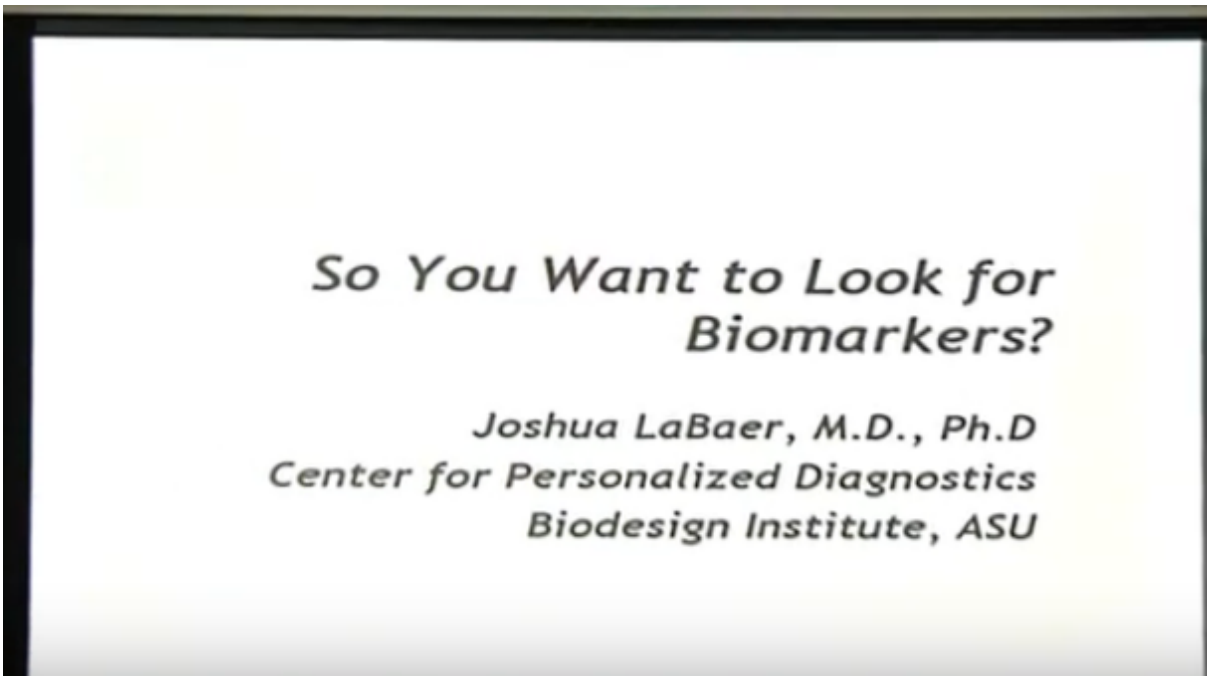
Lecture – 06

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

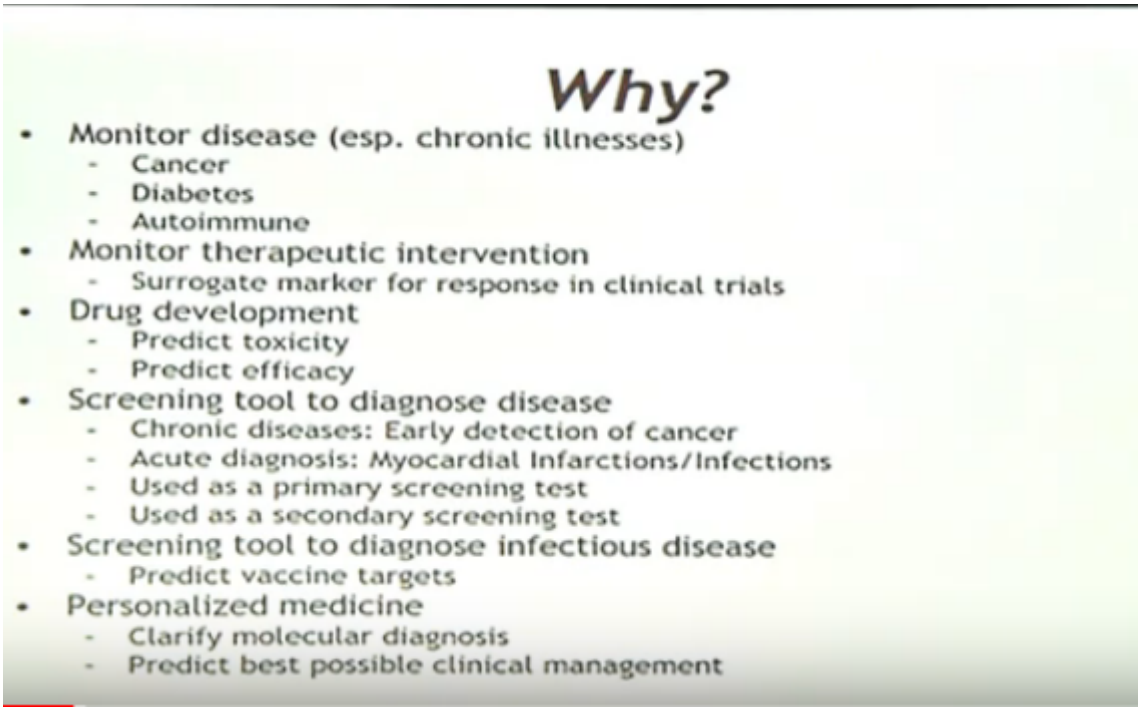
Welcome to MU codes, on applications of in Tractomics, using genomics and proteomics technologies. Today Mr Josh La Baer, will discuss about, biomarkers in more detail. And also different phases of, biomarker detection. He'll then continue, talking about, the importance of, statistics, for the biomarker 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. One need to do, many ways of, data analysis, to ensure, that a given protein or a given candidate bio molecules, 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 producible data, and the cells out, of these experiments? I hope today's lecture. Will give you more insight. An itty gritty detail, about, how to do, biomarker discovery based research. So let us welcome Dr. Josh La Baer, for today's lecture. Alright.

Refer slide time: (2:03)



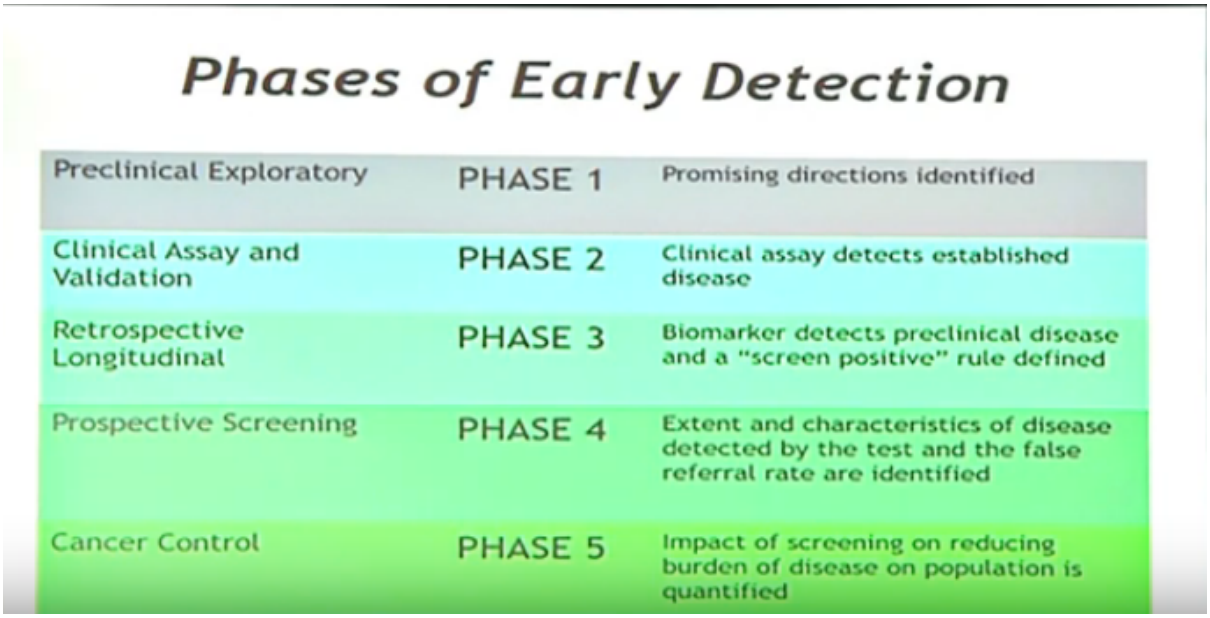
So now we're gonna talk a little bit more specifically about, biomarkers. Okay. Alright.

Refer slide time: (2:09)



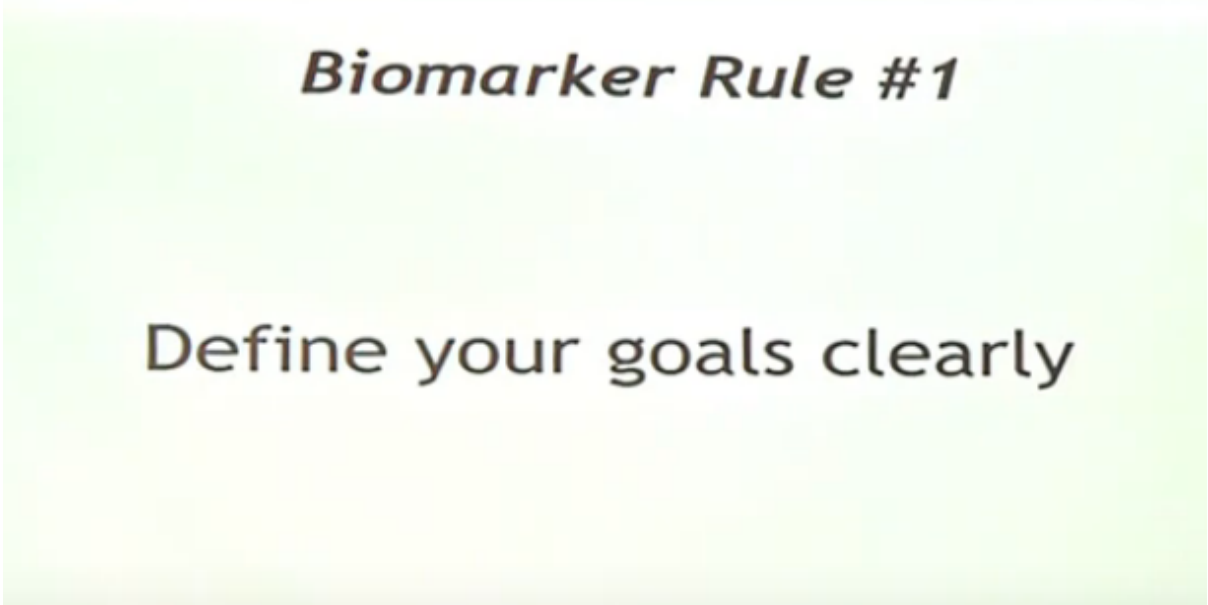
So, I won't go through this part a lot, because we kind of did this. So we talked about, why you would do it? You want to monitor disease, you might want to monitor, whether if therapeutics are, working properly, you can, you might be able to, use the markers to predict, toxicity of drugs or efficacy of drugs. We talked about, the use of a, to screen for disease, to the early detection or even acute diagnosis. Patient shows up, in the, in the hospital, with, you know, crushing, substernal pressure in their chest and you want to know, is this patient, do they just eat some bad food or do they actually have a, ongoing a heart attack and a blood test would be very useful, in that setting. And there are a couple of blood tests, but they're still not, fast enough. You might need a test to look for, infectious disease. Yesterday if you went to the symposium, you heard, about the need for, blood test for tuberculosis. This is an illness that infects a third of the population on our planet. And it's one of the top 10 killers, of all people. And yet, it's very difficult to diagnose. And then, you know, to personalize treatment of therapy, again biomarkers may be helpful, for that. So, all of these are reasons, why you'd want, biomarkers.

Refer slide time: (3:26)



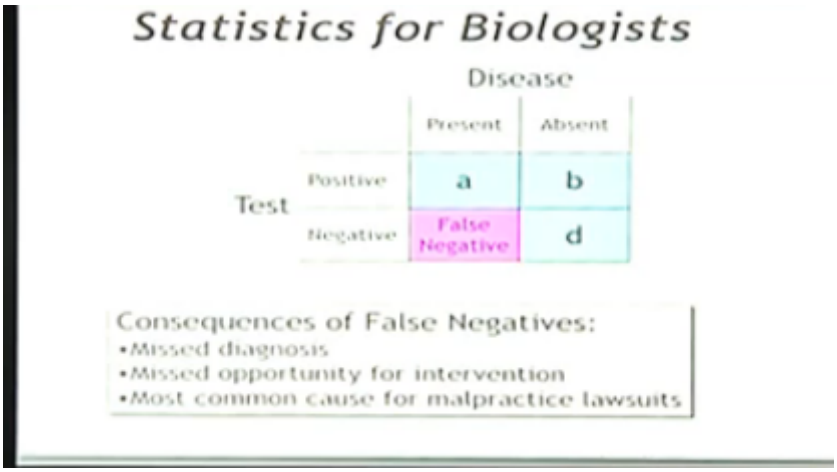
So, this is sort of a different way of saying, what I've told you earlier. This was a, this is based on a publication, from the early detection research Network, at, at the National Cancer Institute, in the US. The, this, this basically, outlines, if you're going to develop an early detection marker, 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, that the assay can detect the, disease. Then they would say, do a retrospective on Steudle study, so you look, these may be, old samples, but you're 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's phase three. Phase four 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, and then cancer control would be to implement the use of that marker, in a large-scale screening population. Okay. So, I'm gonna walk through about, six or seven rules, for biomarkers and then, let's,

Refer slide time: (4:42)



Let’s 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?

Refer slide time: (4:57)



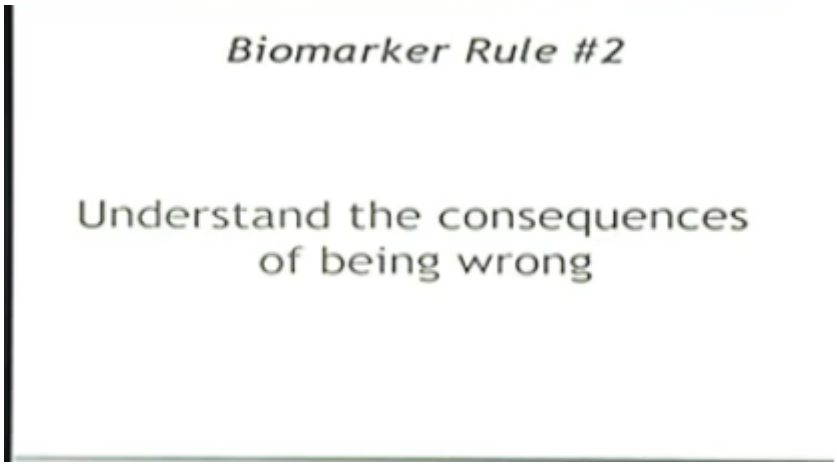
So let's, let's diverge now and talk a little, about the statistics, of biomarkers. Okay. So, this is obvious. Right? You have a population of people. Some of the people have the disease and some people, don't. Right? That's true, of any population anywhere. You got some people in that population, that has it and some people that don't. Right? And we're gonna, for the moments now, let’s 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. Okay? But as you know, nothing is ever perfect. So let's 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. Okay, so.

First we've got this group over here. So we call those, the true positives. True positives means, the test was positive and they actually had the disease. So the test got it right, that’s, that's, as it should be. Right? Okay, 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's when the test is working well. It does what it's supposed to do. Right? So that, that group of people is, a and this group of people is, D. Okay? So, what about this? That's a false positive. Right? What's a false positive? Okay, I got a lot of answers. Over here. Yeah, okay, yeah, the, pit, but the test is positive. Right, the test says they have it, but they don't really have it. Alright. So why do we care? Why do we care? Is it, is it bad, to be false positive? So they might get inappropriate treatment. What else? Right, right, so you're gonna, particularly if it's a, if it's a Ds like, cancer, there's a lot of emotional anguish, to thinking that you're a cancer patient, when you don't really have cancer. Right? Right? And then in some cases, it's also, you put them through needless testing, just 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're actually testing your, your drugs, if you're, if the marker said, that they have the disease, but they didn't have the

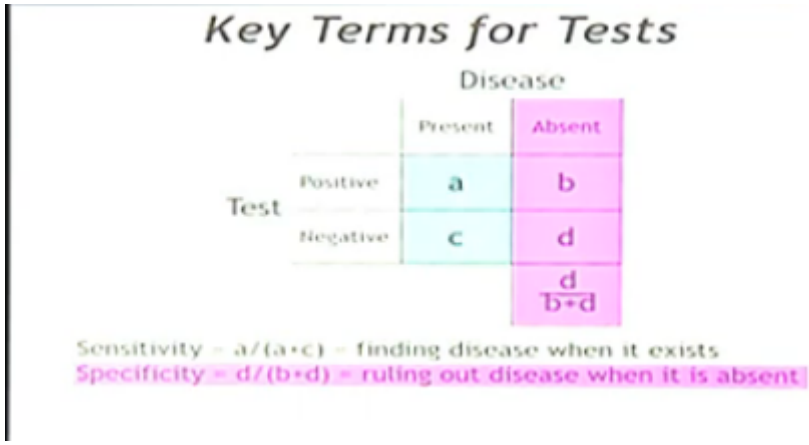
disease and your, your drug won't cure those people and so you'll, you'll get inappropriate results. Okay? And then this group down here, we call those, false negatives. The test was negative, but in fact, they really have the disease. So what's the consequence, of a false negative? Right. That's right. You missed the disease. The patient is ill, you told them, you know, you know what? You're perfectly healthy, go about your life, don't worry about it and then six months later, they have the disease. Right? So, so, this is, the misdiagnosis, it's 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 causes, of huge, you know, loft lawsuits, in the US. And so you don't want to be wrong, about this. The consequences of a false negative are big.

Refer slide time: (9:26)



So rule number two of biomarkers is, understand the consequences of being wrong. You need to know, why it's important to have a good biomarker. Okay.

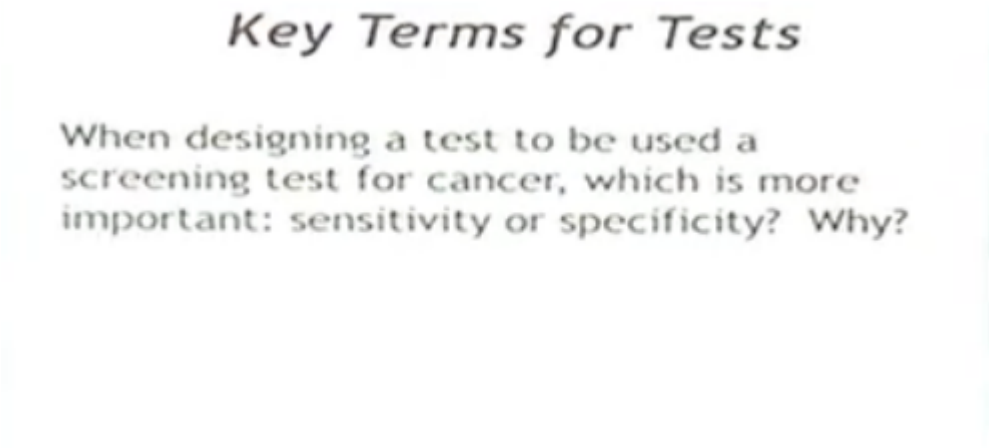
Refer slide time: (9:39)



So now, how do we calculate the probability of disease? Well you take, right, so this is the disease, and, and these are the people, that have the disease. So what's the probability of disease, mathematically here? Right, right. So A, A and C, divided by everybody. Right? So that's, that is the probability of disease. So, in your population, this will tell you, how often the disease occurs. Okay, now, the next thing we want to talk about is sensitivity. Okay? Sensitivity, we define as, a positive

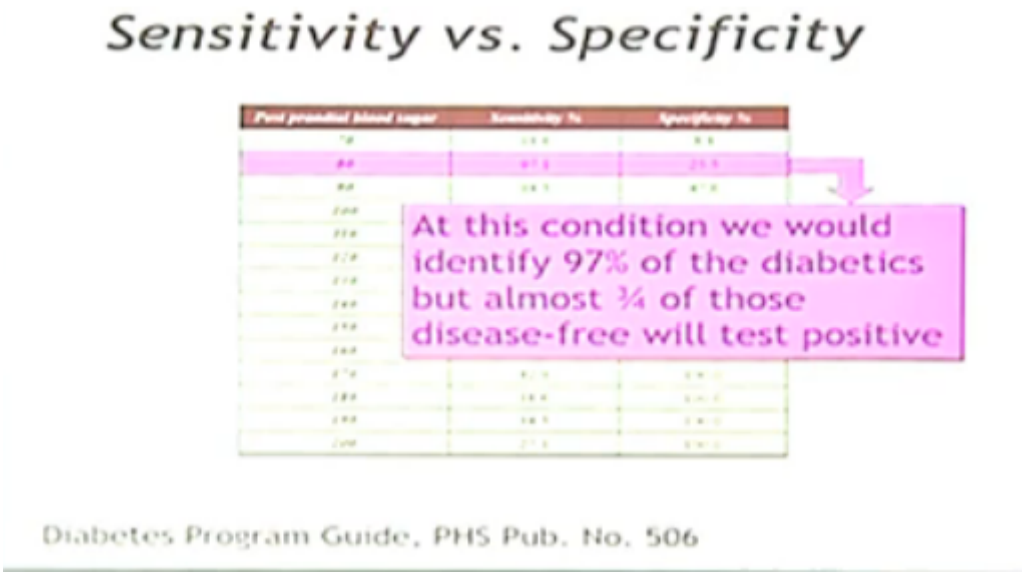
test, in the presence of disease. That's sensitivity. And in this case, mathematically, it's A, over A plus C. So you're saying, these, the denominator is, everybody with disease and A is just, the people who, the tests were positive, for. The closer that A, is to, A plus C. Right? That means the smaller, the negative, the false negative, the false negatives, the better, the test. Right? So that's called, 'Sensitivity'. Finding disease, when it is present.
I make all my students memorize this. Because people often, forget this, stuff. So, so, this is a good measurement, of, how good the test is, at finding it, when it's there. Okay. Specificity is something different. Specificity is ruling out the disease, when it's not present. Okay? So, so, sit or the way that sensitivity. Oh, yeah, okay. Yeah. So this is specificity. You're looking at, the false; we're 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? And so we measure it by, D over B plus D and that, that is, that's the equation here. So it's ruling out disease, when it's absent. Okay?

Refer slide time: (11:58)



Let's do a little quiz question. If you're going to design, a test, to be screened for cancer, which is more important? Sensitivity or specificity? I'm 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 don't want to be wrong, if you tell someone, that they're cancer-free and they're not cancer-free. So in the case of cancer detection, sensitivity is probably, the most important thing. You're willing to tolerate, some false positives, if you have to, to, to make sure, that you don't miss, anybody. Okay. Now let's, let's talk about a different circumstance. Imagine someone going to a doctor, they're coughing up blood, they have weight loss, they have night sweats. Right? And, 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? Here, why? Okay, raise your hand?
So I know who did. Okay. Right. For the TB? Yeah. I mean the, the point is that, sensitivity isn't an issue here. Because the patient's right there, in front of you. You already know this person's sick. That's not the question anymore. The sickness is already a given. What you want to know, is, is it TB or not? Right? You already suspect it's TB and here what you're relying on, is the test to be very specific, to say, yes, it really is TB and not some other. You know? Some other, illness. Okay.

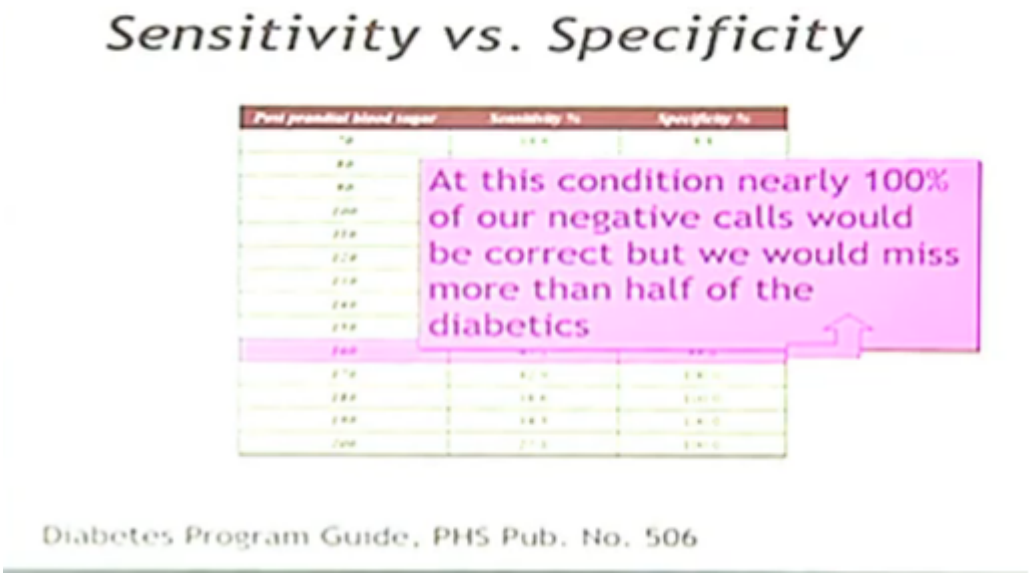
Refer slide time: (13:55)



So now I'm 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 it's above this value, I'm gonna say, it's positive, if it's below this value, I'm gonna say it's 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, in, in opposing ways. So, the, 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. Okay? You'll see why, in a minute. But it is a useful test, to look at this, because it does illustrate, the concept a little bit. Okay? So, so these are the blood sugars, after, after eating a meal, ranging from 70 milligrams per decilitre, up to 200 milligrams per decilitre. And, and here, if you do, if you use this value, as the cut-off, in other words, if you say that if you're above a hundred, you have diabetes.

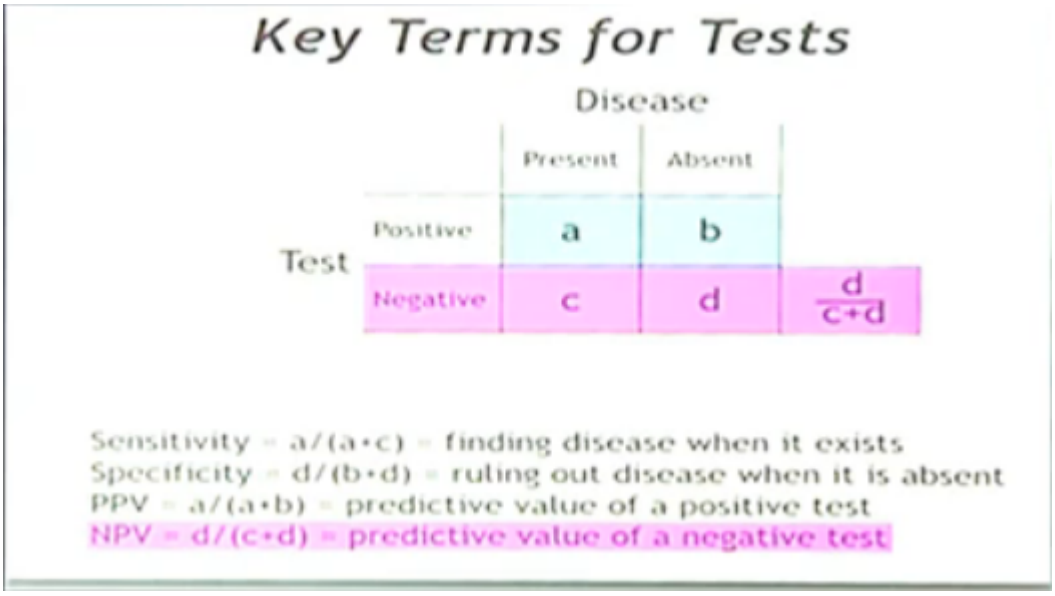
Then this will be your sensitivity and that will be your specificity. Okay? So let's look at this example, here. So you at 80, if you, if used 80s are your cut off, you're going to be 97% sensitive. Right? But you're gonna be only, 25% specific. So one goes up, the other goes down. So what that means is, that, that you're going to identify, 97% of the actual diabetics, the test will be positive, in the presence of disease, 97% of the time. But almost three-quarters of the people, that, that test is disease-free, will also have diabetes. So you won't be very specific. I mean we'll also test positives. So 3/4 people who have no disease, will test, as if they had a huge amount of false positives. Right? Okay?

Refer slide time: (16:19)



Bike, bike, by comparison, let's say well, okay, that was too lenient, let's, let that you allowed too many people in. Let's set a more strict number. Let's say it's 160. Alright? So now this the sensitivity is, 47%, but the specificity is 99%. Okay? So what that means is; that, if you make a negative call, if you say, that they don't have diabetes, you're gonna be almost, always right 99% of the time, you're gonna be correct. But you're gonna miss half the diabetics, you're gonna, miss out, okay, so you're gonna have a lot of, false negatives. And so that's 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's not what you care about. You don't care, how good the test is, what you care about. What's happening to me? 'Tell me about me, I don't want to know, about your test. I don't know, what, how am I doing'? Right? And so, um so, there are two statistical terms, we use to describe, what's happening to me? Alright?

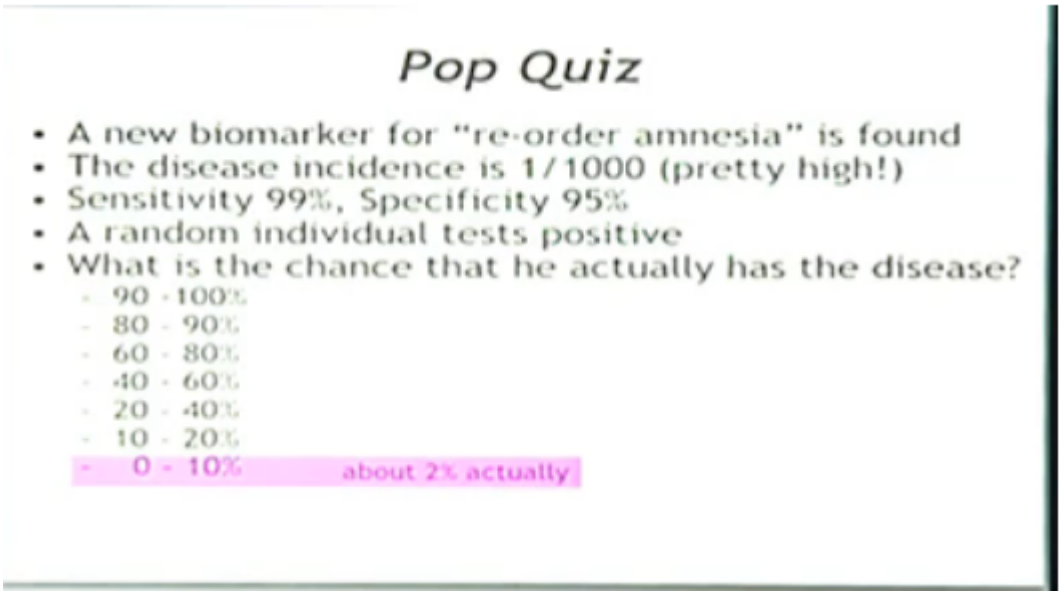
Refer slide time: (17:31)



The first one is that, the, the positive predictive value. Okay, so what do I mean by the positive predictive value? The positive predictive value is if the test is positive, what's 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's shown here. It's basically taking, all the people who actually have the disease, divided by all the people, who were tested, as having a 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 it, 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 don't have cancer or you don't have, your child doesn't 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, is negative. So positive predictive value and negative predictive value, this is, what doctors care about, this is what patients care about. What's happening to me? How am I doing? Okay?

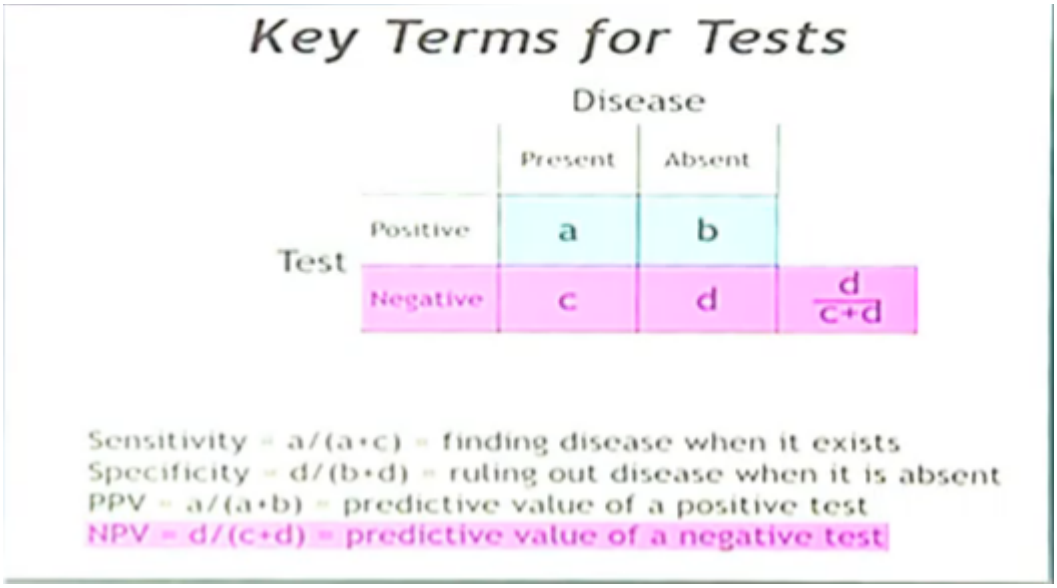
Refer slide time: (19:05)



So ray if we can do a little quiz, now. Okay. So this is a quiz, for in this case, a test called, 'Roberto amnesia'. The disease is, occurs in 1,000 people. So it's a pretty common disease. The sensitivity of our test, is 99% and the specificity, is 95%. Okay? We test a random individual, for the disease, what's the chance that, he actually has the disease? Okay? Got it? Sensitivity 95, 99, specify, specificity, is is 95. So how many people, think that, there's an 80 to 90% chance, that he has the disease? Okay. Got one of those.

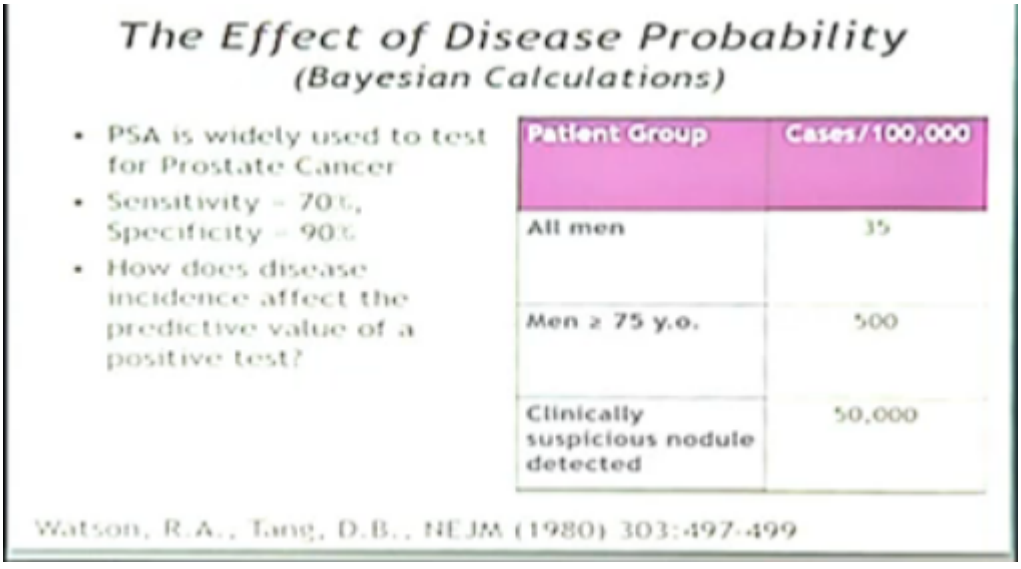
How many people think, it's 60 to 80% chance, that he actually has a disease? How about forty to sixty? I'm going to Assume; you got it all wrong, if you didn't get it. 20 to 40 percent? I got one, 20 for 40. So far, okay. How about 10 to 20? How about 0 to 10? Got a few of those. The rest of you all think that it's 90 to 100. How many think it's 90 to 100? Okay, got a few 90 to 100s. Alright. Alright, it's about 2%, yeah. It's, it's about, 2%. Right? Because, the, remember, what, what affects you here, is the, is the incidence of the disease, it's 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.

Refer slide time: (20:57)



Remember that sensitivity and Specificity were down in, in these columns, here. Right? They, those terms, do not depend, on the population. It doesn't matter, how often the disease occurs, for them. They strictly measure, the value, of the test, on whatever specific population, they're being tested, on. But positive predictive value and negative predictive value, they depend on, how often the disease occurs. And I'm gonna walk you through that, in a minute. But it's 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? Okay.

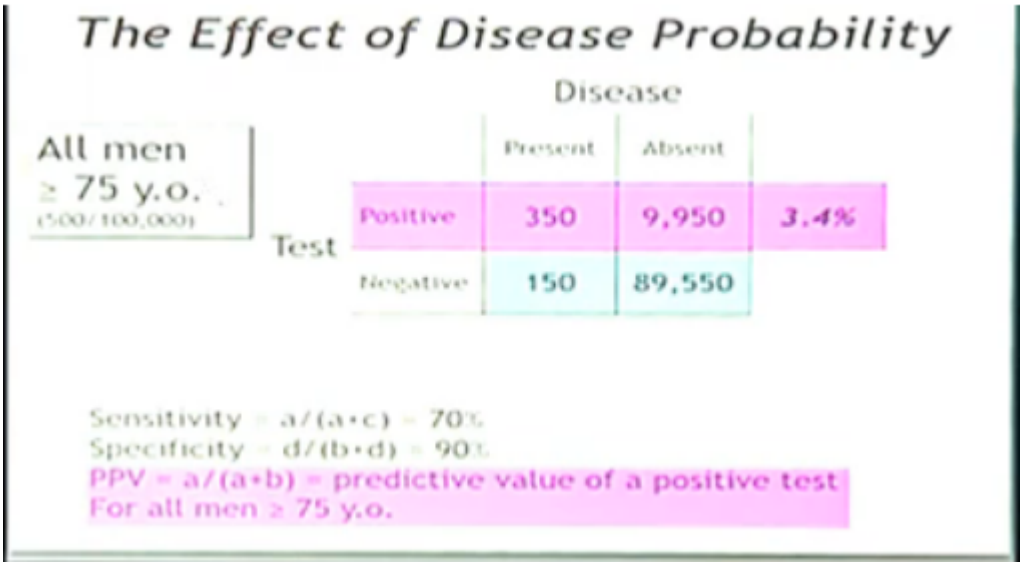
Refer slide time: (21:41)



So let's walk through that. So this is um, this is, this has to do with, what's called, ‘Bayesian calculations’, which includes looking, not only at the probability, but also at, what's called the, ‘Prior Probability’, which is, when you begin your test, what was the likelihood, to start with? And we're gonna use as an example, the prostate specific, and antigen test. It's a very common test, used to detect, prostate cancer. It has a sensitivity of around, 70 percent and a specificity of 90%. That's one of the best values, you'll see anywhere. You know, that's a pretty typical marker.

When people prove, when I told you before, but people published, ninety nine percent, ninety nine percent, you don't believe it. Numbers like this, that's kind of what you'd expect, from a, pretty good marker. So now we're gonna ask the question; How does incidence or prevalence, affect the positive predictive value of a test? We're gonna consider, three different Populations. We're gonna consider all men, in which case, the incidence of, prostate cancer, is thirty five cases, in a hundred thousand. We're going to consider men, who are over seventy five, in which case, the prevalence of the disease goes up to, five hundred, four hundred thousand. And then we're 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's a about a 50% chance, that they have cancer. Okay? So three different populations, these are the instance. Remember I told you, the probability of disease, A plus C over, A plus B, plus C, plus D? That's what, these numbers are, right here. Okay? So let's look at the first case.

Refer slide time: (23:21)

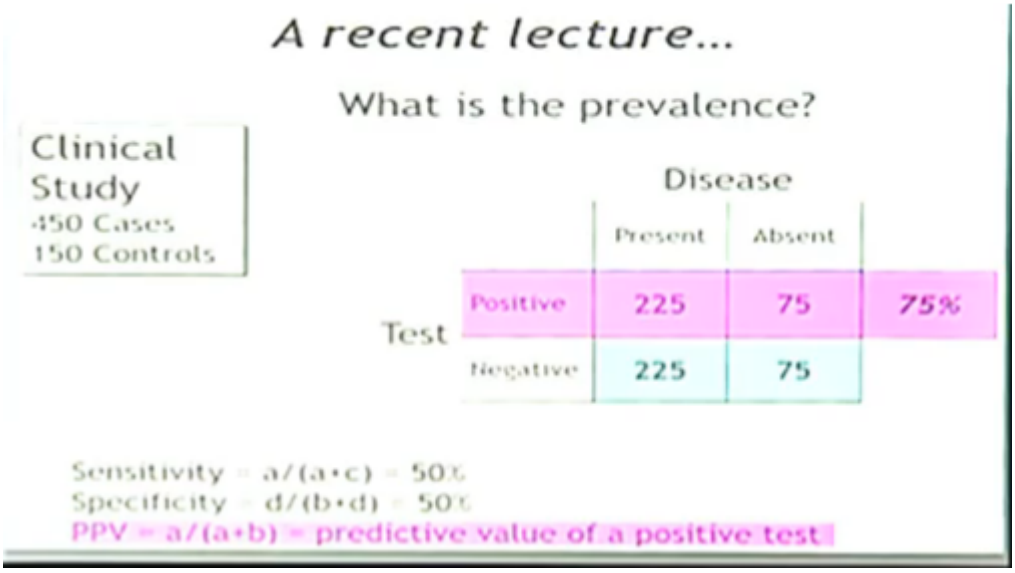


In this case, we're looking at, the clinical nodule, a 50%, likelihood, to start, that this person has cancer. Right, so, you so notice, that I have, that, this number here, and that number there, add up to 50,000. Right? So 50,000, remember I said that, out of a hundred thousand men, fifty thousand had it? So fifty thousand have it and fifty thousand don't. So that's appropriate. Right? Remember I Said, that it has a seventy percent, Sensitivity? So that means, of this number here, seventy percent or 35,000, are Positive.

And I remember, I said that it had ninety percent specificity? So this number, forty five thousand, doesn't 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, the positive value is, eighty eight percent. So even though, you already have a suspected mass and even though, this test has a set percent specificity, to 95 percent sensitivity, specificity the predictive value is, still not 100%, it's still about 88%. Okay? Now let's look at a very different population, we'll go to the other end of the spectrum. Let's look for men, who, all men, 30, 35 men in a hundred thousand that have the disease. So now let's do the math again. The population that has the disease is, thirty five. The population that doesn't is everybody else. Right? Out of a hundred thousand. Still, we have a 70 percent specificity, sensitivity here and here, we still have, a ninety-five, a ninety percent, and specificity. Look at the, how, how good that, test is. Point two percent, point two percent. So the, the take-home message, here is, that, depending on the population, the positive predict, the positive predictive value, changes dramatically. We didn't change

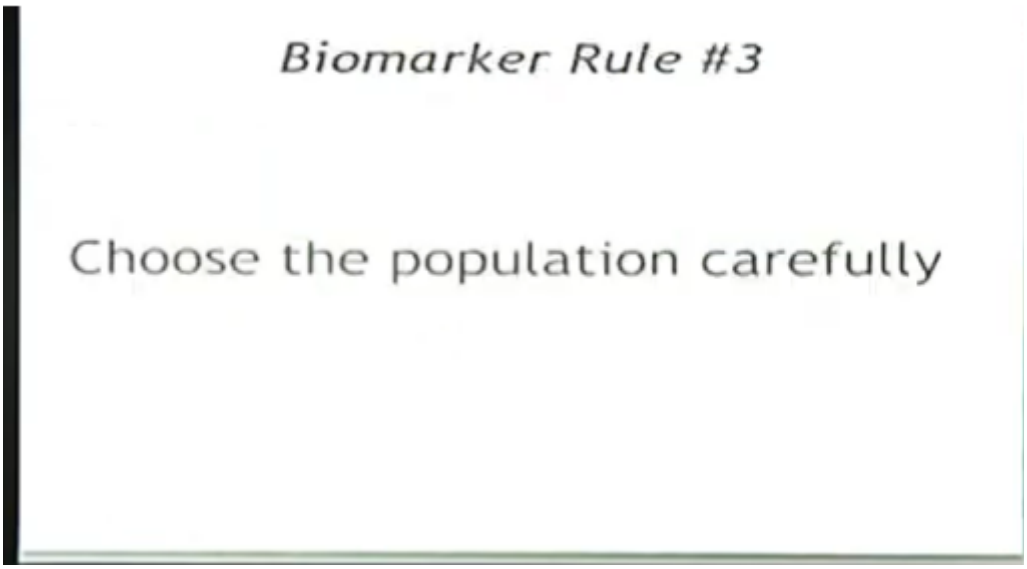
these numbers at all. Those numbers stayed the same, throughout the whole discussion. The only thing that we change was, how often, the disease occurs. And if the disease is rare, then, the predictive value of the test drops, quite a bit. This is, one of the reasons, why, at least in the US, we don't recommend, that young men do, treadmill tests, for our heart disease. Because, the treadmill test was designed for, you know, older men. Where it has, good predictive value.
But you know, when the incidence of the disease drops, like it does, here, and then the predictive value drops, precipitously. And then, the risk of a false positive becomes much higher. Okay. And then, this is just a shorter sort of, show you, the kind of more, general circumstance of, 500 in a, hundred thousand. So this is, not far from the, what you know, one in a thousand, 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're dealing with. Okay.

Refer slide time: (26:22)



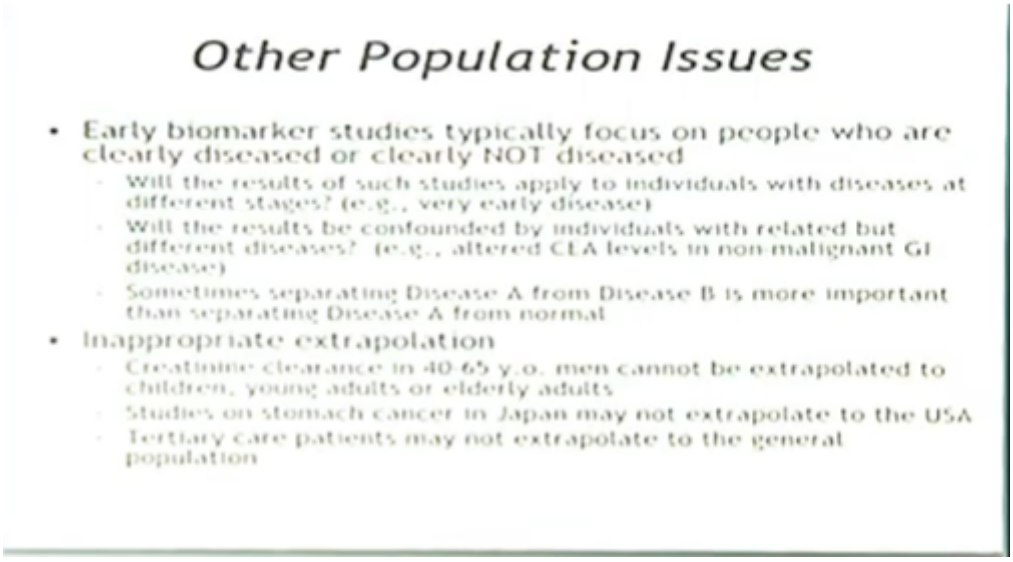
so I, I sat through a lecture, in my Institute, Where, someone was boasting about, his, test that he had developed and he, hit, this is the clinical study he did, 450 cases, in a 150 controls. So the prevalence, in this Population, is what? So the, the prevalence is very high. Right? Because you, you're, you're 3/4s, of the people, in your study, have the disease, 3/4 of them have it. Right? So, so he did that, he had this, positive test and he said that his predictive value was 75%. Right? And, and I looked at the numbers here and it turns out that, if he had zero, if the tests were equally split, between, positive and negative. Right? He would have, like it, half the time it's positive and half the time it's negative. He would have still had a, predictive value of, 75%. So he had to do, nothing. The test had to, had zero predictive value, and in a sense and it would still have given him a positive predictive value of 75%. So is pretty lame presentation.

Refer slide time: (27:32)



Okay, so rule number three; choose your population carefully. Right, alright.

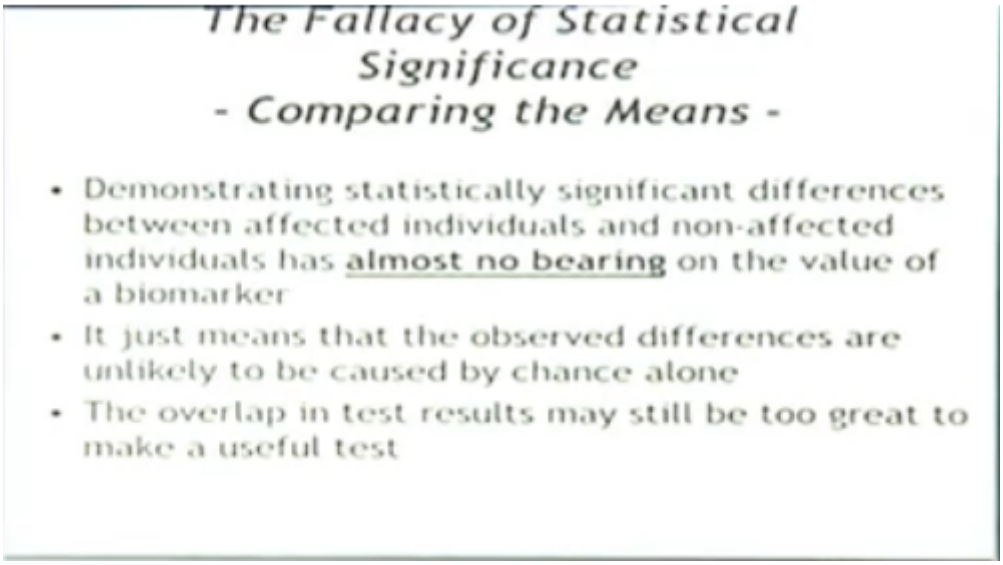
Refer slide time: (27:40)



So, if you're going to do an early biomarker study, then make sure, you pick people, who have early stage disease. Because that's 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 CEA levels or have, non-malignant, GI disease? And, just remember that, sometimes it's a more important, to separate disease A from B, then, disease A from normal. So, imagine if you're in a clinic and someone walks into your clinic and they have abdominal pain and they tell you that, they've had abdominal pain, for months and you know they've been, losing weight. Right?
In that case, you're 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's ill. They've been suffering from, GI symptoms for months. So they're you know,

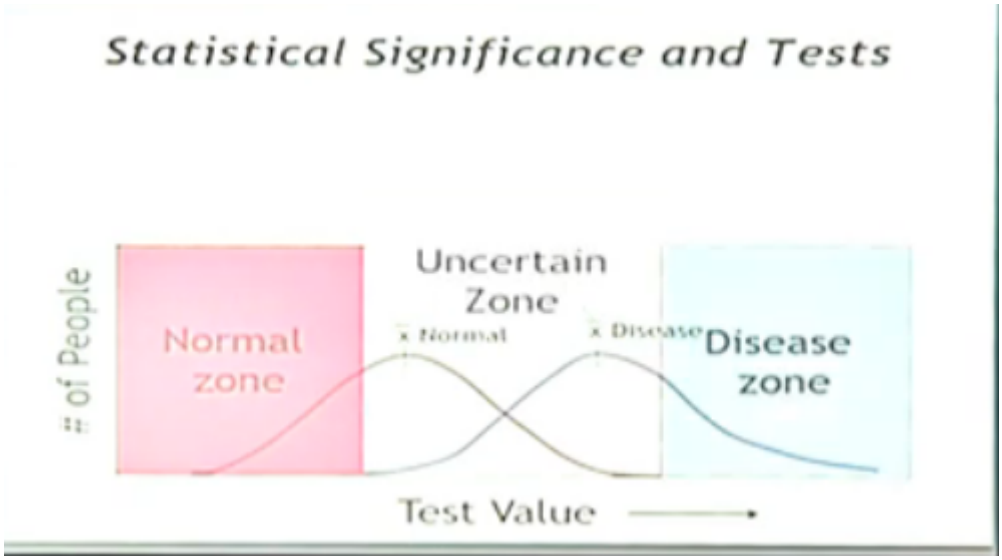
there's something wrong. You're not separating normal from, from dizzy, from cancer, you're separating, cancer from, other GI diseases. And so always remember that, if you're going to do a study, to find a biomarker, you should find, you should use a population, of maybe, people with, with, non cancer GI diseases, from GI, from cancer diseases. You need to make sure, you don't, extrapolate inappropriately, if you, if you develop a test, that's good, in one population, it might not work, in another population. If, if for example, their, their kidneys don't work, as well, in older people. If it's something that's excreted by the kidneys, the tests may work in a twenty year old, it may not work, in the 60 year old. Diseases on stomach cancer, for example, don't 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, from healthy people. Okay.

Refer slide time: (29:46)



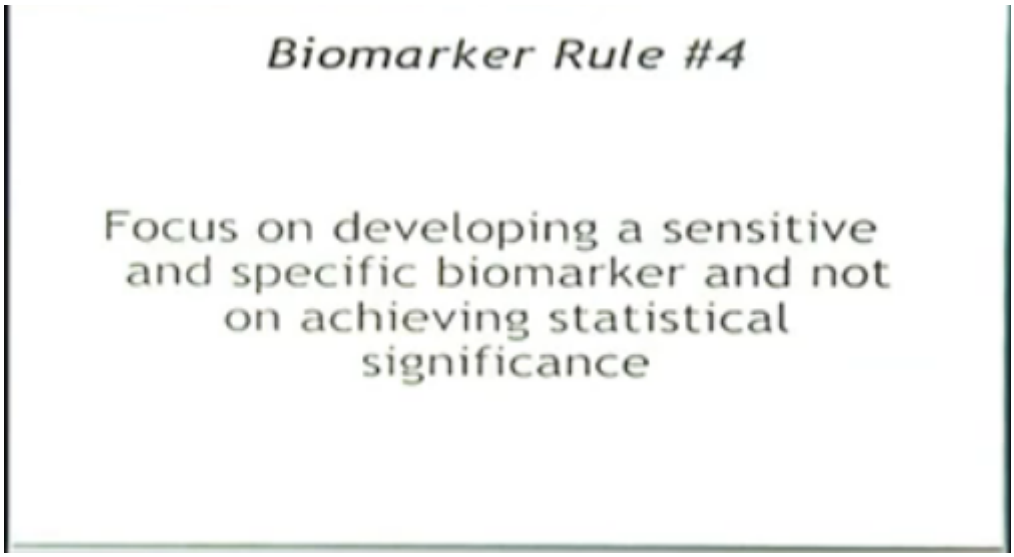
And then, this is something that we talked about, a little bit earlier, and already. This is, what I call the fallacy of, of, of statistical significance. And so we, we kind of covered that already. Just because there's a, there's a good p-value, between a and B, doesn't mean, that they're good, biomarkers. You should be using, sensitivity and specificity, not p-values and,

Refer slide time: (30:05)



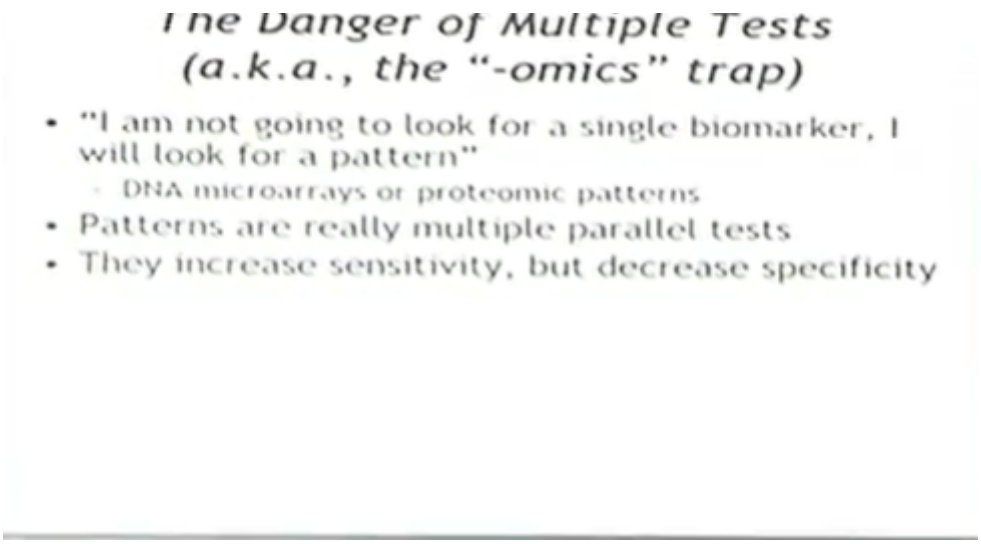
And that's really shown on this thing here, which we've already covered, so I'm going to skip that. Alright, alright.

Refer slide time: (30:15)



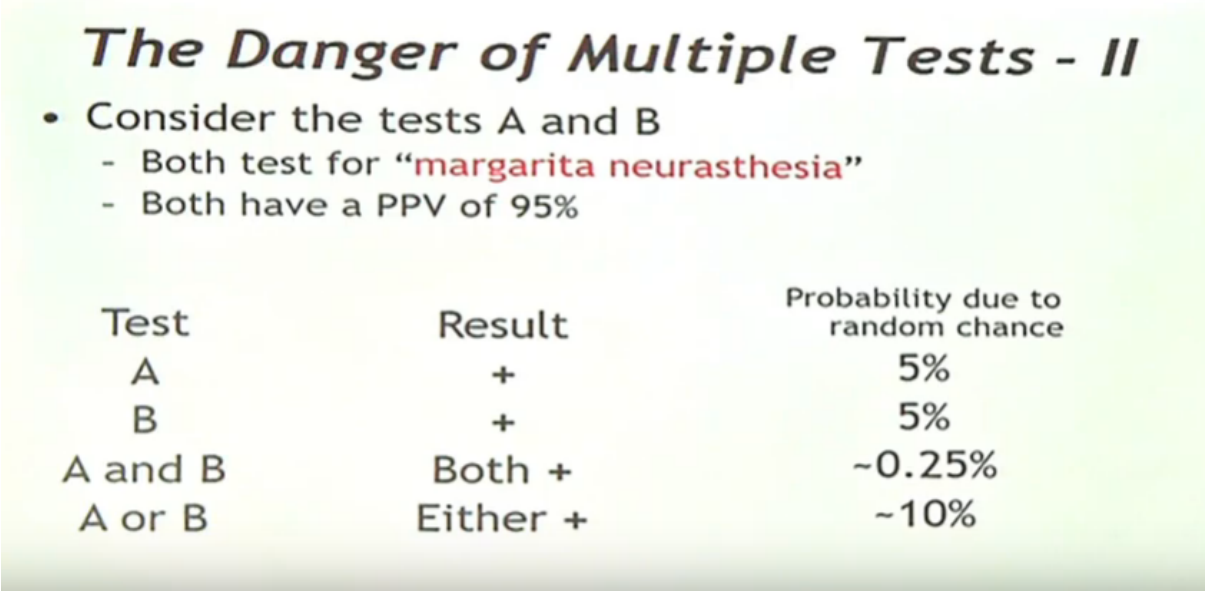
So focus on sensitive specificity markers and not on statistical significance. Alright. That’s fair, alright.

Refer slide time: (30:20)



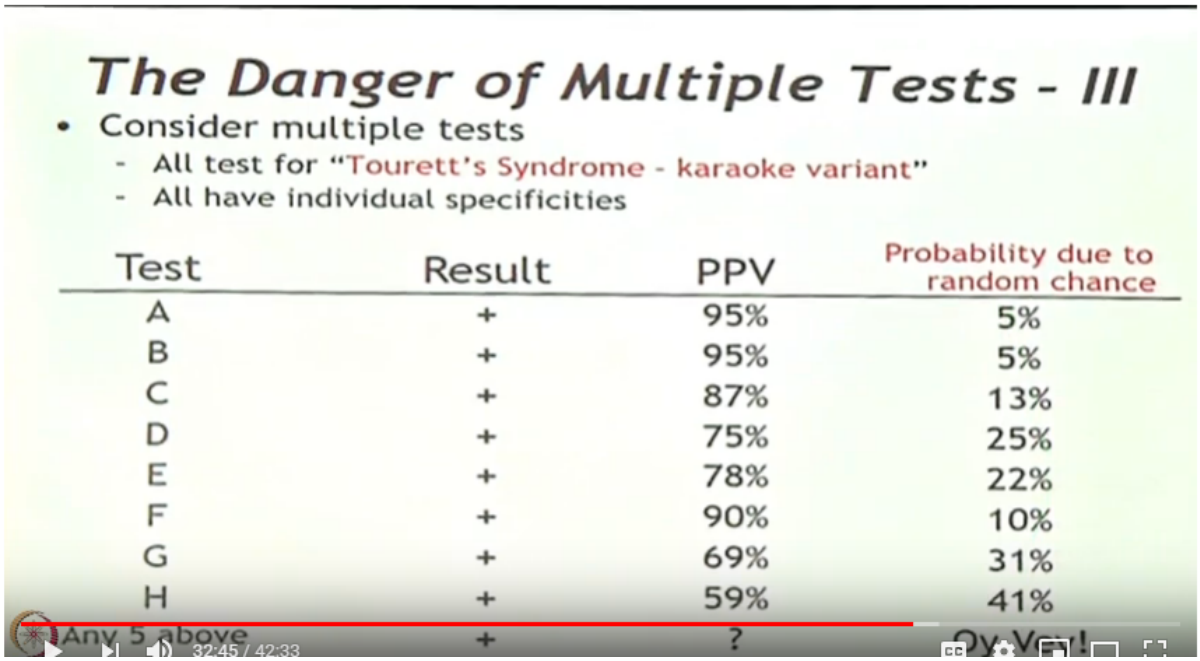
So now I want to mention a little bit, the and we're coming to the end here. The, the Omics trap. Because all of you are, many of you are, going to be doing Omics studies. That's what we all do these days. And you often hear this statement, from people, in the Omics studies. I'm not going to look for a biomarker, I'm going to look for a pattern, I'm gonna look for a signature. They might be doing it, on DNA, microarrays or protein arrays. But you have to remember that, a, a pattern is really, multiple, parallel tests. They're doing a bunch of different, molecular, statistical studies. And they, by doing multiple tests, they increase your sensitivity. Because each tests, 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're going to do multiple parallel tests or look for patterns, my, my biggest advice is to get a statistician. Because you're gonna need, more careful statistics. And, we, this class is not prepared, we're not going to do, those statistics here. You just need to be aware, that, when you get to that stage, it's time to engage somebody. So, we have two tests. Imagine that, that this test,

Refer slide time: (31:40)



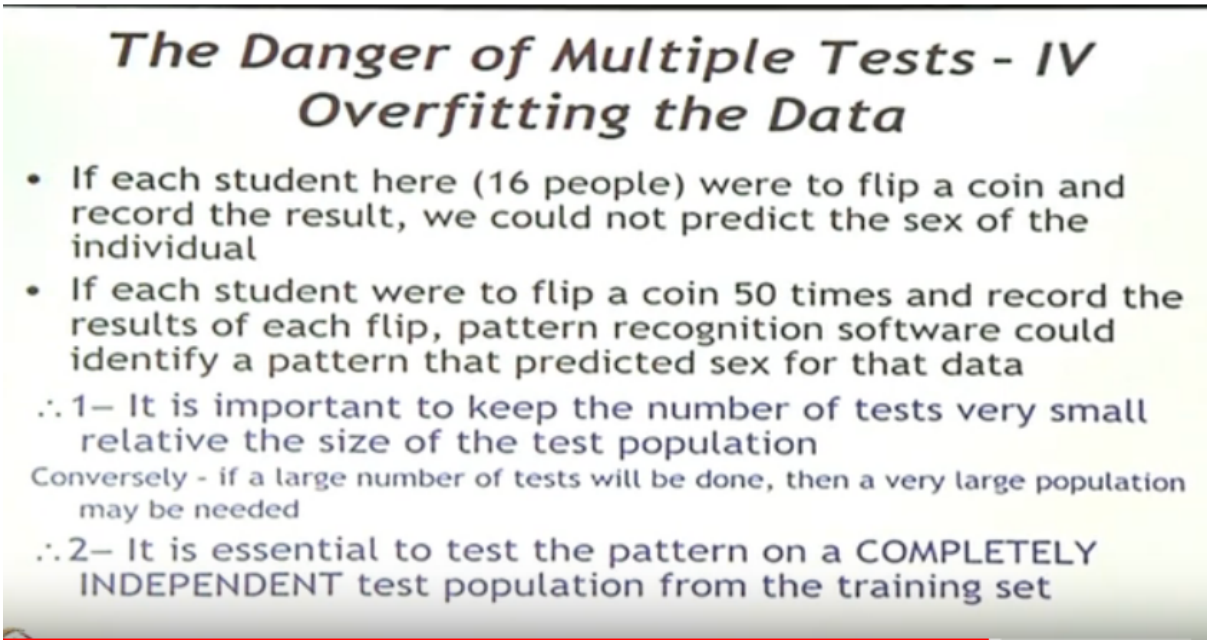
They have 2 tests, for the same illness and they're testing for a positive and they both have a positive particularly of, 95 percent. So, imagine test A, is positive and it has a probability, the probability that it's going to be positive is, 5%. Test B might be positive, so its chance, to chance alone is, 5%. 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 not have to add the two effects together. Okay, yeah.

Refer slide time: (32:25)



So this is even, it now imagine if you do this, with multiple tests. So now you have a whole series of tests. I'm going to just go. So now, each of these is gonna have a, different positive, predictive value, they're going to have all kinds of different, due to random chances and if you add them all up, you, numbers get to be, outrageous.

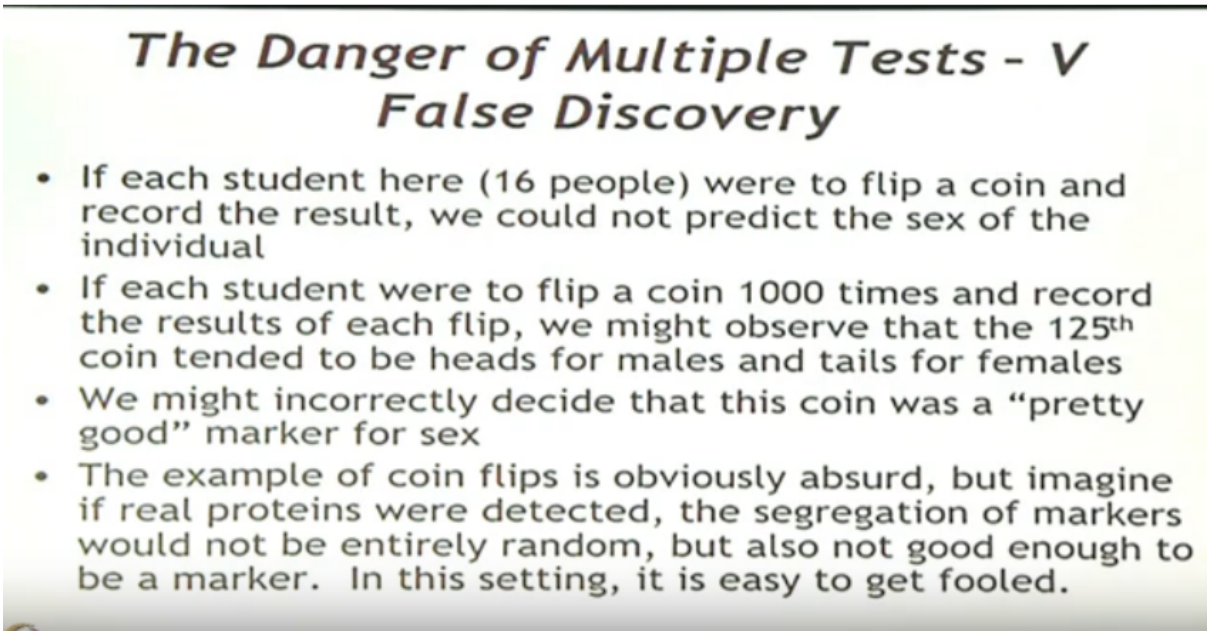
Refer slide time: (32:49)



So again, the take-home message is, get, get a biostatistician. So here's the example that, that I like to remind, people of, when they're 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, the result on the coin, would predict, the gender of the individual, who flipped the coin? Right? Nothing, right? Okay, now, let me change that. Let's say, that I gave you each, 10,000 coins to flip. And you went one by one and flipped every coin and you mark down heads or tails. What's the chance, that among those 10,000 flips, that one of them, maybe the five thousand six hundred and thirty fifth, of them, would correlate, with sex of the individual? There's a chance, right? Might not be perfect? But among those ten thousand tries, maybe one of them, by chance alone, would, align, maybe not perfectly, but it would align with the gender of the individual. And you would say, 'Aha, I found that biomarker'. If you, if you flip a coin, five thousand six hundred thirty five times, that one will predict the sex, of the individual. But you'd be wrong. So how would you prove that you'd be wrong? You repeat the study; you do it a second Time, ten thousand. Right? And now, the five hundred six hundred, doesn't work anymore. Now it's the hundred and twenty third. Right? It, it's just random chance, some of them will happen to work. And, and that is, what we do with, Omic studies. We test, ten thousand things, we get one that works and we say, 'Aha, I found a biomarker'. But you tried ten thousand times.

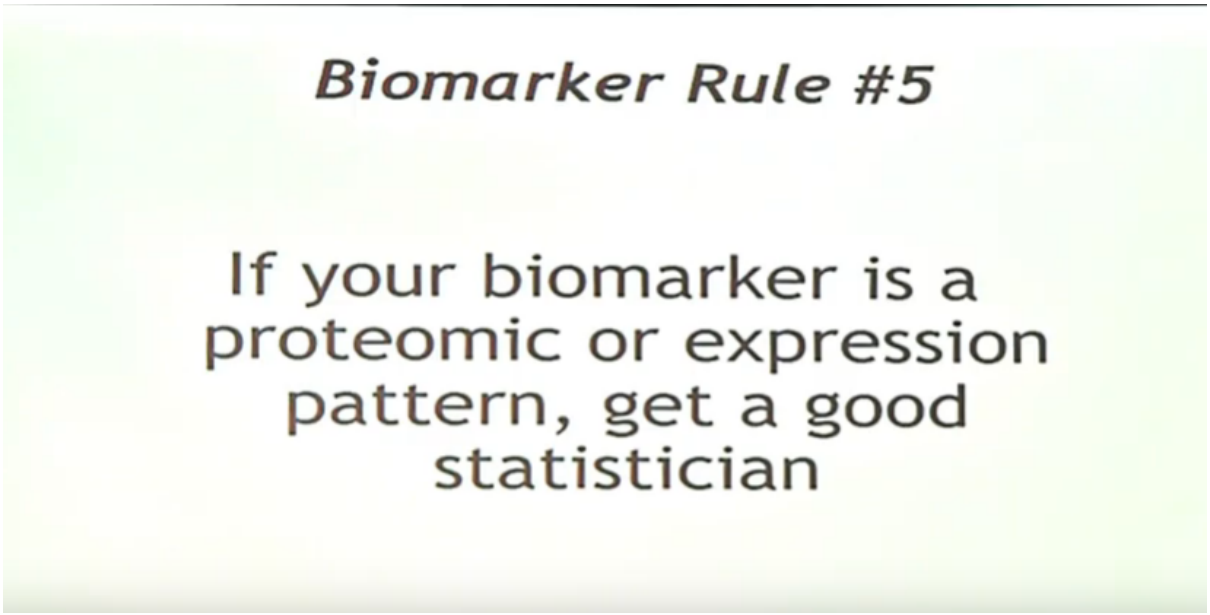
So you have to, you have to, adjust for that, by doing some kind of false discovery rate, adjustments. So, so that's, that's kind of what, what, what I did here. So, imagine if people did all these studies. Right? Right? Uhm, right, you have to keep this, you have to keep the number of the population small. this is especially a problem, when the size of your population, is small, relative, to the number of variables, you're trying. If you're, if you, if you have a study, of a hundred individuals, 50 cases and 50 controls, but you're testing, 10,000 variables? You have this risk of what's called, over fitting. And, and then,

Refer slide time: (35:22)



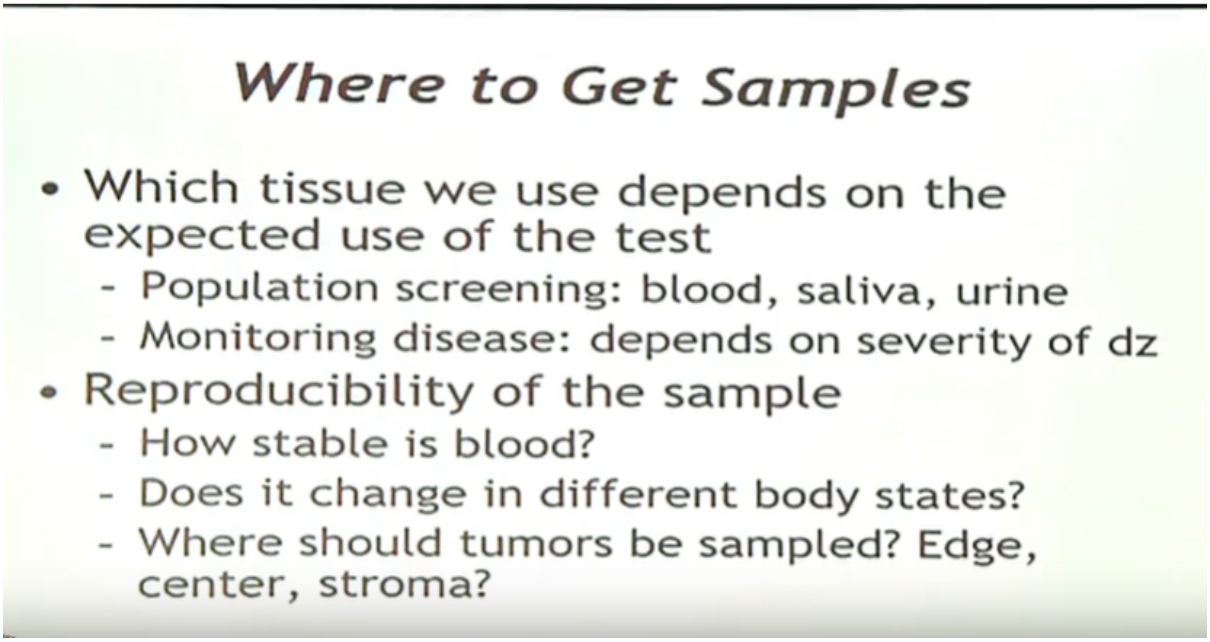
That's why, if you repeat the study doing, a completely, different, population. Alright, so I, I kind of, went through this. Alright,

Refer slide time: (35:29)



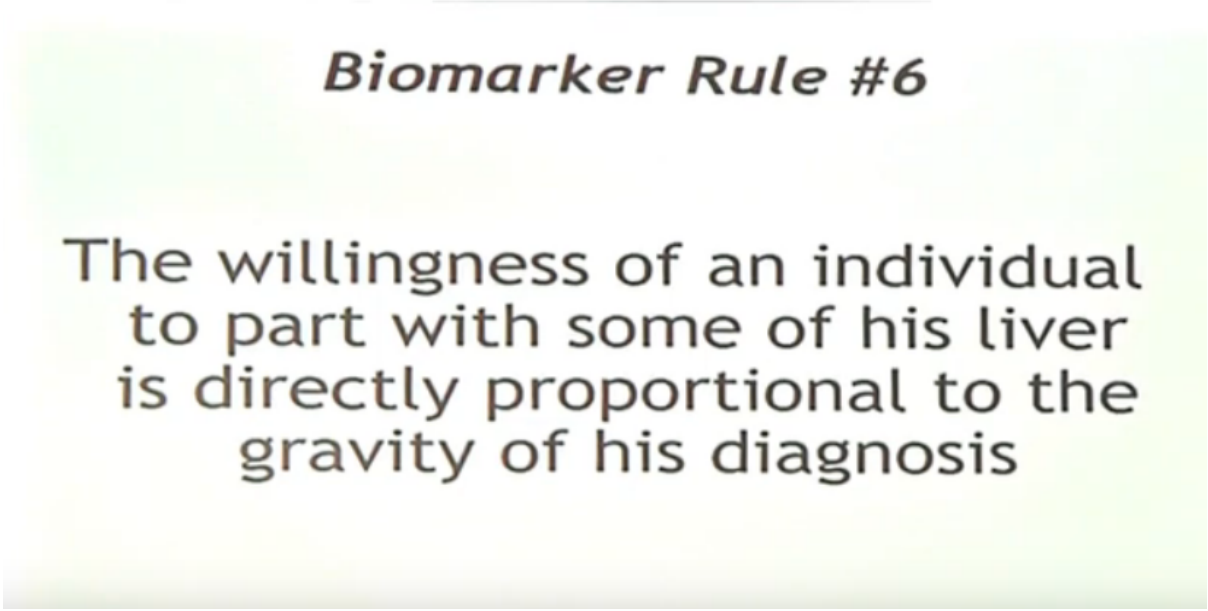
so if your, if your biomarkers of proteomic or expression pattern, the bottom line is, get a good statistician. Okay?

Refer slide time: (35:35)



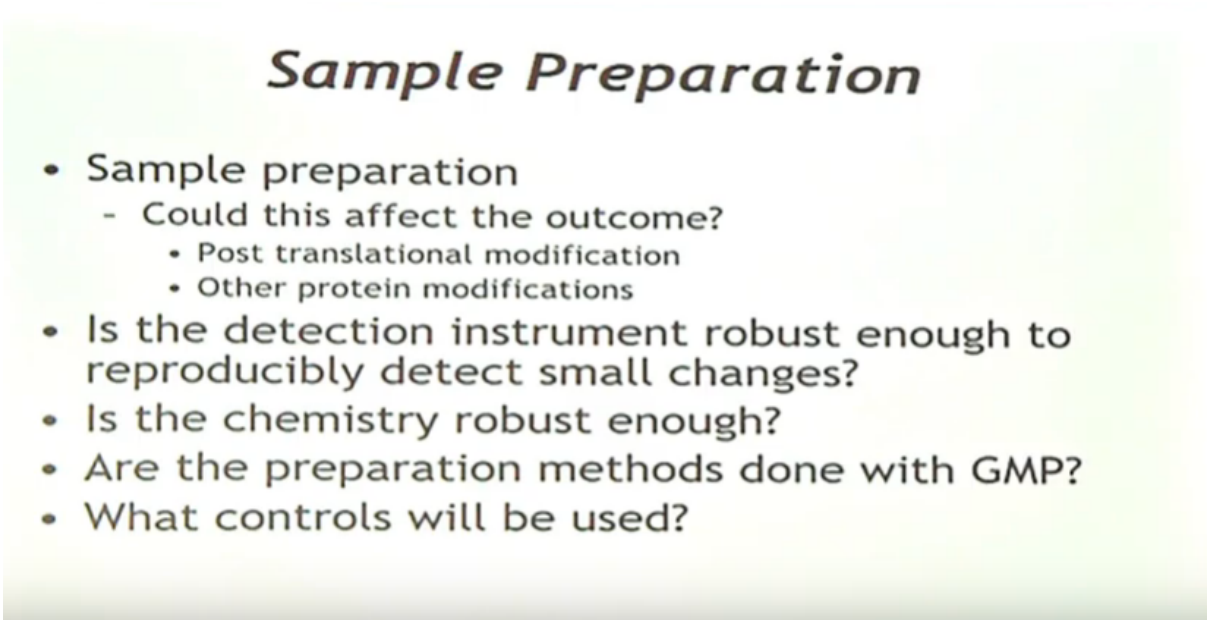
The last couple things, I'm going to mention is, where do you get your samples? Make sure that, the sample that you use is relevant, for the use of the test. So imagine, if you're gonna do, a biomarker, on early these, disease detection, right? We said that you're really going to be testing a healthy 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? Right? You know, if you're going to take a test for healthy people, it should be a simple test. Urine may be blood, you know, it's got to be something that you can measure easily. Maybe saliva. You can't, 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're, you have to look at whether this sample, will be stable? If it's a biomarker and blood, will it be stable in blood? Part of, 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're 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's something that, you have to consider. And then of course, if you're, measuring samples from a tumour, you need to look at, where you're, you're taking your biopsy from.

Refer slide time: (37:17)



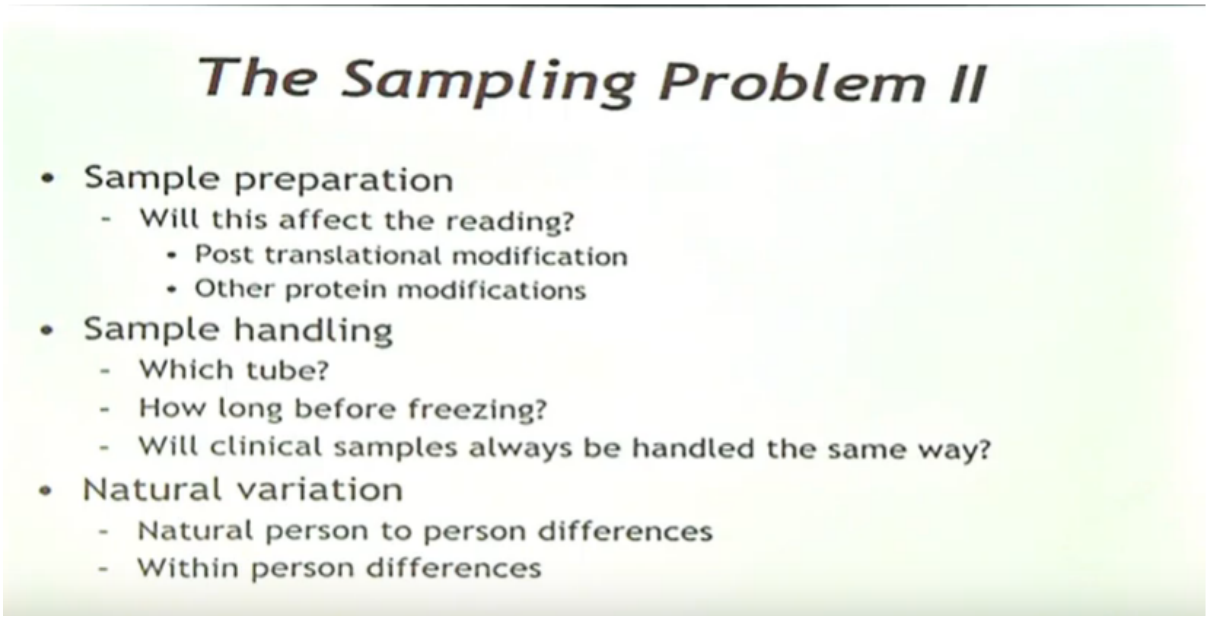
So rule number 6 is, the willings of an individual department, some of his liver, is directly proportional, to the gravity of his diagnosis. People do not give up, parts of themselves, easily. They only do so, when they're really sick. So that's good to remember that.

Refer slide time: (37:26)



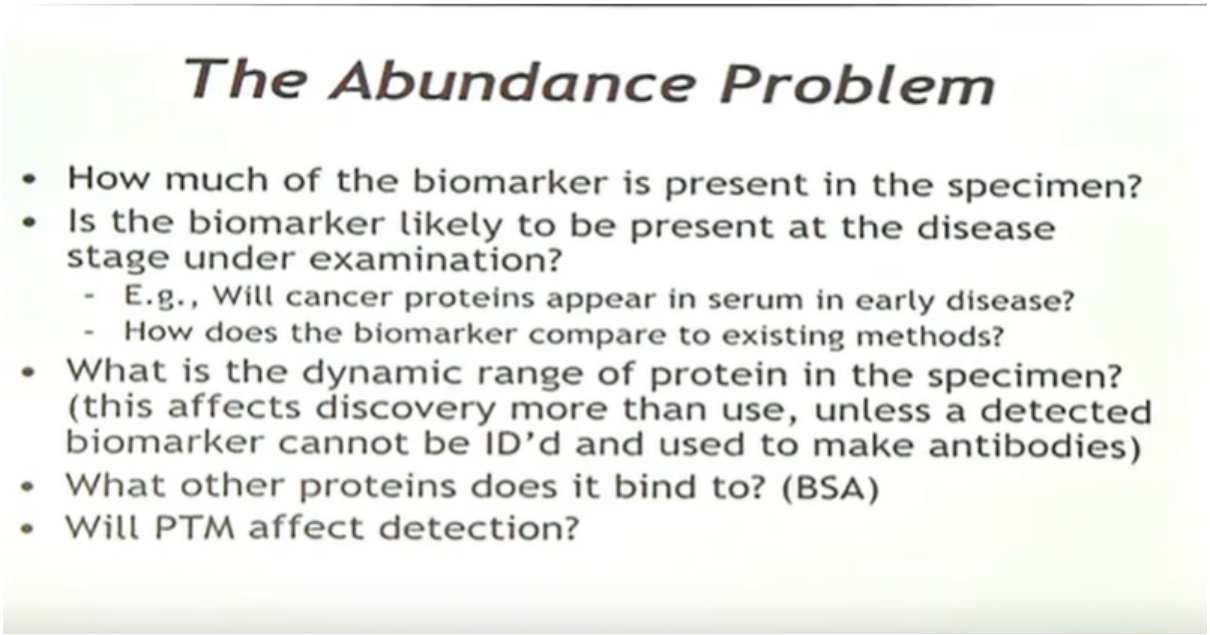
Okay, part of biomarkers, knows how to prepare your samples. How are you going to preserve it? That could dramatically affect outcome. I already gave you the example. With Paul Tempst. Is the instrument, robust and reliable? Is it going to give you the same answer, every time you measure it? Is the chemistry robust? It well, if you Ship, 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're going to use?

Refer slide time: (37:55)



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

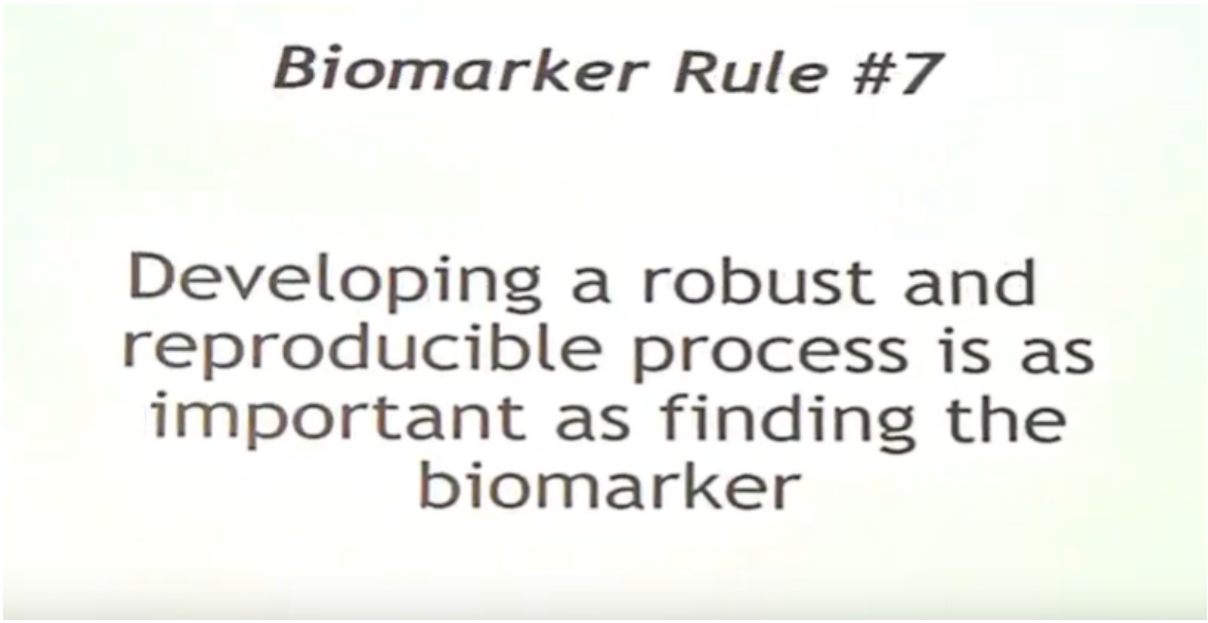
Refer slide time: (38:23)



And then, then there's 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

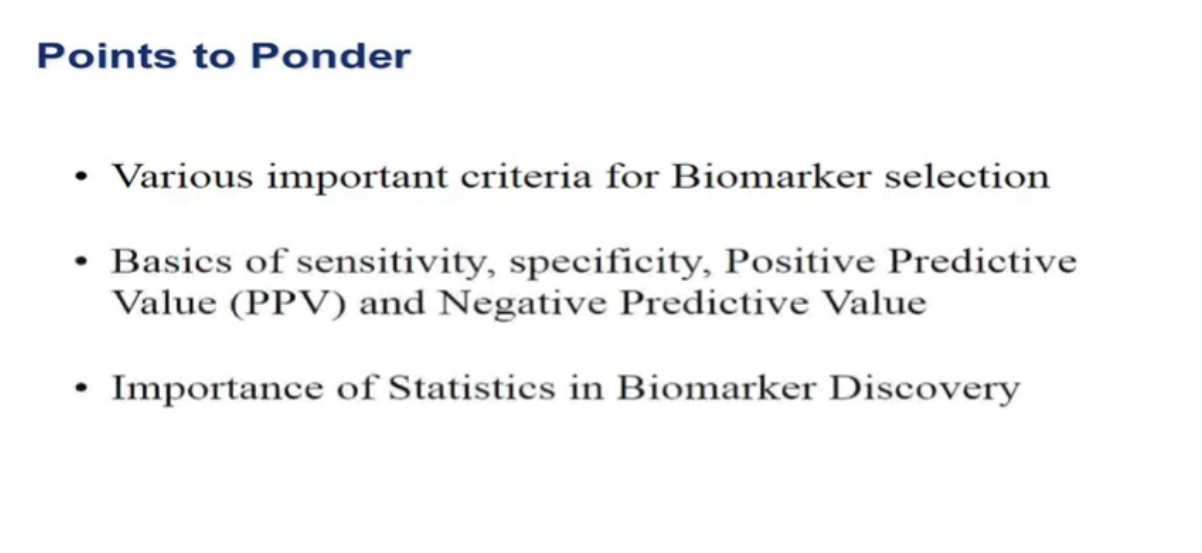
detect it? So in the case of early biomarkers, early detection biomarkers, is there going to be enough there, in an early specimen, from people, with early disease, that you can actually detect it? So the marker may be very good at picking up cancer, but it may be too weak, to be able to pick it, in early disease. That was the case with the CA-125 that I mentioned earlier. It was a good marker, for distinguishing ovarian cancer; it's just not abundant enough and early disease to pick it up. Developing a robust, reliable test is half the game. Just because you found a molecule, that looks good, doesn't mean that,

Refer slide time: (39:14)



You've got a biomarker. What you need now is, to develop it, into an actual diagnostic test. And then the last thing, I'm 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. 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, which fit with what you think, is going on the disease. If it's a sort of a random molecule, it'll be a lot harder, to validate it. So I will stop there. Alright,

Refer slide time: (39:43)



So by now you are quite familiar, with the importance of, studying Biomarkers. And you've 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's why, actual clinical translation, of biomarkers, it's 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 this 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've 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 concentrations, which are very crucial and I must say, that despite, all the odds, despite all the challenges, this is, the many bio markets, which are now getting translated, to the clinics, they're getting approved by the US, FDA. And there are some success stories, especially, the OVA1 and OVA4 and so the other protein, which are now coming to the markets, giving you the motivation, that, if we do these kind of discovery workflow properly, probably the eventually, it may be, translatable to the clinic's. Thank you very much.