and that is the predictive value of the positive test right and that that matters a lot to patients. Sometimes this other value matters even more this is the what we call the negative predictive value.

So, if you had a test we did a test for cancer or we did a test for birth defects in your child, how sure are we that you do not have cancer or you do not have your child does not have birth defects right. So, what how confident is a negative value in telling you that you are disease free. And that is defined as taking all the people who are truly negative divided by all the people who are tested as negative. So, positive predictive value and negative predictive value this is what doctor's care about this is what patients care about what is happening to me how am I doing ok.

(Refer Slide Time: 18:41)



So, where we are going to do a little quiz now ok. So, this is a quiz for in this case a test called reorder amnesia, the disease is occurs in one thousand people. So, its a pretty common disease. The sensitivity of our test is 99 percent and the specificity is 95 percent ok. We test a random individual for the disease what is the chance that he actually has the disease got it.

Sensitivity 95 99 specificity is 95. So, how many people think that there is an 80 to 90 percent chance that he has the disease ok?

Student: (Refer Time: 19:28).

I have got one of those how many people think its 60 to 80 percent chance that he actually has a disease? How about 40 to 60?

People raise your hands some answer at least.

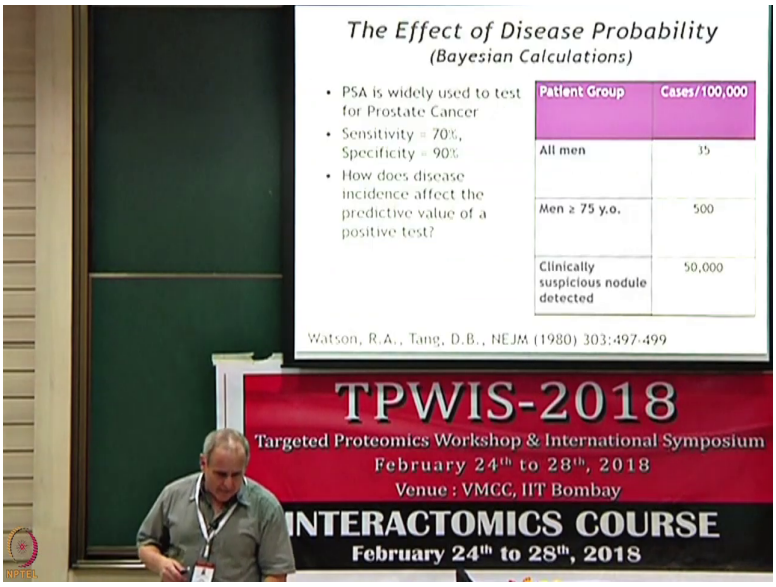
I am going to assume you got it all wrong if you did not get it. 20 to 40 percent I got one 24 40; so, far how about 10 to 20 about 0 to 10.

Yes (Refer Time: 19:55) answers.

Got a few of those the rest of you all think that its 90 to a 100 I only think its nine 100 got a few 90 to a 100 alright its about 2 percent. Yeah its about 2 percent right because the remember what effect you hear is the is the incidence of the disease, its very low and that is that it turns out that this is an important thing to remember about these statistics and let me go back a second and point that out. Remember that sensitivity and specificity were down in these columns here right. They those terms do not depend on the population it does not matter how often the disease occurs for them, they strictly measure the value of the test on whatever specific population they are being tested on.

But positive predictive value and negative predictive value they depend on how often the disease occurs and I am going to walk you through that in a minute, but its really important to remember that. When you hear somebody boast about the positive predictive value of a test, the first thing you need to ask was what population did you test how prevalent was the disease in that population ok. So, let us walk through that.

(Refer Slide Time: 21:19)



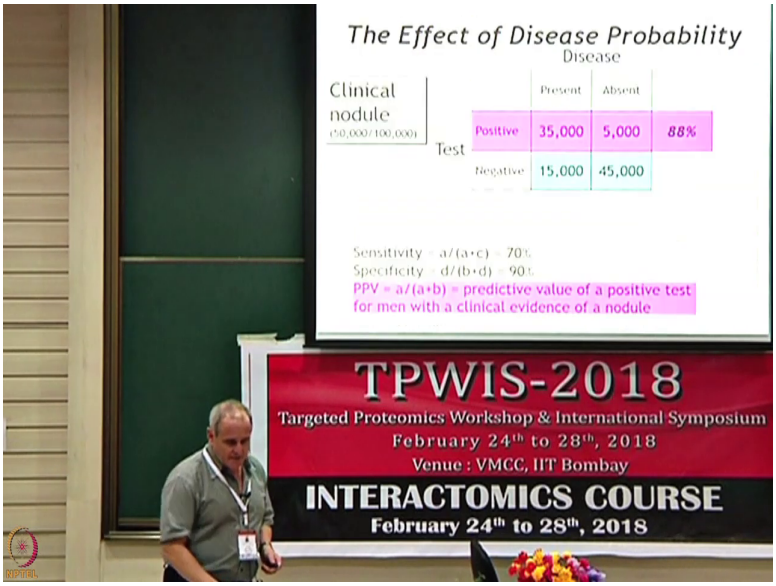
So, this is this has to do with what is called Bayesian calculations, which includes looking not only at the probability, but also at what is called the prior probability which is when you begin your test what was the likelihood to start with. And, we are going to use as an example the prostate specific antigen test is a very common test used to detect prostate cancer, it has a sensitivity of around 70 percent and a specificity of 90 percent that is one of the best you will see anywhere.

You know that is a pretty typical marker when people prove when I told you before, but people published 99 percent 9 9 percent you do not believe it numbers like this that is kind of what you would expect from a pretty good marker. So, now, we are going to ask the question how does incidence or prevalence affect the positive predictive value of a test. We are going to consider three different populations, we are going to consider all men in which case the incidence of prostate cancer is 35 cases in a 1, 00,000 we are going to consider men who are

over 75 in which case the prevalence of the disease goes up to 500 400000 and then we are going to consider men who already have a clinically suspicious nodule a doctor did an exam and found a mass.

So, that in that case there is a about a 50 percent chance that they have cancer ok. So, three different populations these are the incidents remember I told you the probability of disease a plus c over a plus b plus c plus d that is what these numbers are right here ok. So, let us look at the first case.

(Refer Slide Time: 22:57)

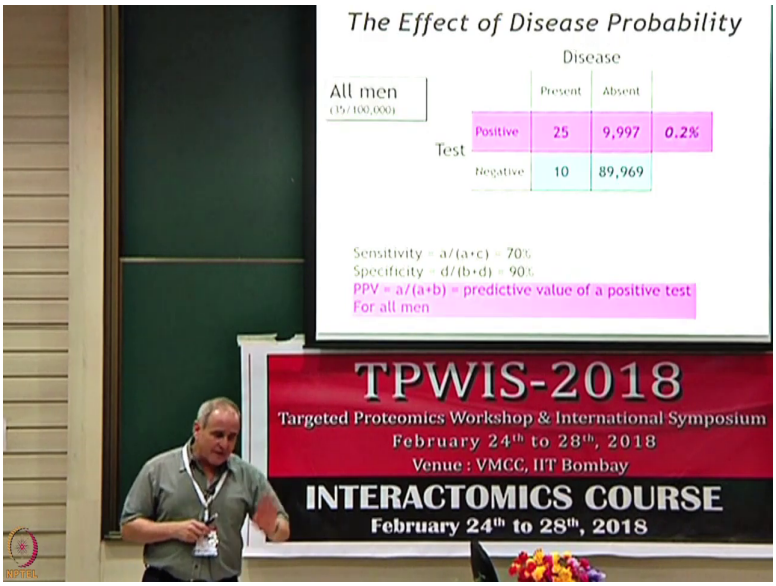


In this case we are looking at the clinical nodule a 50 percent likelihood to start that this person has cancer right. So, you so, notice that I have that this number here and that number there add up to 50,000 right. So, 50,000 remember I said that out of a 100000 men 50,000 had it. So, 50,000 have it and 50,000 do not.

So, that is appropriate right remember I said that it has a 70 percent sensitivity so; that means, of this number here 70 percent or 35,000 are positive and I remember I said that it had a 90 percent specificity. So, this number 45,000 do not have it right. So, these numbers all add up to these numbers here; you believe me? So, now, do the math if you do the math the positive Richter value is 88 percent.

So, even though you already have a suspected mass and even though this test has a 70 percent specificity to 95 percent sensitivity specificity, the predictive value is still not 100 percent its still about 88 percent ok.

(Refer Slide Time: 24:11)



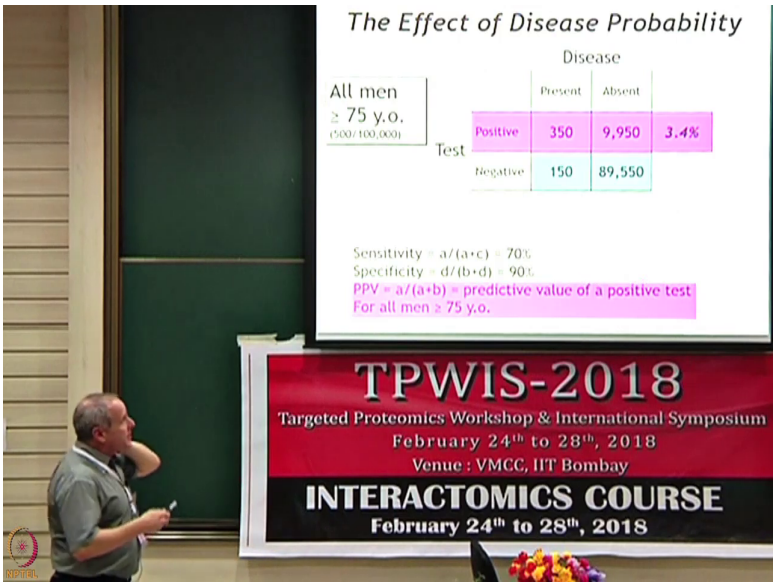
Now, let us look at a very different population, we will go to the other end of the spectrum let us look for men who all men 30 35 men and a 100000 that have the disease. So, now, let us

do the math again the population that has the disease is 35 the population that does not is everybody else right.

Out of a 100000 still we have a 70 percent specificity here and here we still have a 95 a 99 percent specificity look at the how good that test is 0.2 percent 0.2 percent. So, the take home message here is that depending on the population, the positive identity positive predictive value changes dramatically we did not change these numbers at all. Those numbers stayed the same throughout the whole discussion the only thing that we changed was how often the disease occurs and if the disease is rare then the predictive value of the test drops quite a bit.

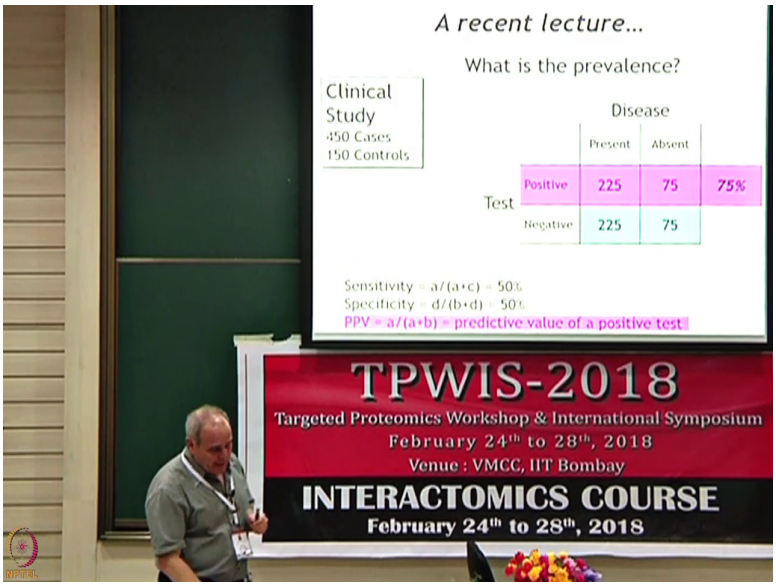
This is one of the reasons why at least in the US we do not recommend that young men do treadmill tests for heart disease because the treadmill test was designed for you know older men, where it has good predictive value. But what you know when the incidence of the disease drops like it does here, then the predictive value drops precipitously and then the risk of a false positive becomes much higher ok.

(Refer Slide Time: 25:41)



And then this is just a shorter sort of show you the kind of more general circumstance of 500 in a in a 100000. So, this is not far from the what you know one in a 1000 we looked at in that quiz question and again here the test is around 3.4 percent. So, it all has to do with the population you are dealing with ok.

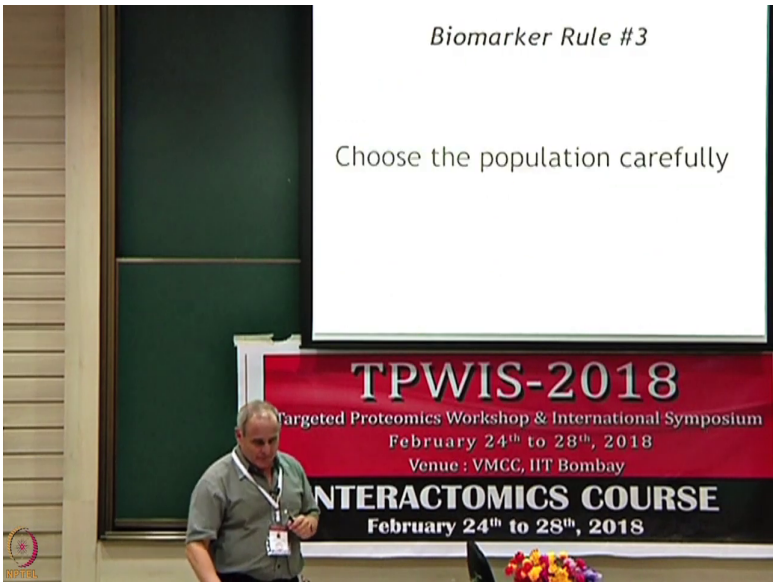
(Refer Slide Time: 26:01)



So, I in my institute, where someone was boasting about his test that he had developed and he hit this is the clinical study he did; 450 cases in 150 controls. So, the prevalence in this population is what? So, the prevalence is very high right because you are you are three forth of the people in your study have the disease, three forth of them have it right. So, he did that he had this positive test and he said that his predictive value was 75 percent right.

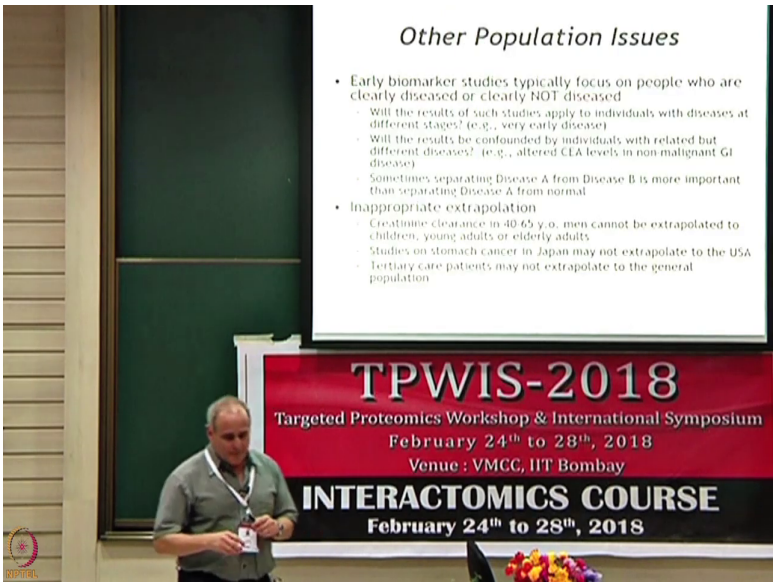
And, I looked at the numbers here and it turns out that if he had 0 if the tests were equally split between positive and negative right, he would have liked it half the time its positive and a half the time its negative, he would have still had a predictive value of 75 percent. So, he had to do nothing the tests had to had zero predictive value in a sense and it would still have given him a positive predictive value of 75 percent.

(Refer Slide Time: 27:09)



So, is pretty lame presentation ok. So, rule number 3 choose your population carefully right alright.

(Refer Slide Time: 27:19)



So, if you are going to do an early biomarker study and make sure you pick people who have early stage disease because that is when you want to get the disease will it apply you know if the test will apply to people with different stages of disease, if it could be confounded by people with different diseases maybe they have other things that could alter their CA levels or have nonmalignant GI disease, and just remember that sometimes its a more important to separate disease A from B, then disease A from normal.

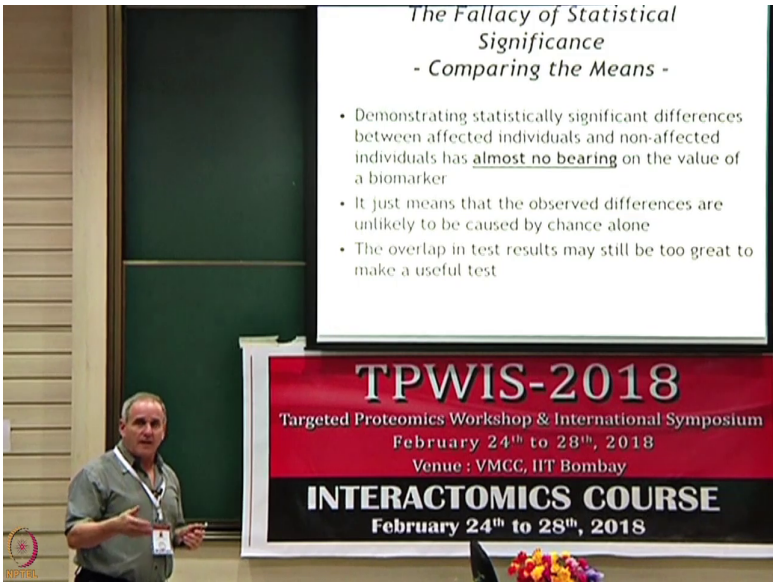
So, imagine if you are in a clinic and someone walks into your clinic and they have abdominal pain and they tell you that they have had abdominal pain for months and you know they have been losing weight right. In that case you are not necessarily interested in distinguishing colon cancer from healthy people. You might be more interested in

distinguishing colon cancer from inflammatory bowel disease. You know the patient is ill they have been suffering from GI symptoms for months.

So, they are you know there is something wrong, you are not separating normal from cancer you are separating cancer from other GI diseases. And so, always remember that if you are going to do a study to find a biomarker, you should find you should use a population of maybe people with non-cancer GI diseases from GI from cancer diseases.

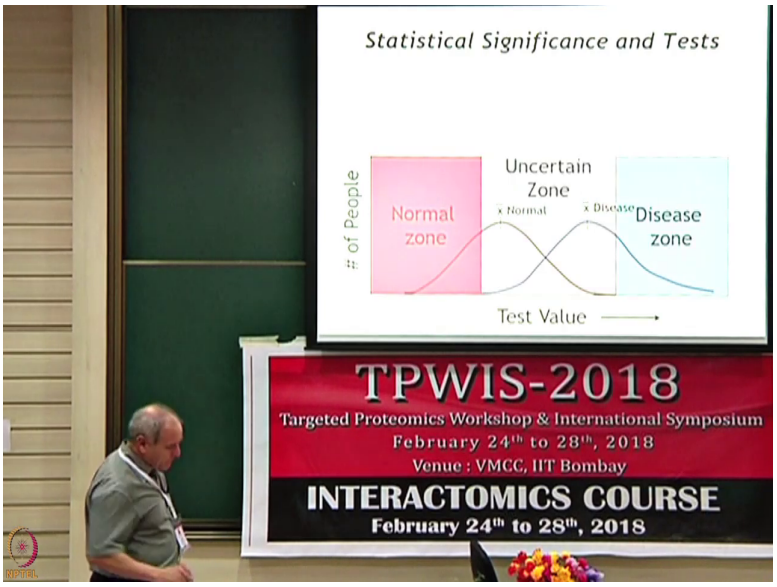
You need to make sure you do not extrapolate inappropriately if you if you develop a test that is good in one population, it might not work in another population if for example, their kidneys do not work as well in older people, if it is something that is excreted by the kidneys, the tests may work in a 20 year old, it may not work in the 60 year old. Diseases on stomach cancer for example, do not extrapolate to the USA, the risk factor for stomach cancer are much higher there that population is different and of course, cancer patients in the hospital are different from healthy people ok.

(Refer Slide Time: 29:23)



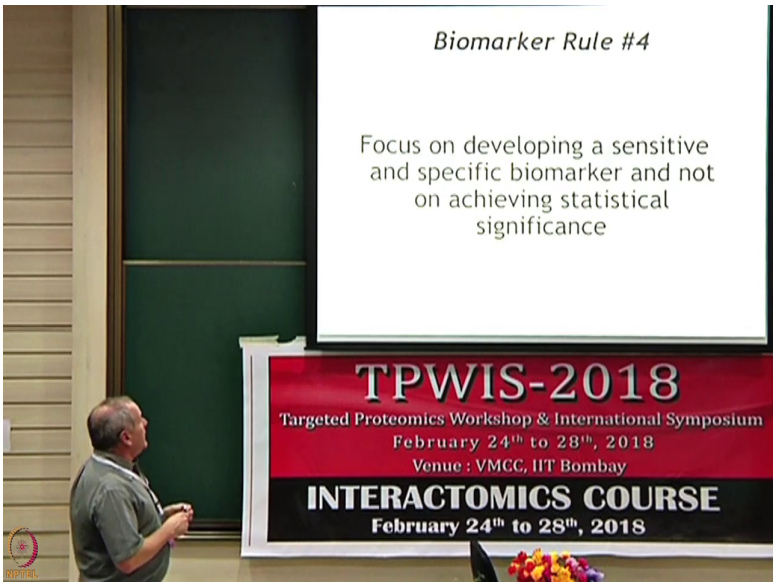
And, then this is something that we talked about a little bit earlier already, this is what I call the fallacy of statistical significance. And so, we kind of covered that already just because there is a good p value between a and b does not mean that they are good biomarkers you should be using sensitivity and specificity not p values and that is really shown on this thing here which we have already covered.

(Refer Slide Time: 29:43)



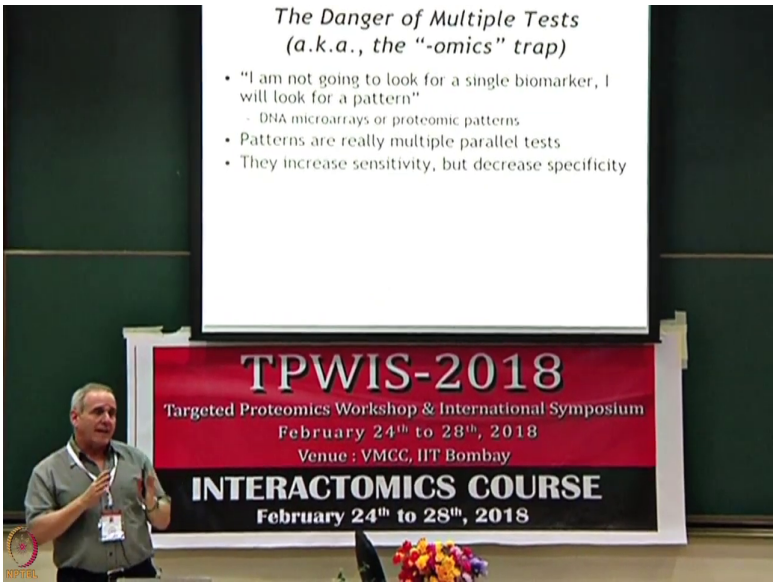
So, I am going to skip that alright.

(Refer Slide Time: 29:51)



Alright. So, focus on sensitivity specificity markers and not on statistical significance alright that is fair alright.

(Refer Slide Time: 29:57)

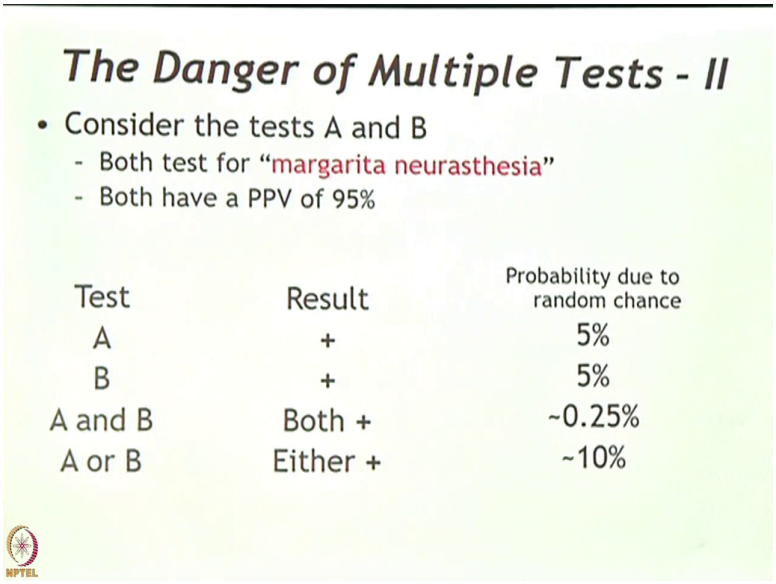


So, now I want to mention a little bit the and we are coming to the end here, the omics trap because all of you are many of you are going to be doing omics studies that is what we all do these days. And you often hear this statement from people in the omics studies. I am not going to look for a biomarker; I am going to look for a pattern I am going to look for a signature.

They might be doing it on DNA microarrays or protein arrays, but you have to remember that a pattern is really multiple parallel tests. They are doing a bunch of different molecular statistical studies and they by doing multiple tests, they increase your sensitivity because each test has a chance of being positive, but they reduce your specificity because you have a higher now rate of false positives right.

So, if you are going to do multiple parallel tests or look for patterns, my biggest advice is to get a statistician because you are going to need more careful statistics and we this class is not prepared we are not going to do those statistics here, you just need to be aware that when you get to that stage its time to engage somebody. So, we have two tests imagine that that this test.

(Refer Slide Time: 31:19)

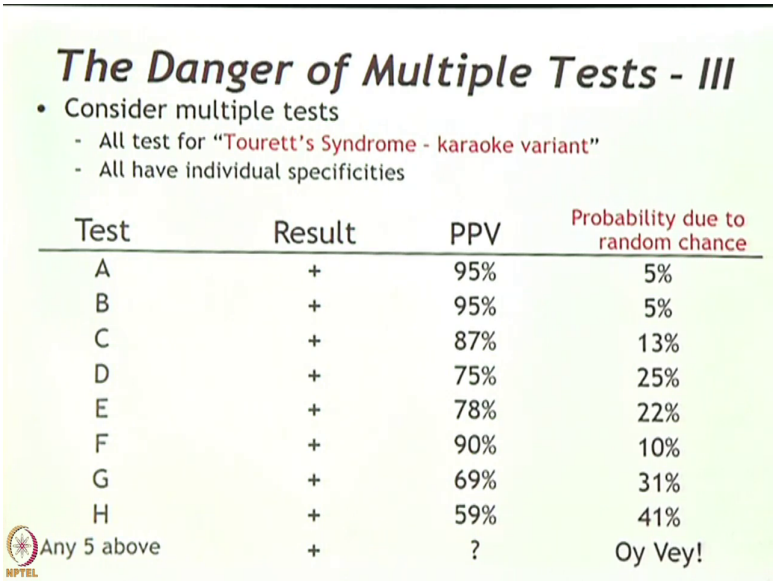


They have two tests for the same illness and they are testing for a positive and they both have a positive particular value of 95 percent. So, imagine test A is positive and it has a probability the probability that it is going to be positive is 5 percent, test B might be positive. So, its chance to chance alone is 5 percent.

If you do both A and B now if you require them both to be positive now you reduce now your test is getting more stringent because the chance of a false positive is much lower now. But, if

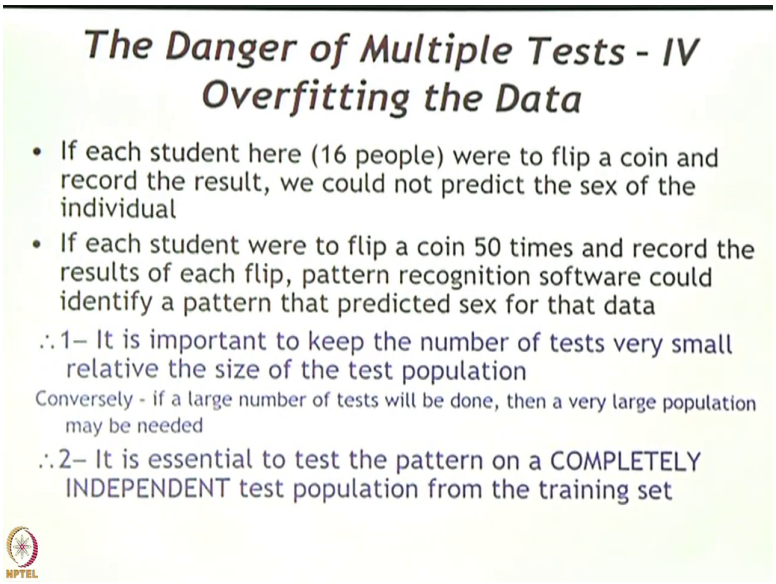
you accept either one now the chance is much higher because you now have to add the two effects together ok.

(Refer Slide Time: 32:03)



Yeah. So, this is even now imagine if you do this with multiple tests. So, now, you have a whole series of tests. So, I am going to just go.

(Refer Slide Time: 32:27)



So, now each of these is going to have a different positive predictive value, they are going to have all kinds of different due to random chances and if you add them all up you numbers get to be outrageous. So, again the take home message is get a biostatistician. So, here is the example that, that I like to remind people of when they are doing multiple testing and this is a lot like what you would see in an omic study.

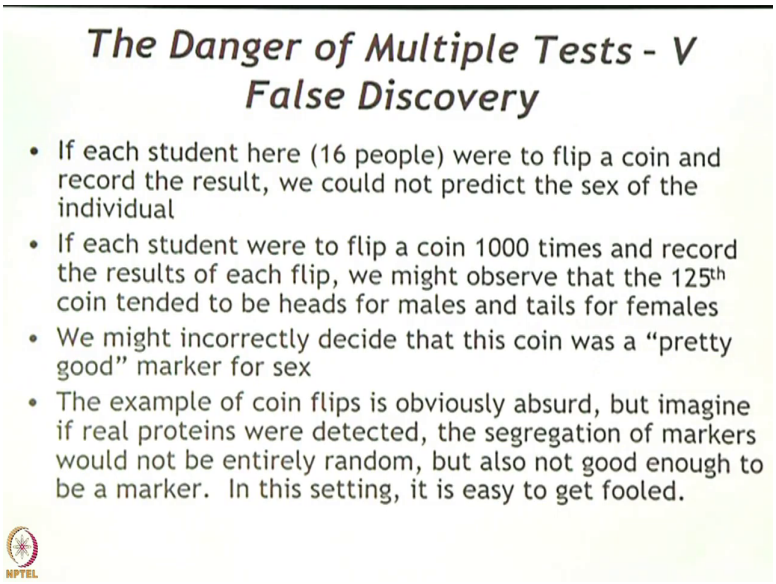
So, if I asked every one of you to take out a coin and flip it, and mark down whether you got a heads or a tails. What do you think the likelihood is that the result on the coin would predict the gender the individual who flipped the coin right nothing right ok. Now let me change that let us say that I gave you each 10,000 coins to flip and you are one by one flip every coin and you mark down heads or tails, what is the chance that among those 10,000 flips that one of them maybe the 5635th of them would correlate with sex of the individual.

There is a chance right might not be perfect, but among those 10,000 tries maybe one of them by chance alone would align maybe not perfectly, but it would align with the gender the individual. And you would say aha I found a biomarker if you if you flip a coin 5635 times that one will predict the sex of the individual.

But you would be wrong. So, how would you prove that you would be wrong. You repeat the study, you do it a second time 10,000 right and now the 506 are three that does not work anymore now its the 123rd right. It its just random chance some of them will happen to work and that is what we do with omic studies we test 10,000 things, we get one that works and we say aha I found a biomarker, but you tried 10,000 times. So, you have to you have to adjust for that by doing some kind of false discovery adjustments.

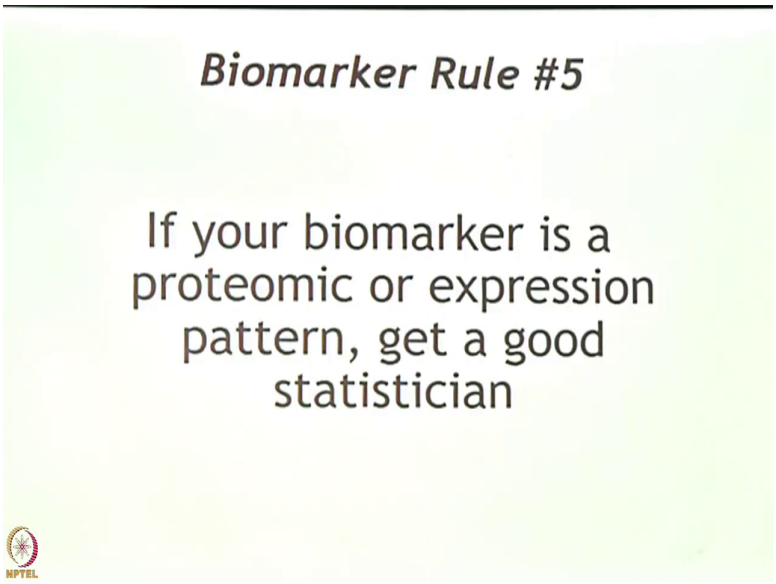
So, that is kind of what I did here. So, imagine if people did all these studies right you have to keep this you have to keep the numbers the populations small, this is especially a problem when the size of your population is small relative to the number of variables you are trying. If you are, if you have a study of a 100 individuals, 50 cases and 50 controls, but you are testing 10,000 variables you have this risk of what is called over fitting and then that is why if they repeat the study doing a completely different population alright.

(Refer Slide Time: 35:03)



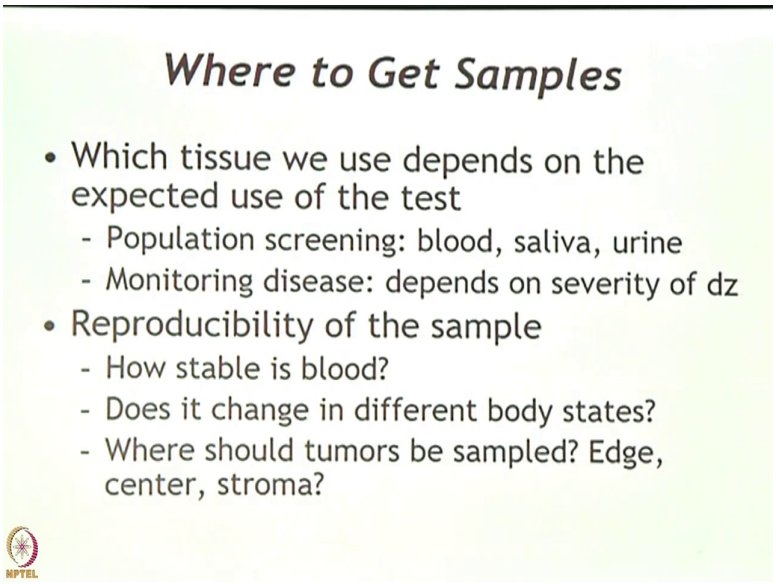
So, I kind of went through this alright.

(Refer Slide Time: 35:07)



So, if your biomarkers of proteomic or expression pattern, the bottom line is get a good statistician ok.

(Refer Slide Time: 35:13)



The last couple things I am going to mention is where do you get your samples make sure that the sample that you use is relevant for the use of the test. So, imagine if you are going to do a biomarker on early disease detection right, we said that you are really going to be testing a healthy population. Healthy people are not going to be interested in giving you biopsies nor would it be appropriate to put them through that risk right.

You know if you are going to take a test for healthy people, it should be a simple test urine may be blood you know it is got to be something that you can measure easily maybe saliva, you cannot rely on doing biopsies. On the other hand if they already have cancer then of course, they might be willing to do that.

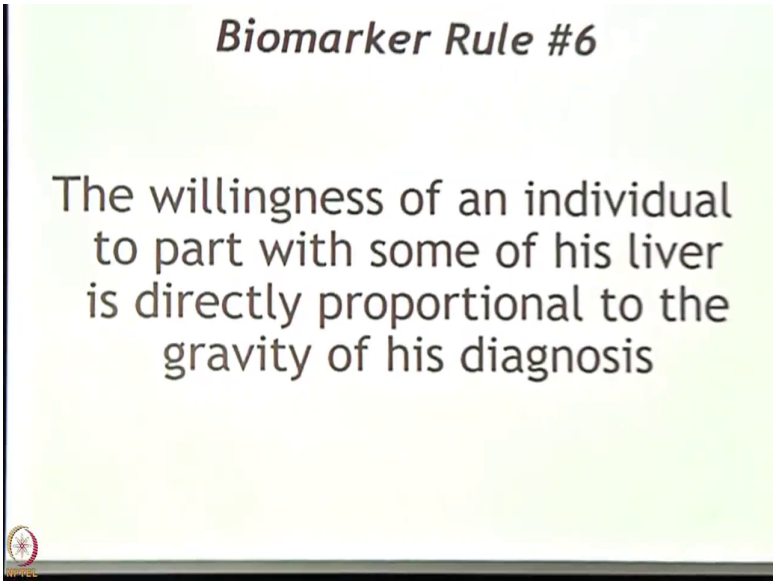
If you are you have to look at whether this sample will be stable, if it is a biomarker and blood will it be stable in blood part of the remember I told you about Paul Temps and the

study where he could tell the difference between the tubes, well what it turns out is that one of the tube types was inhibiting a protease and the other one was not and what was causing the difference was proteolysis in the sample.

So, in that case the material was not stable. So, you need to know that what you are measuring is stable in your in your samples. You need to know if it changes in body states if that molecule goes up and down after a meal if it goes up and down with a sleep cycle again that is something that you have to consider. And then of course, if you are measuring samples from a tumor you need to look at where your you are taking your biopsy from.

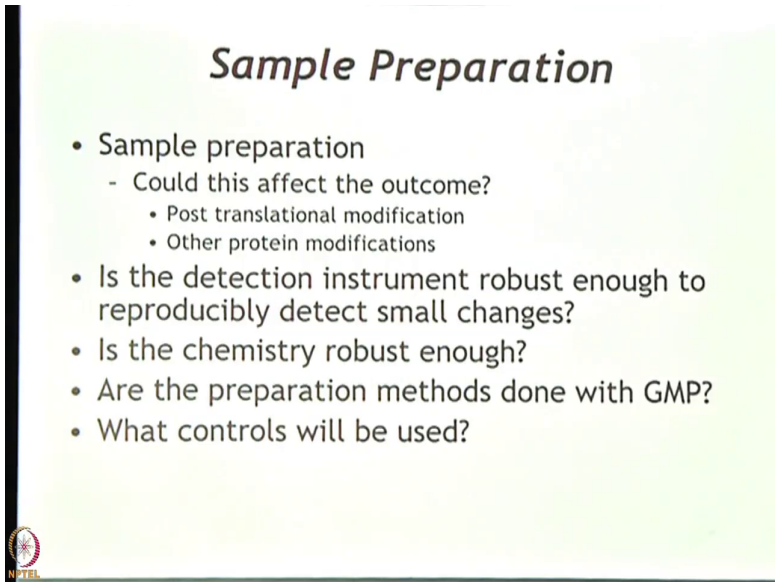
So rule number 6 is the willings of an individual department some of his liver is directly proportional to the gravity of his diagnosis.

(Refer Slide Time: 36:55)



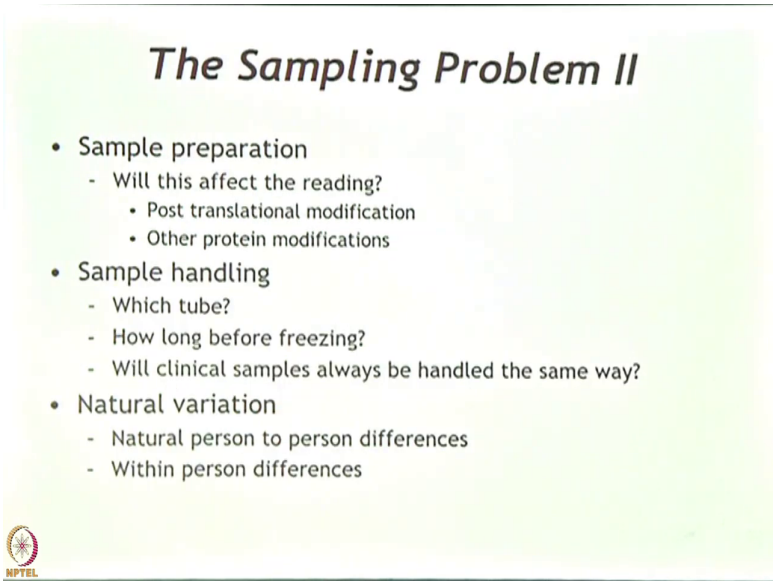
People do not give up parts of themselves easily they only do. So, when they are really sick so, that is good to remember that ok.

(Refer Slide Time: 37:01)



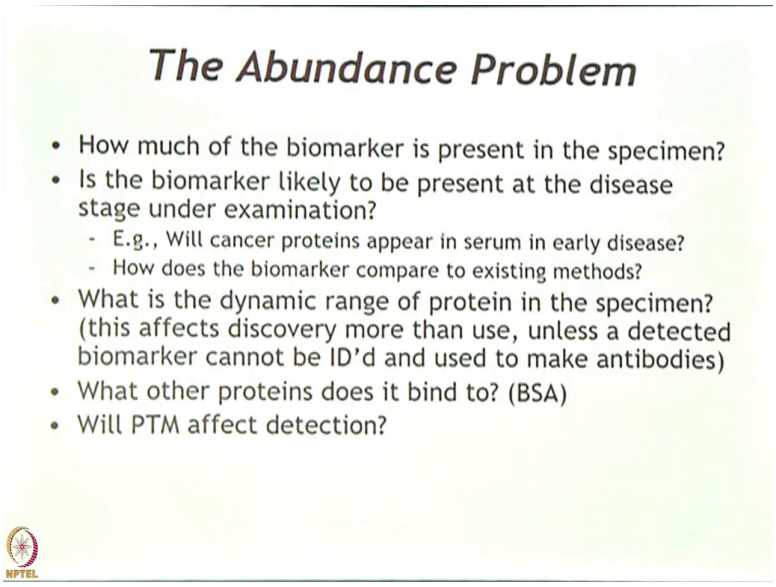
Part of biomarkers is knowing how to prepare your samples. How are you going to preserve it, that could dramatically affect outcome I already gave you the example with Paul Tempest is the instrument robust and reliable is it going to give you the same answer every time you measure. It is the chemistry robust well if you shift this sample to a hospital far away will they get the same answer that your hospital gets here and then you of course, what controls you are going to use.

(Refer Slide Time: 37:33)



So, these are just some of the general things to think about. So, we will sample preparation affect the reading are you handling the samples properly are you going to freeze them. And then of course, you need to know if they are natural variations of the biomarker you are testing from person to person because that is going to that if there is a, if there is a lot of natural variation, even among normals that is going to make it more difficult to use that as a biomarker.

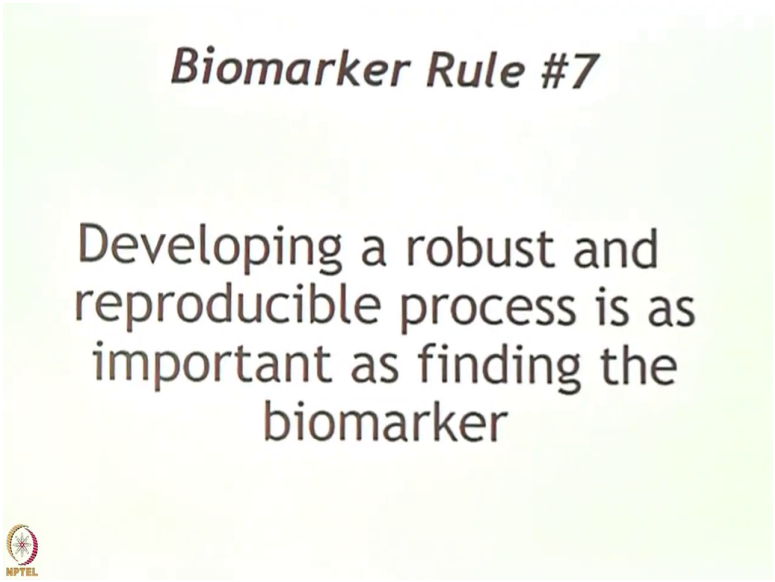
(Refer Slide Time: 37:59)



And then there is this question of abundance of the biomarker is there enough of it in the sample that you can measure it, is it likely to is it will you be able to detect it when you want to detect it. So, in the case of early biomarkers, earlier textured biomarkers is there going to be enough there in an early specimen from people with early disease that you can actually detect it.

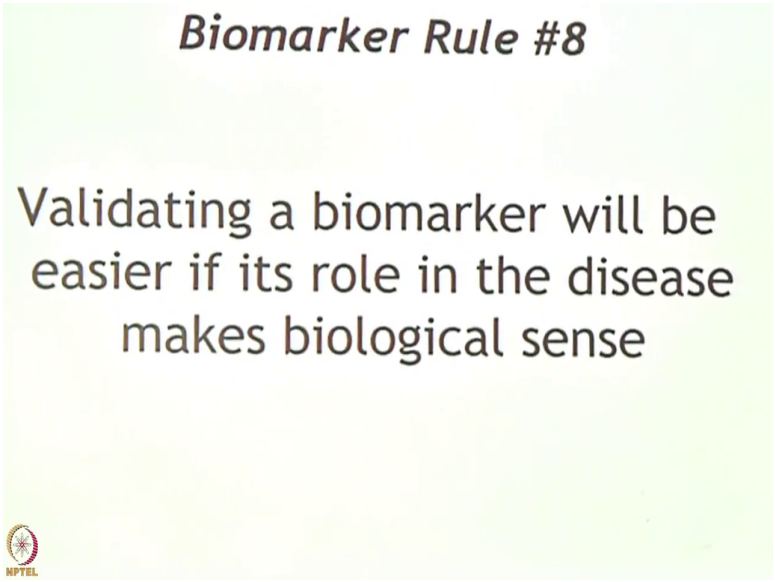
So, the marker may be very good at picking up cancer, but it may be too weak to be able to pick it in early disease that was the case with the CA 125 that I mentioned earlier. It was a good marker for distinguishing ovarian cancer it is just not abundant enough and early disease to pick it up developing a robust reliable test is half the game. Just because you have found a molecule that looks good does not mean that you have got a biomarker what you need now is to develop it into an actual diagnostic test.

(Refer Slide Time: 38:53)



And then the last thing I am going to mention is this one which is that your markers are likely to be more believable, if they relate to the biology of the disease.

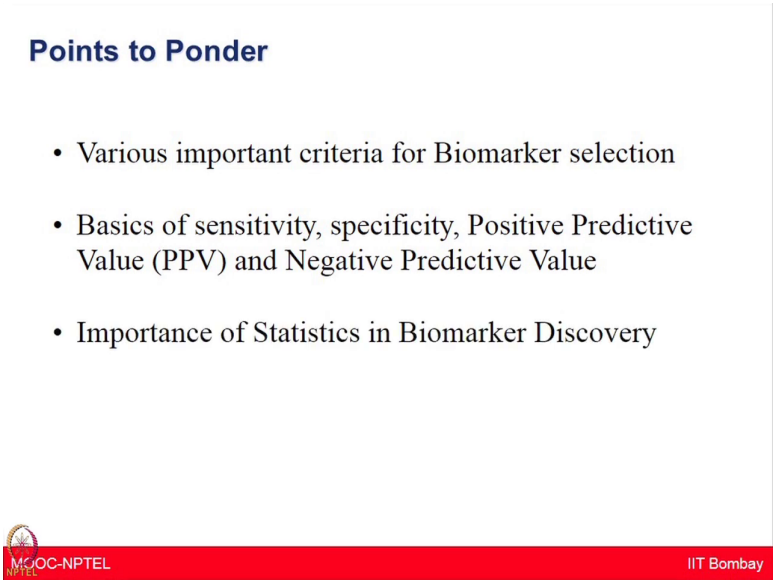
(Refer Slide Time: 38:57)



And, I think a couple of you have already mentioned that, but just keep that in mind that that if you want the marker to make sense look for markers that that fit with what you think is going on the disease if it is a sort of a random molecule it will be a lot harder to validate it.

So, I will stop there.

(Refer Slide Time: 39:23)



Alright. So, by now you are quite familiar with the importance of studying biomarkers and you have also seen the challenges of performing the experiment to discover new biomarkers. And, I must say that you note is very very challenging journey and that is why actual clinical translation of biomarkers it is not easy and not very successful. Either the lot of candidate proteins have been discovered which have potential for the biomarkers especially for early detection of disease or prognostic values or even therapeutic values.

However, many of the biomarkers are not easily translatable to the clinics reason that you know you need to do lot of validation to ensure that from the discovery work what the biomarkers have been identified, they really fit the purpose of the clinical assays and they are able to serve the utility for the large patient populations.

Therefore, the biomarker discovery program even if it is performed on the small number of samples, you need to now scale up to the really large number of samples to do the validation that these proteins are actually showing the kind of expression pattern which you have discovered from the initial workflow.

You have also learned the need to have a good team involving clinicians, who can give the right samples to test your hypothesis, the right type of technology platforms, where you can execute these experiments and then involve the scientists who are good in doing the big data analysis who can now make reproducible and sense of your data without compromising on the data quality.

So, these are the considerations which are very crucial and I must say that despite all the odds despite all the challenges, this is there many bio markets which are now getting translated to the clinics, then getting approved by the USFDA. And, there are some success stories especially the ova 1 and ova 4 and so, the other protein which are now coming to the markets giving you the motivation that if we do these kind of discovery work flow properly, probably the eventually it may be translatable to the clinics.

Thank you very much.