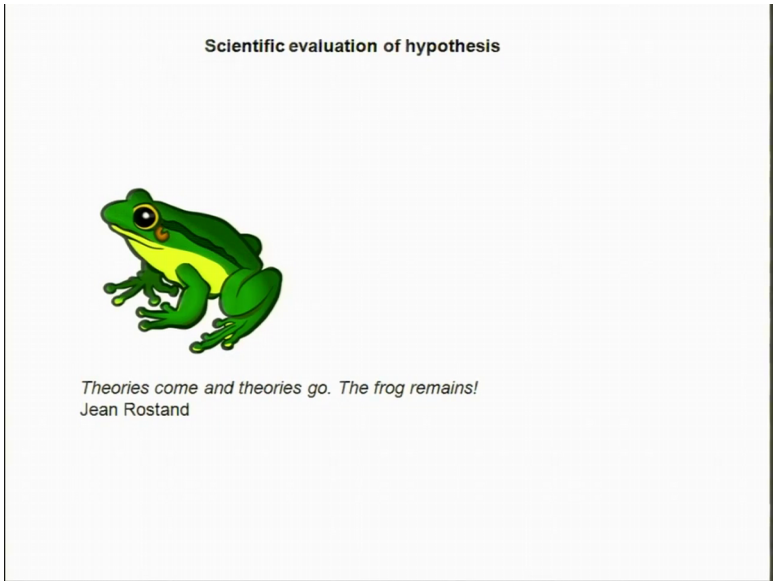
Introduction to Professional Scientific Communication
Prof. S. Ganesh
Department of Biological Sciences & Bioengineering
Indian Institute of Technology, Kanpur

Lecture - 06
Testing the Hypothesis

So, welcome back to this course professional scientific communication. So, in the previous lectures we looked into how do you develop hypothesis, now we are going to look into how do you know you continue the topic on hypothesis and see few more examples as to why hypothesis are important.

Frog is an important model for biology especially those who are spreading on developmental aspects, frog is extremely useful model because they lay egg outside, fertilization takes place outside, the (Refer Time: 00:47) grow outside therefore, we can understand most of the developmental processes and. In fact, much of our understanding as to how they (Refer Time: 00:54) develop, how come from frogs.

(Refer Slide Time: 00:58)

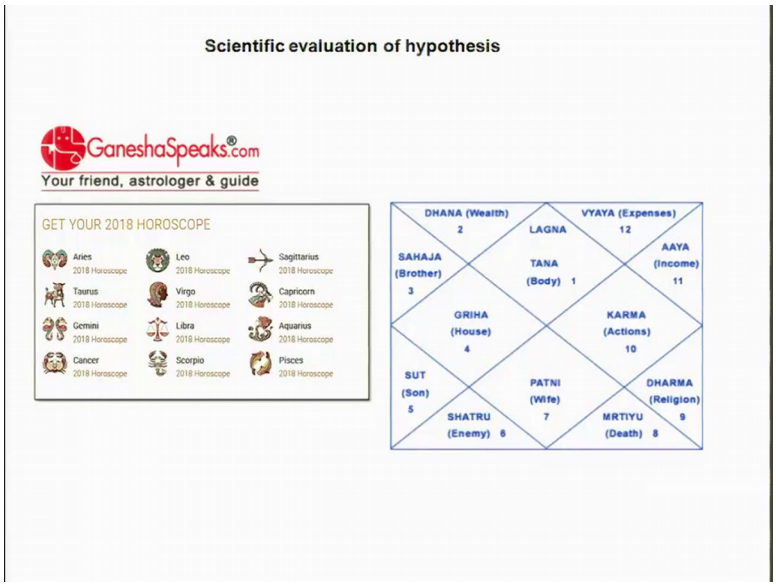


So, there is a famous saying by one scientist Jean Rostand, the commenter the theories come and theories go the frog remains meaning. A number of theories are come in based on to experiments that are carried out on frogs, and they have been disproved meaning theories have been broken therefore, new theories have been developed tested you do get knowledge, but again the theory was you know again proven wrong and so on.

It is enacted just to tell how you know in science you always develop something tested, and the theories no longer viable theory no longer your hypothesis no longer you know attractive, but that hypothesis did contribute to the advance mental science you understood something, and again you develop certain hypothesis and go on you know dream more experiments.

So, what we are going to do is that how do you evaluate an hypothesis, because your hypothesis has to be very, very strong. So, let us take some example this may not sound scientific, but I thought these all examples by which I can explain as to why it is important to have very good observations data to generate hypothesis ok.

(Refer Slide Time: 02:16)



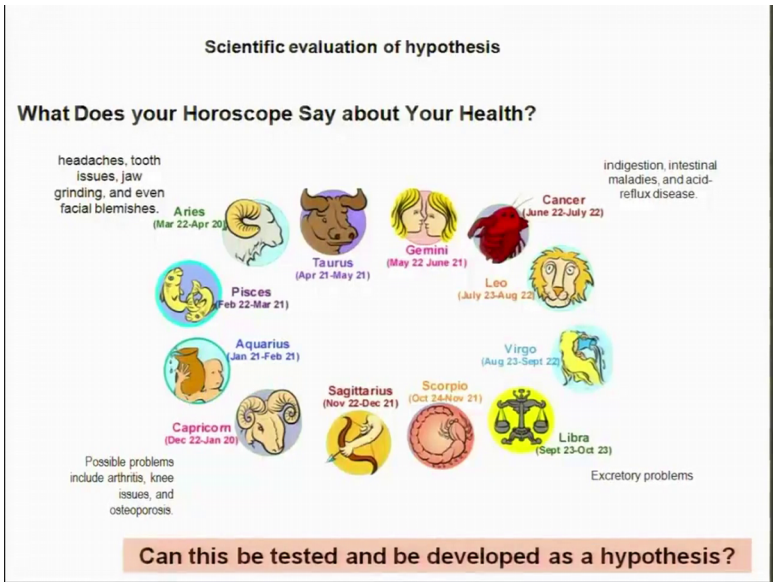
So, hypotheses have predictive value, so you say this is what happens and it has a predictive value and you can test it that is what hypothesis. So, same thing that is in a very similar things have been told in astrology as well for example, you know your horoscope 2018 soon coming, or when you are you know seeing this video already it is 2018, January, February, March, whatever and you may be wondering how this year is going to be firm right.

So, then meaning you and then you may look at your, you know horoscope based on your date of birth or other aspects that are practiced and then it will tell you good bad, ugly, whatever it is depending on what is your luck and so on right. So, this is again the prediction, can this prediction be tested may be yes because you know it says in the

beginning of the year that if you are for example, a Leo they are going to have these, these things they are going to have find it difficult, you are monatory you know financing problem will have where somebody in your family will have health issues and.

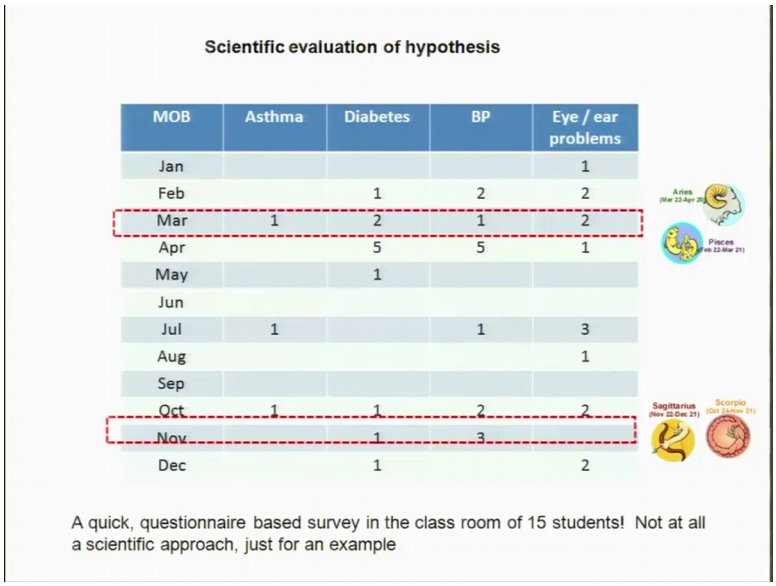
So, on end of the year in December you can go back and ask whether all these you know happen to you or not, then then many of will prediction if they come true would you believe that that is not a hypothesis. So, this is how you saw all how validate right so let us look into some all this what does your horoscope say about your health right let us look into. So, these are again there are you know various ways by which you know people predict this and then one is using the zorac you know same you know you are what is called. So, the ancient [FL] whatever it is say with depending on the month you know in a year that you are born.

(Refer Slide Time: 04:12)



So, there are certain examples shown for example, that if you are a Cancer then you may have problem in indigestion intestinal maladies, and if your Libra you have excretory problems and so on. So, there are some predictions that are shown all right, now is it true. So, that's the question how valid this predictions are so maybe you can do that can this be tested it can be tested you can do that right.

(Refer Slide Time: 04:36)



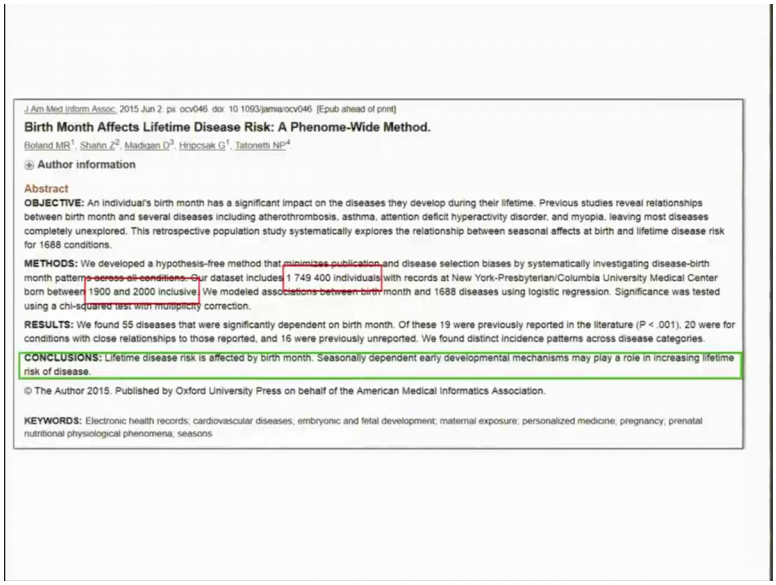
So, this is something that often I give as an example in the classroom, at IIT when I do this course and I ask basically the students to list thing about parents father, mother, which year they are born what kind of elements they develop and try to make a chart which is something that he may also wish to do talk to all your friends. And then 20, 30 number whatever you know ask about the parents, what in which month they were born and whether they have some of these conditions, which are very common asthma diabetes blood pressure eye ear problem and so on.

And you can put plus minus in sort of you know correlate and then test, it is completely random for example, then regardless which month you are born you should have enderbiving, you know these elements, but it looks like often it is not that random for example, this is an indication that is I got it the recent class this year. When I ask this kind of questions out of 15 students this is what they have sort of listed right its interesting, because you know in some month for example, in March or March, Feb and October, November, December, you find you know those that are born in this you know in these period of the year.

They have developed or at least those who are developed these a conditions are happened to the born during you know majority of month in Feb, March, April or October, November, December you does it really make, one possibility is that since the sample size is too small possibly majority of the (Refer Time: 06:20) in this month, there

is a possibility, but this is how you develop the an hypothesis and then test it, but if it is a large sample what we have collected.

(Refer Slide Time: 06:30)



And you can test this hypothesis this is I am just showing you one of the papers, the paper is birth month affect t life time disease risk. A phenome wide method meaning they looked at the phenome type they look at the large number of patients, and they have able to show some interesting correlation with regard to you know everyone you know year or whatever month birth month does it give you know a more fiscal developing certain disease later in your life.

What is interesting is this this particular papers is come from a group in New York, and have looked at all those incidents of the New York City in the US. They looked at a large number of individuals you can see that you know runs in thousands and lakhs, and then they looked at for almost 100 years period like, 1900 to 2000.

So, that is very long period its possible because you know most of you know developing counties have very good hospital backup is a social the government takes care of their health therefore, they have very good records with regard to each individual as to what kind of element they had. So, basically looked at hospital records and they come up with certain conclusions. So, what they have done is they looked at the records of these many individuals right, and then they asked as to whether there is any correlation. With regard to for example, then the birth month and provenance of (Refer Time: 08:05) diseases. So,

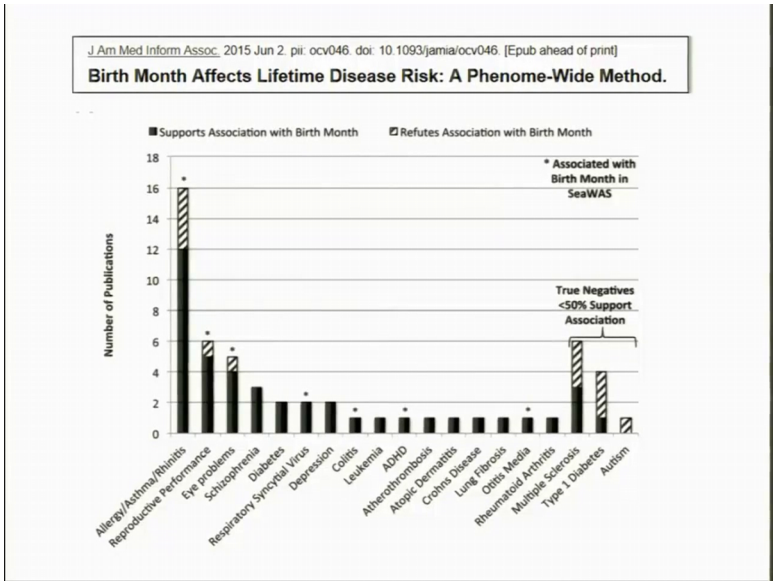
they have looked at 55 diseases you know you can go and study read this paper it is available fully.

And then they try to you know correlate and some other interesting correlations are I will discuss, but the conclusion is the life time disease risk is affected by birth month, if you are born on certain you know month for example, he may be at higher risk or developing certain disease not here in India, but if it happened to be in New York, I will come to that little later seasonally dependent early development mechanism may play a role in increasing life time risk, what they say is that you know any for example, New York or you should look into most of the cities in India also you have seasonal variations January could be cold, and May, April may be very hot and then it became humidate.

And then you have certain months that have you know the flowers blooming you have coming or a lot of dust and things like that. So, it depends on when you are young you are exposed to different kind of environment, when you are just born that possibly make some difference in your physiology which, now you know it is called as a physianist, which could possibly make you know having a kind of having a higher risk for developing a disease or you know kind of a protection we do not know.

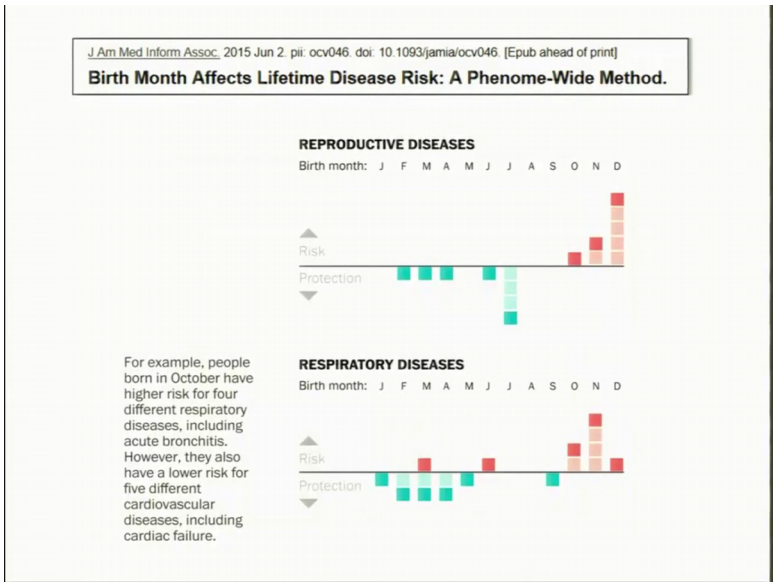
So, this is you know this kind of study can really help us to come up with certain very good hypothesis.

(Refer Slide Time: 09:31)



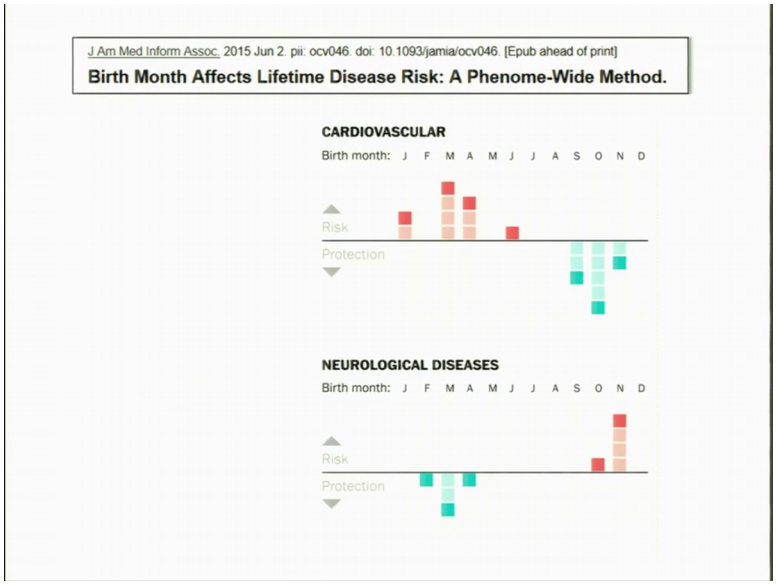
What I am showing is if you know data, this is the bar diagram showing birth month you know affects in you know basically you know individual that are born in a given month whether they have higher risk of in having a disease. So, wherever it is with a star after these are associated with birth month in the sense that allergy asthma rhinitis these are the conditions that shows strong correlation for individual that are born in a given month right and so on.

(Refer Slide Time: 10:06)



So, will go to some other more details this is what it is for example, reproductive diseases. So, one that are shown in the orange color the bar that goes up, if you are born in December, November or October you have a higher risk of developing you know reproductive disease. And on the other hand, if you are born in July then it sought of protects then the probability that you may develop reproductive disease in your if you are born and brought up over there, goes down significantly if you are born in the month of July likewise there are some you know prediction for based on the data, that these are correlation for respiratory disease, for cardiovascular disease, neurological disease and so on.

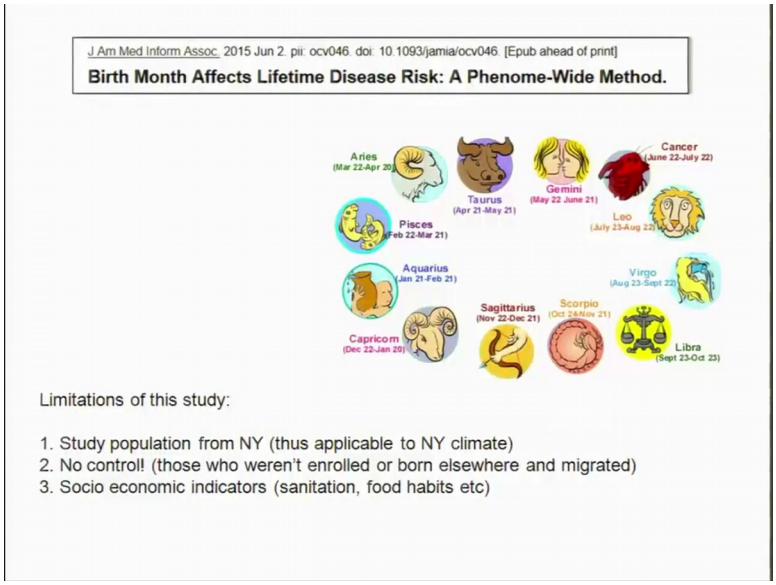
(Refer Slide Time: 10:26)



So, that gives you a good indication as to the birth month can modulate your physiology such that either you have an increased risk of developing a disease, or it can confer protection to you in developing the disease right.

So, that is a you know it helps you to develop an hypothesis because this that i strong, it is not from 15 individual it is from ones in 17 lakhs, you know and then 100 years of you know records. And it is a very, very significant data all this individuals are from new york one particular you know major metropolitan in the us and you can correlate everything with that. So, now, you can you know now ask after all the birth month is what link to the different zodiac signs, now kindly go on say now well over said about the zodiac sign and having higher risk of developing certain disease, could be true the prediction is correct and indeed an hypothesis.

(Refer Slide Time: 11:52)



What you need to understand is that the zodiac sign is applicable to anybody you know who is in the world regardless which country you are. As long as you are born in the month of November or you are born in August or born in November whatever it is then it is a sign, but this the climatic change in the seasonal changes that happen in one city may be very different from the seasonal changes that happen in different city therefore, the environmental factor that modulate your development when you are young true either develop risks, or to have some protective function in the life time diseases you know is depending on environment in which you grew.

So, there are some limitations you have to understand that the limitations of the study are study population is from New York, thus applicable to only the New York climate right. No controls because you never said those who are enrolled or born elsewhere and migrated for example, you know for example, the New York has got huge number of Indians, who migrated after becoming a graduate in India they went and settled there to work.

So, and they moving there they are already twenty five to thirty years now they were born and part of the India. Now after moving in there whether it changes, whether they also behave very similar way as in they are the good controls and then what is not looked that is socio economic indicators for example, sanitation your food habits your lifestyle they also could be modulators. So, if you are exposed to dust for example, that means,

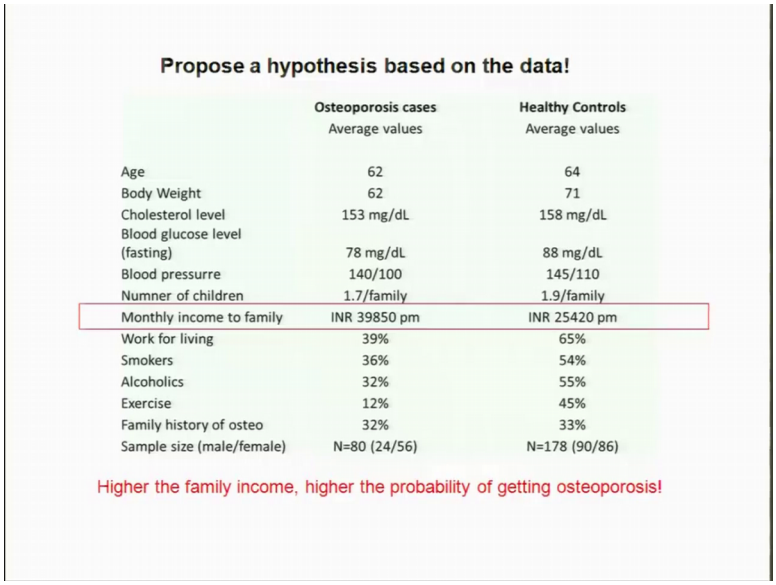
that your house is not well protected, you are exposed to cold for example,; that means, your house is not insulated you do not have heaters.

If your house is exposed to extreme weather conditions for example, have heat in summer again your house is not air conditioned for example, someone who is wealthy good offered all the luxury would have protected the house such a way that, inside working temperature remain constant the air is brewed and. So, on it may not affect as much as somebody who lives in house that is not protected.

So, there are several different concepts you need to look into and then you arrive at certain conclusion, but as how you develop hypothesis and then to test the hypothesis, now based on the study of New York population this particular paper you can develop hypothesis, now you can test you seeing variety of models and control group and so on. And validate indeed whether your birth month can affect the risk you know lifetime disease risk or not, in there are emerging reports which suggest that indeed that is a case again, but need to consider what kind of you know climate that you are exposed to when you are growing up.

Now I am moving into a different problem in developing a hypothesis these are again random, again I use it in the classroom to explain as to how effectively you can develop hypothesis.

(Refer Slide Time: 14:58)



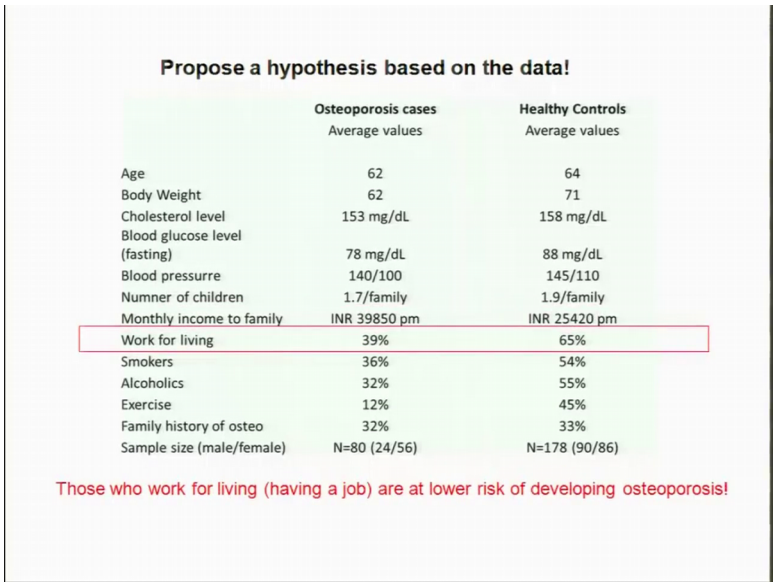
And you can look at the table here, it basically look at two groups of individuals. So, one is called cases osteoporosis a condition in which that bone becomes you know better and then there are healthy controls, so not having such kind of condition.

So, these are the two groups and we are looking into a large number of parameters, some of them are biochemical parameters, some are their habits and others you can see that age is even 62, 64, the body weight, cholesterol level, blood glucose level, blood pressure, number of children that they have monthly income, working for living whether the individual who is having this you know condition he is working, or smoker, alcoholics, whether they do exercise, whether they have any family history.

Because there is a family history then then there is a genetic component, basically this is what we look into and then you need to analyze the data and develop an hypothesis. So, by looking through you know the data what basically you look at where are the differences between these two groups, based on these whether you can develop certain hypothesis right.

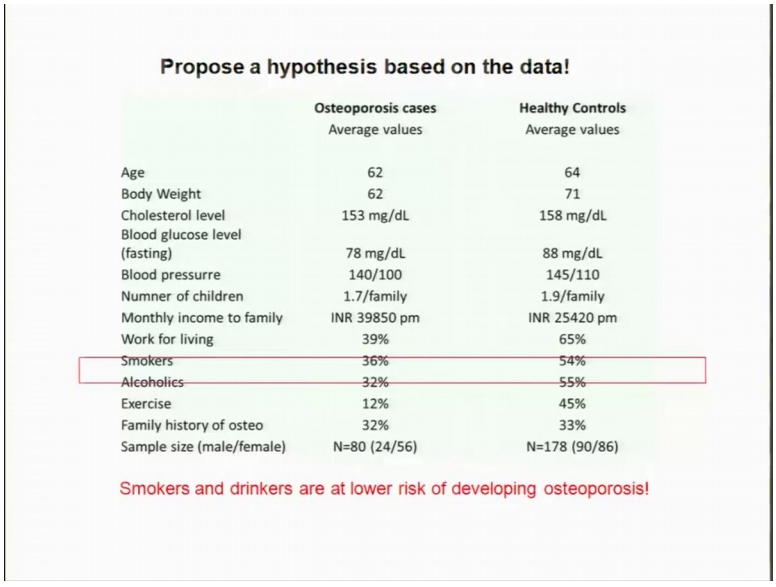
So, let us look in that way first one there is a huge difference between the osteoporosis cases and healthy controls, in terms of the monthly income you can see here the healthy controls you know earned less on an average as compared to the osteoporosis case. In other words higher the based only on the data, then your hypothesis would be higher the family income higher the probability of getting the osteoporosis is that true that you need to analyze the data.

(Refer Slide Time: 16:40)



Second whether the individual who has developed this osteoporosis, whether they work for living or they are just home makers and again you find those who work for living having a job or at lower risk out of getting osteoporosis; that means, you get out of home and then do work is it an hypothesis is a correct hypothesis.

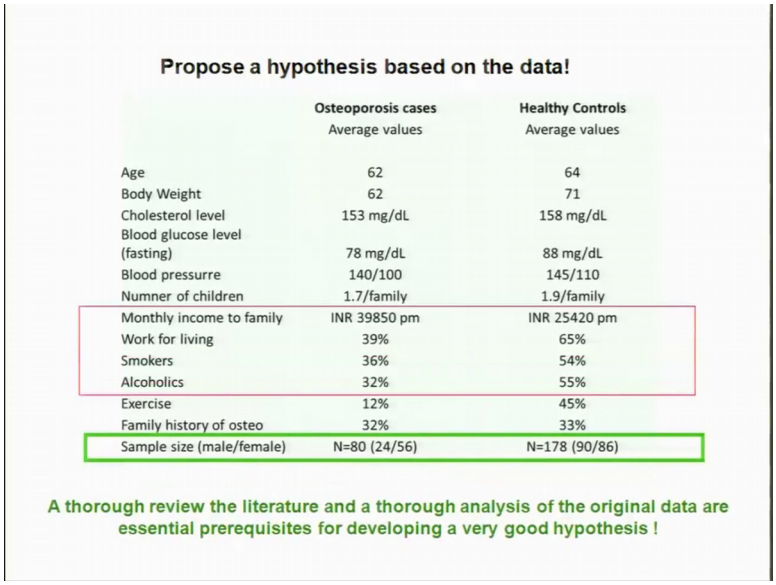
(Refer Slide Time: 17:00)



And third one when you look at smokers and alcoholics the percentage is very high again healthy controls as compared to osteoporosis; that means, smoker's drinkers are at lower risk. In other words you smoke and drink for the riskal developing osteoporosis goes

down, now the data suggests like that, but what do you need to look at is you have to critically evaluate the data because you know is a data do not say anything.

(Refer Slide Time: 17:28)



These are raw data, but you have to analyze what you need to look at is sample size and the for example, the male female ratio really look into the demography then you may not be able to get the real pictures. So, one of the indicators here is that the sample size first osteoporosis is smaller nearly 80 cases that you have looked at, and the control is 178 pretty large.

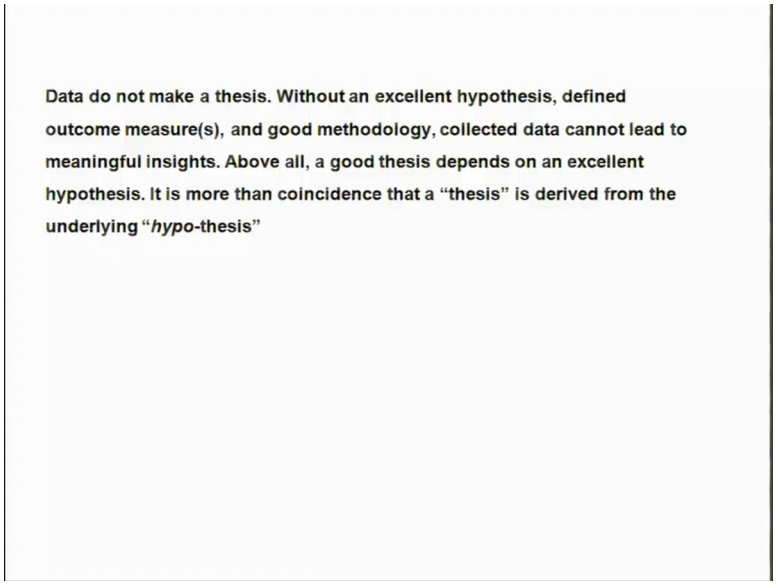
So, basically to avoid any sample by us you should ever had larger sample size, second if we look into the male female ratio in the healthy control it is almost like 1 is to 1, right 40 percent approx there is to female, but if you look in here again it is male female ratio is also very you can clearly see that the female is over represented in the cases as compared to the healthy controls.

So, what does that mean in a typical middle class family what do you expect you know often the mother is at home taking care of the home, she is not going out for a job. So, in that sense she is not working for living, but she is working for everyone else living right. So, she is not earning income and then often you will find that they do not smoke they do not drink, but you may go to the healthy control they have higher proportional than male apart they go out for earning they also drink they have smoke.

So, there is, that is where the bias come in right if you look in for example, again family income here both of them are not working therefore, income is low in the healthy control probably both of them are working the income is high. So, here there is we know that if you are you know if you are a female, human then you are at higher risk or developing osteoporosis because of the physiology and indeed that is what we reflected because there is bias in the sampling.

So, this is what is important when you develop one any hypothesis you simply it is not that you by any data, or by any logics you have to evaluate critically and see whether the data is reliable whether it is done with all the important controls whether I can base my hypothesis based on the data. So, data do not make a thesis remember that when you do any you know research it is not that you have gotten some data where would make a thesis.

(Refer Slide Time: 20:01)



Without an excellent hypothesis which requires very good survive the literature, and very good thought process defined outcome measures you have one hypothesis. And you have measured as to what you want to do and then this is what you are expecting that they have come, in a good methodology because you had asked good questions you have to test it using certain appropriate methodology and you need to have very good data collected.

You have just now seen I have shown a table which is misleading if you do not look in carefully. And you know if you do not have this can lead to you know when if you do not have these it cannot lead to a meaningful insights, you may not really add much unless you have a good hypothesis defined outcome measures and good methodology and good data collection is not going to make a good outcome.

Above all a good thesis depends on an excellent hypothesis extremely important, it is more than coincidence that the, is thesis is derived from underlying hypothesis. So, therefore, you know it is important that you understand hypothesis you spend lot of time on hypothesis for the very simple reason that you cannot start a resource project without a well-defined hypothesis which can be tested which has high predictive value which can be proven correct therefore, you have contributed immensely to a given field.