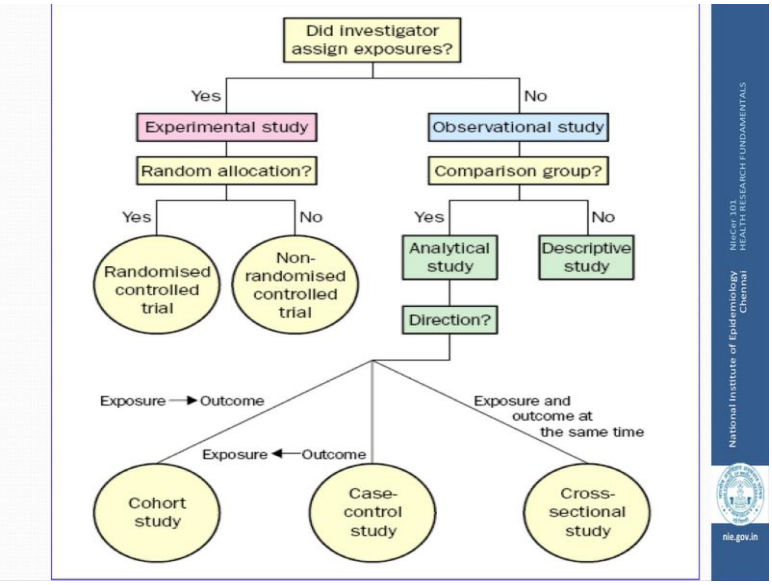
Health Research Fundamentals
Dr. Manoj Murhekar
ICMR - National Institute of Epidemiology, Chennai

Lecture – 06
Analytical study designs

Hello and welcome. My name is Manoj Murhekar and in today's lecture I will give you an overview of cohort and case control studies. In earlier lectures, my colleagues have given you an overview of different study designs.

(Refer Slide Time: 00:27)

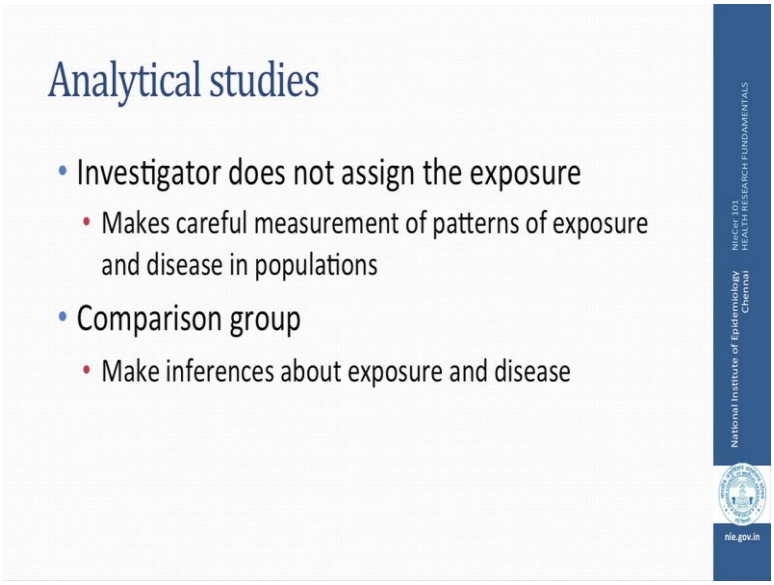


Let us have a quick recap. Epidemiological studies are broadly divided into 2 categories; first is Experimental studies and second is Observational studies. And this categorization is based on these questions; did the investigator assign the exposure? So, in experimental studies investigator assigns the exposure and this exposure could be in terms of new intervention, new drug or new vaccine. These studies are further classified into Randomized and Non-randomized studies, based on Random allocation of exposure.

On the other hand, in observational studies investigator does not assign the exposure. If there is no comparison group in observational studies, such studies are called as

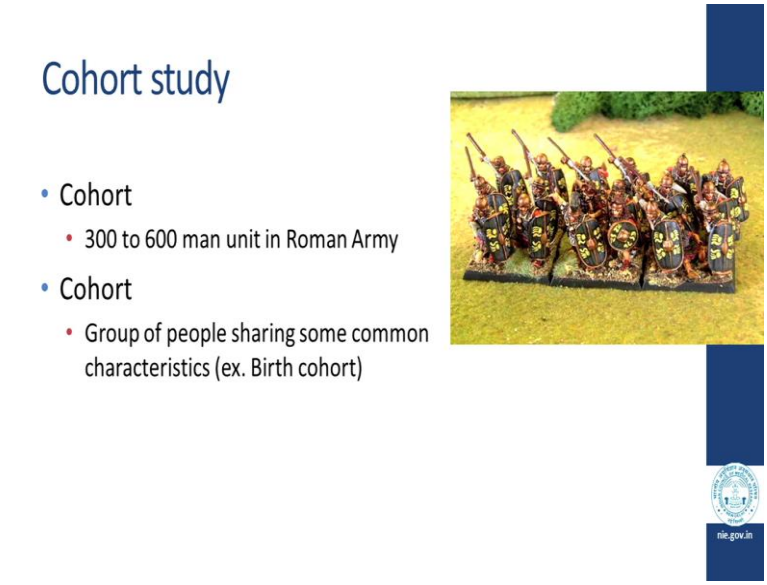
Descriptive studies. And in these studies, we described health event in terms of time, place and person. If there is a comparison group in observational studies, the studies are called as analytical studies, which are further divided based on the direction of the studies. Cohort studies they progress from exposures to outcome, whereas case control study progress from outcomes to exposure and in cross sectional studies we measure exposures and outcomes at same time.

(Refer Slide Time: 01:53)



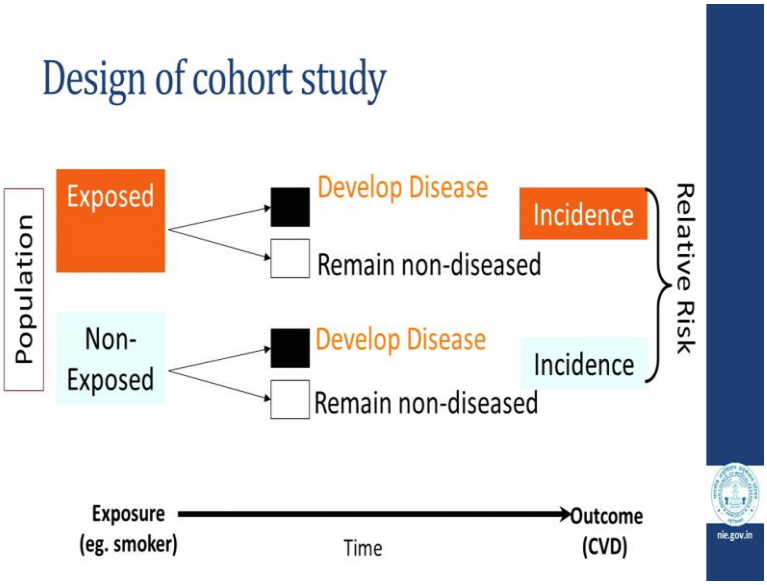
So, in short analytical studies are the one in which investigator does not assign the exposure, there is no randomization. So, what investigator does essentially is he carefully measures the pattern of exposure and disease in population. There is a comparison group in analytical studies and using this comparison group the investigator next inferences about exposure and the disease.

(Refer Slide Time: 02:20)



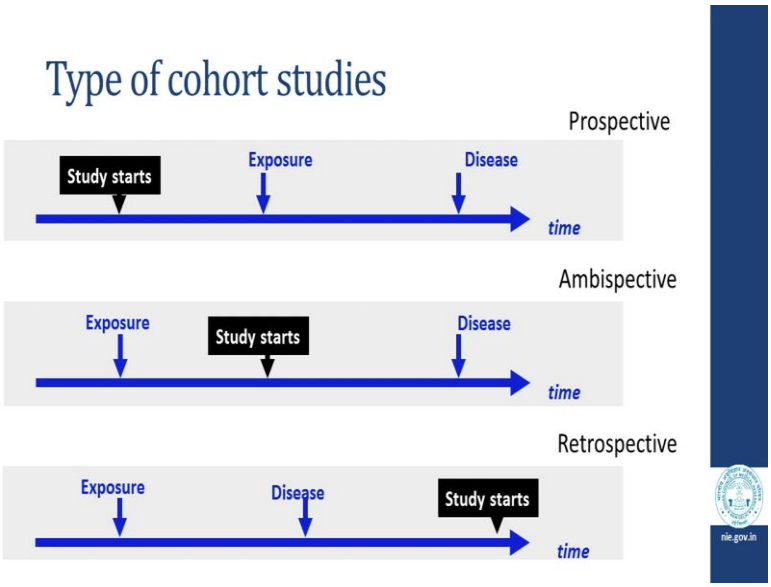
Let me first talk about Cohort study. The word cohort has a military origin, military roots rather than medical routes. In Roman army, a 300 to 600 man unit was called as cohort, whereas in epidemiology, the word cohort is a group of individuals sharing some common characteristic; one such example could be a birth cohort, all the children who are born today will form today's birth cohort.

(Refer Slide Time: 02:55)



Let us see, how the design of cohort study is. As we know, cohort study will progress from exposure of outcome and in this particular example exposure is say, cigarette smoking and outcome development of cardio vascular disease. Cohort studies begin with selection of Exposed and Un-exposed cohort, and in this example, it would be people who are the cigarette smoking and those are not smoking. Once we identify this cohort, these cohorts are followed in time. Some of this exposed and non-exposed individuals will develop the disease that is cardio vascular disease, whereas the remaining people would remain non diseased. We will then calculate the incidence of cardio-vascular disease in exposed population and in unexposed population. And we will compare this incidence using a measure of association called as relative risk. I will talk about this relative risk later.

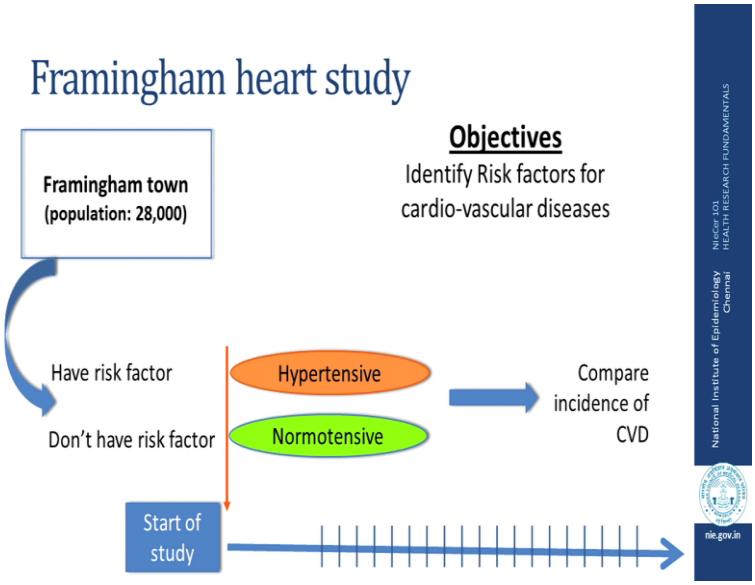
(Refer Slide Time: 03:59)



There are 3 types of cohort studies; the first is Prospective cohorts study. In the prospective cohort study, by the time your study starts exposure and disease has not yet occurred. Whereas in case of your Retrospective cohort study, both the exposures and disease has already occurred when you start the study. And there is a combination of these two approaches, which is called as Bidirectional study or Ambispective study, wherein when your study starts, the exposure has already occurred and then you follow this exposed and un-exposed individuals, till they develop the outcome. Let me explain

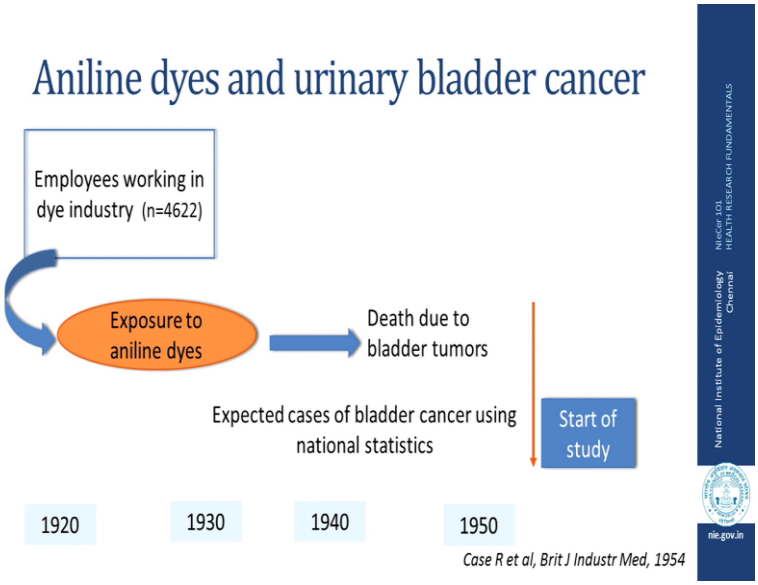
these different types of studies by giving some example.

(Refer Slide Time: 04:51)



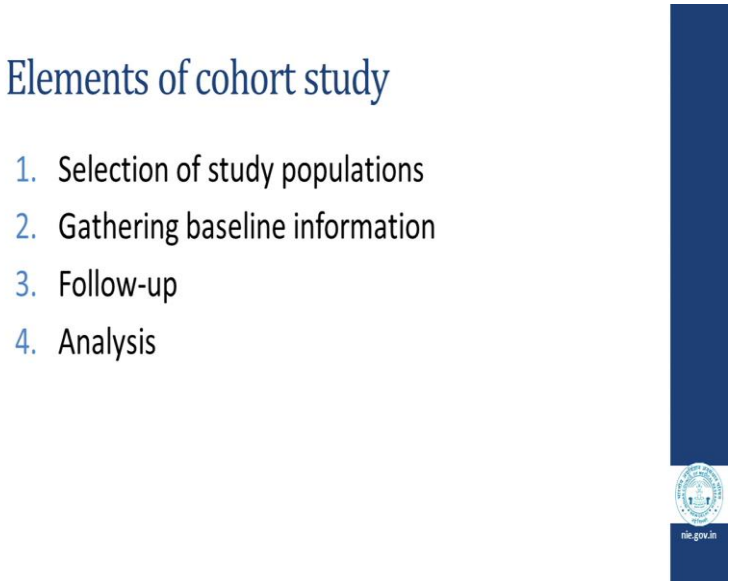
First example is of Framingham heart study, which is one of the oldest cohort study initiated in 1940s. The objective of this study was to identify risk factors for cardio-vascular diseases. This study was conducted in a town of Framingham, which had a population of about 28,000. So, this population; in fact, the sample of this population was then divided into 2, based on those having risk factor and those who was not having the risk factor. And the investigator considered several risk factors, one of which was hypertension. So, for the purpose they classified this population into those who had hypertension and those who did not have hypertension. This cohort was then followed up in time and the incidence of cardio-vascular diseases was compared in two cohorts.

(Refer Slide Time: 05:50)



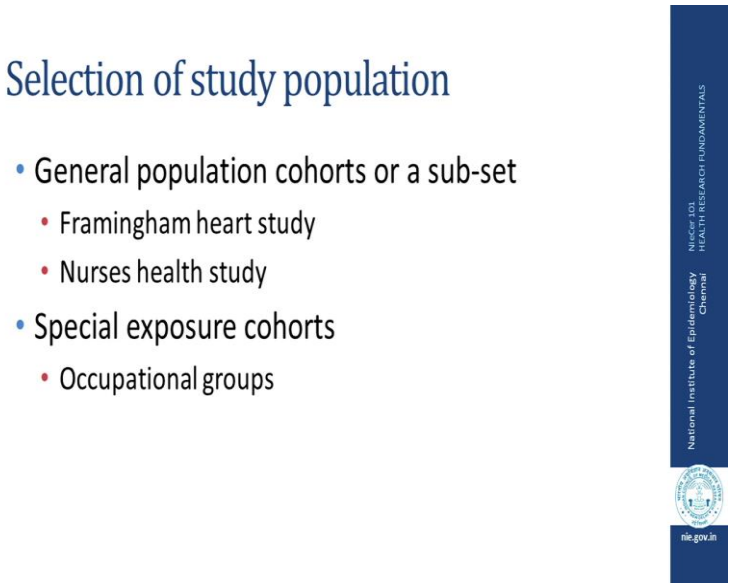
This is an example of a Retrospective cohort study. The objective of this study was to evaluate the role of Aniline dyes or exposure to aniline dyes and development of urinary bladder cancer. So, the investigator for this study recruited about 4622 workers who were working in dye industry between 1920 and 1951. So, this recruitment was based on available records in those factories. Investigator also reviewed the death records of this 4622 individuals and essentially, they looked about any mention of urinary bladder tumors on their death records, and then they compared death rates in these population with that of expected number of deaths of bladder cancer using national statistics. So, by the time the studies started, both exposure and outcome had occurred.

(Refer Slide Time: 06:58)



So, these are the 4 important components of cohort study. First is selection of study population, second is gathering baseline information, third is following up this cohort and fourth is doing analysis.

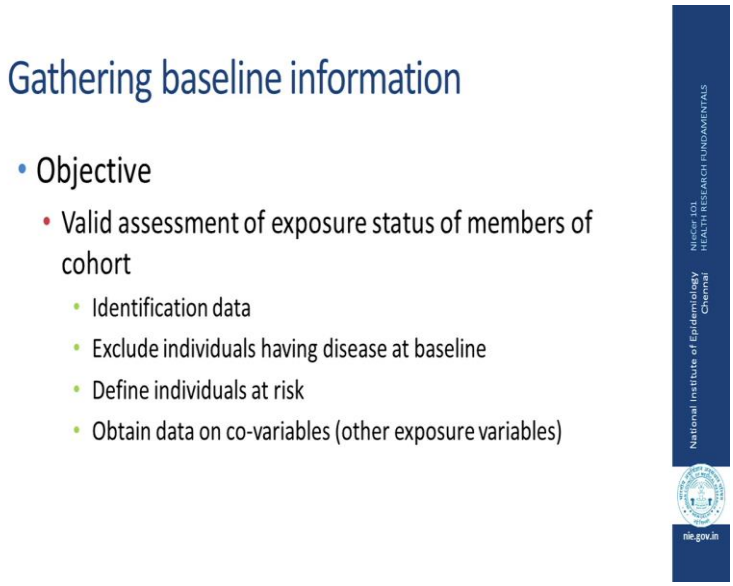
(Refer Slide Time: 07:15)



There could be 2 approaches of selecting study population. You could select your cohort

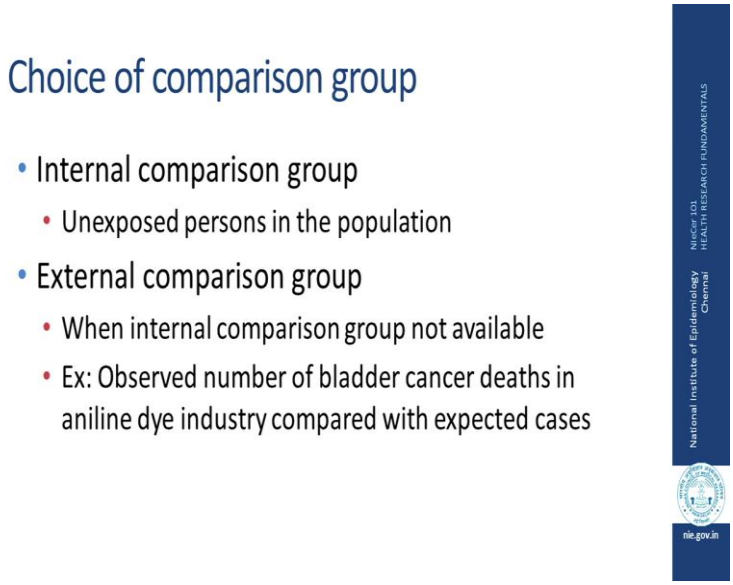
from general population as was done in case of Framingham study or you could select a subset of general population as was done in nurses' health study. The second approach could be selecting a special exposure groups, such as occupational groups.

(Refer Slide Time: 07:39)



Once you select this study population, the next important state is collecting baseline information from this population and the objective of this step is to have a valid assessment of exposures status of members of cohort. And by doing this baseline information we can also collect identification details of the study population, we can exclude those individuals, who are having the disease of interest at baseline so that the population which remains is at risk of developing the disease and we can also obtain the data about other risk factors or other exposure variables.

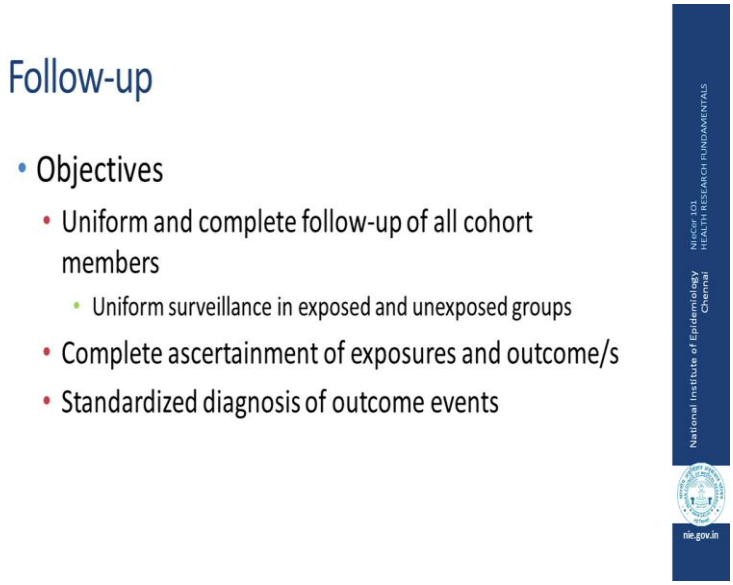
(Refer Slide Time: 08:18)



As we have seen in analytical study there is a comparison group and we have 2 options of having comparison group in case of cohorts. One is internal comparison group and the unexposed persons in the population is taken as an internal comparison group and example could be Framingham cohort study. Wherein, those who had hypertension were considered as exposed population and those who were normotensive was unexposed population.

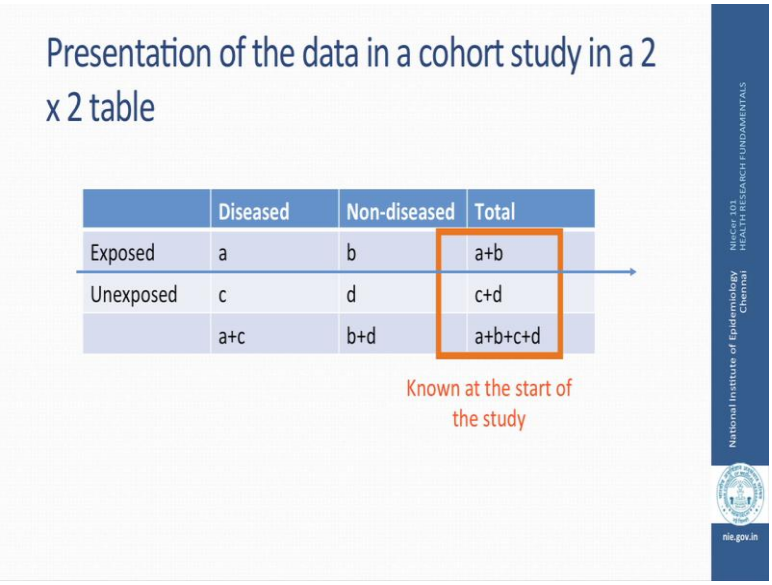
Sometimes it is not possible to have internal comparison group. As we have seen in the case of aniline dye example, everybody in those factories were exposed and it was not possible to have a internal comparison group. And therefore, the investigators compared the death rates with that of general population, so you could take an external comparison group in such situations.

(Refer Slide Time: 09:18)



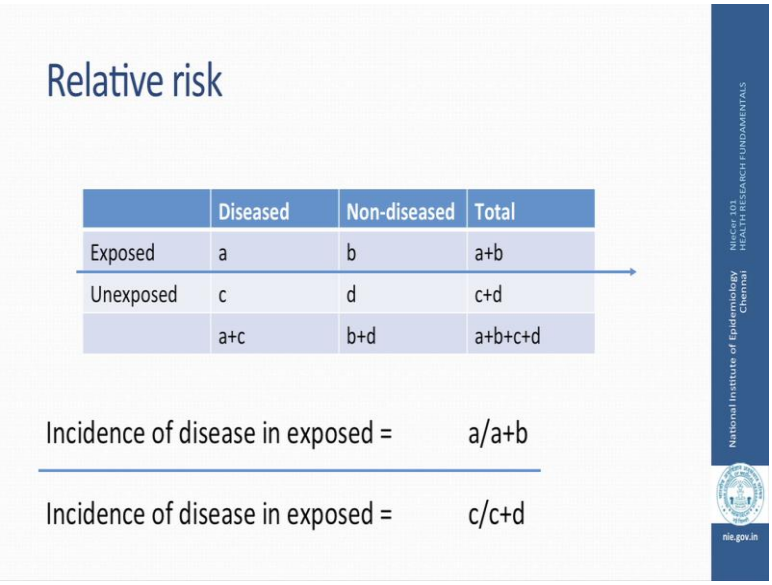
Once you recruit your exposed and unexposed population, the next very important step is doing a good follow-up of these populations. There could be 3 principles of having good follow-up, first is having a uniform surveillance in exposed and unexposed group; having complete assessment of exposures and outcomes and third is using a standardized diagnosis of outcomes, especially since the cohort studies can last for a long period of time.

(Refer Slide Time: 09:50)



This is how the data in cohort study would look like this is how the 2 by 2 table in cohort study would look like. We started the study with selecting people who are exposed, which is a + b and people who are unexposed which is c + d. And we followed this people so a + c developed the disease and b + d remind non-diseased. So, at the beginning of the study we know who were expose and who were not exposed.

(Refer Slide Time: 10:22)



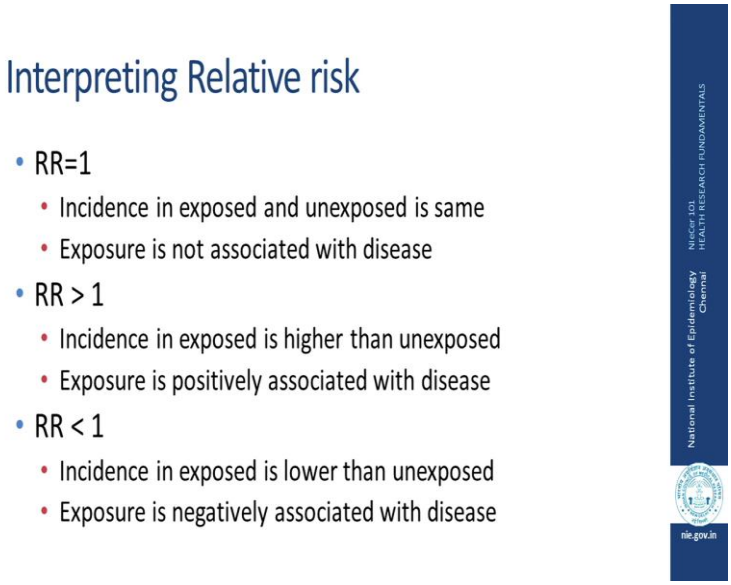
So, we can calculate incidence of disease in exposed population which can be given by formula

Incidence of disease in exposed = a/ a+b and

Incidence in unexposed population = c / c+d

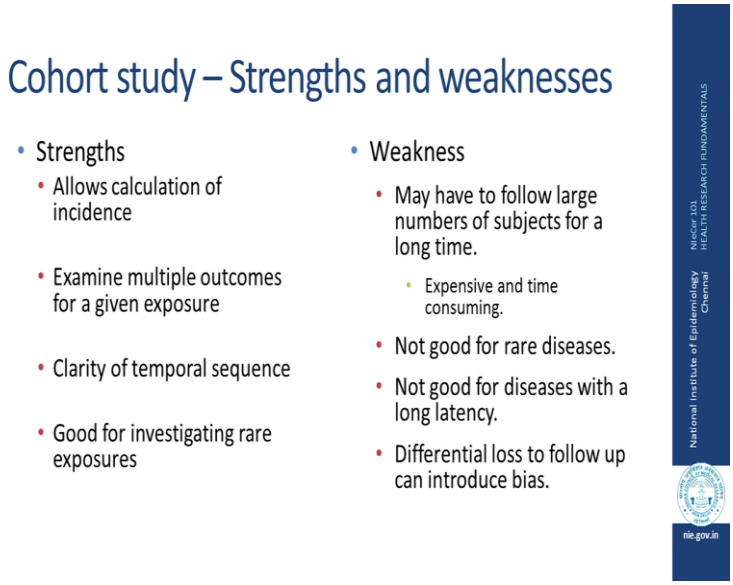
And the ration of these 2 incidence is a relative risk.

(Refer Slide Time: 10:42)



How do you interpreting this relative risk? There could be 3 possible scenarios of relative risk, one is relative risk is equal to 1. If your relative risk is 1, it means that incidence of disease in exposed and unexposed population is same and we can interpret that the exposure is not associated with the disease. Relative risk could be more than 1, which means that incidence of disease is higher in exposed population as compared to unexposed population and we can interpret that the exposure is positively associated with the disease. Relative risk can also be less than 1, which means that incidence of disease in exposed population is lower than unexposed population and here we can interpret that exposure is negatively associated with the disease.

(Refer Slide Time: 11:39)



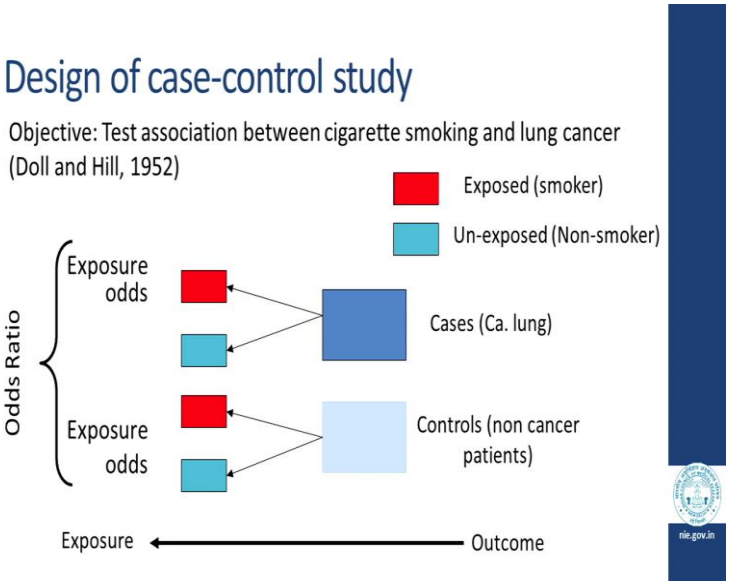
Cohort studies have certain strengths as well as certain weaknesses. So, what are their strengths? They allow calculation of incidence because when we start the study, we start the study with selecting exposed population and unexposed population and we follow them in time therefore, it is possible for us to calculate the incidence of disease. We can examine multiple outcomes for a given exposures. We are very confident about temporality in case of cohort study, and last is that this studies are especially useful for rare exposures.

So, what are the weaknesses? The sample size for cohort study could be very large; we need to follow these people for a very long time and therefore, cohort studies could be expensive and time consuming. They are not recommended for diseases which are rare or diseases which have very long latency, and if you do not have good follow up or differential follow loss in exposed and unexposed population it could introduce certain amount of bias in your study.

(Refer Slide Time: 12:53)

Let us now see Case control study. Case control studies are exactly opposite to that of cohort study when it comes to the direction or logic of the study.

(Refer Slide Time: 13:06)

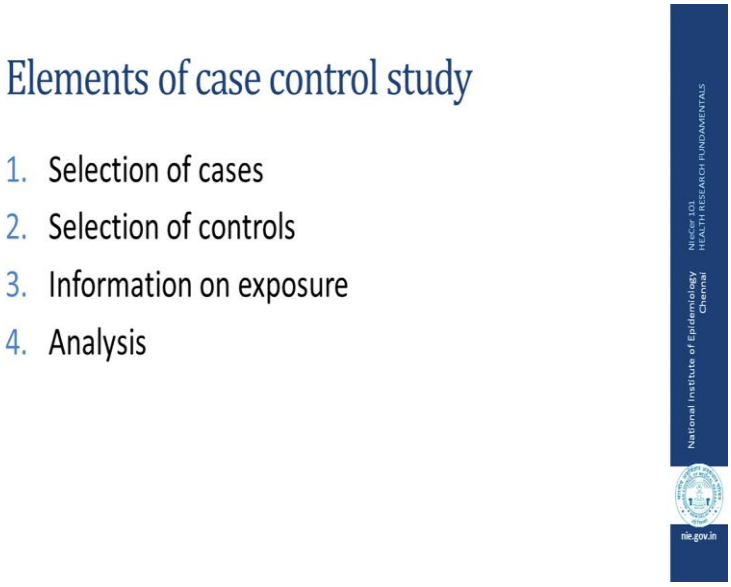


Let us first see the design of cohort study. And to explain the design, I will use the example of one of the very old case control study conducted by Doll and Hill. The

objective of this study was to test the association between cigarette smoking and lung cancer. As the name suggests case control, we start with selecting cases and control is the one who does not have the disease in question. So, the first step is selecting cases, so for this study Doll and Hill selected lung cancer cases, who are admitted in hospital in about 20 hospitals in London. These cases were all histopathologically proven cases of lung cancer. So, for each case they selected a control which was a non-lung cancer patient admitted in the same hospital and these cases and controls are then interviewed to find out their prior exposures.

So, Doll and Hill found out how many of the cases were cigarette smokers, they had a detailed questionnaire to ask about history of cigarette smoking. They asked how many of them were smoking. What is the age of starting smoking? What type of cigarette they were smoking? And so on and so forth. So, we find out how many of the cases are exposed? How many of the controls are exposed? And same way how many of the cases are unexposed? And, based on this data we calculate what is called as Exposure odds among cases and Exposure odds among controls and then we calculate what is known as odds ratio as a major of association between exposure and outcome.

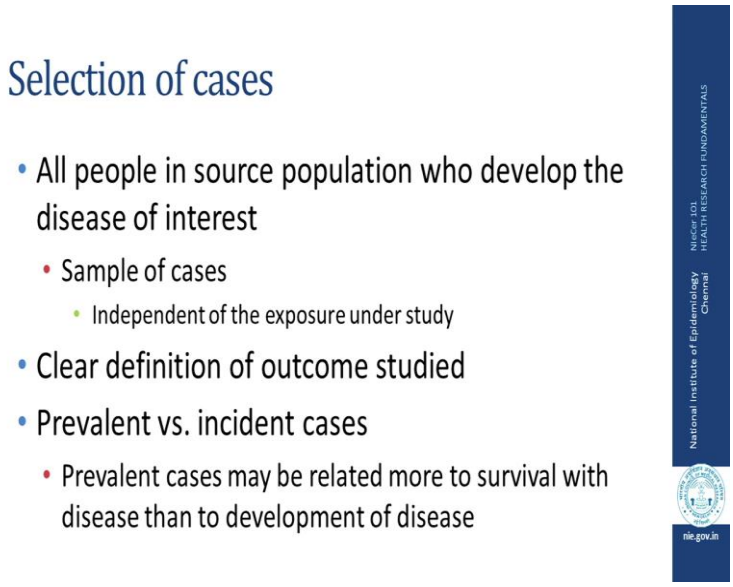
(Refer Slide Time: 14:50)



Like cohorts study, there are four important elements of case control study. First is

selecting cases, second is selecting controls and third is collecting valid information about exposures and then doing the analysis.

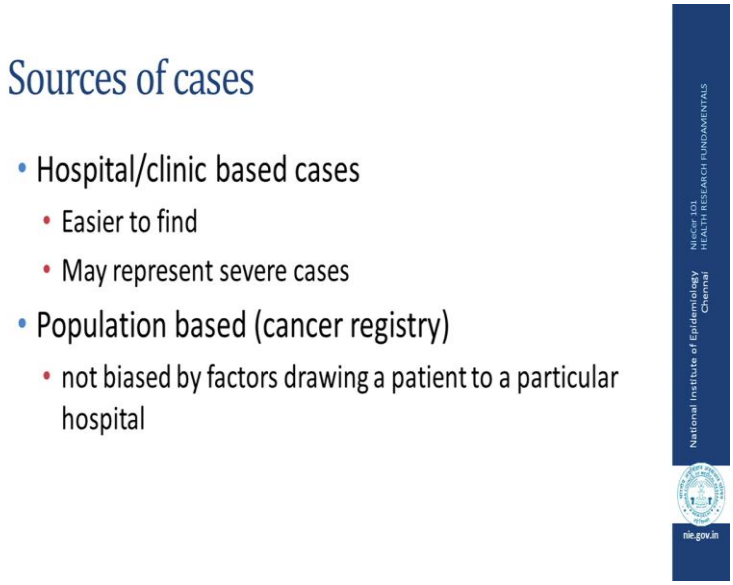
(Refer Slide Time: 15:03)



Coming to the selection of cases, theoretically all people in the source population who develop the disease of interest could be included in your study or you could sample these cases. However, one thing you should keep in mind that, the selection of cases should be independent of exposure under study. We need to have a clear definition of outcome to be studied. One also need to decide whether to include prevalent cases or should include only incident cases.

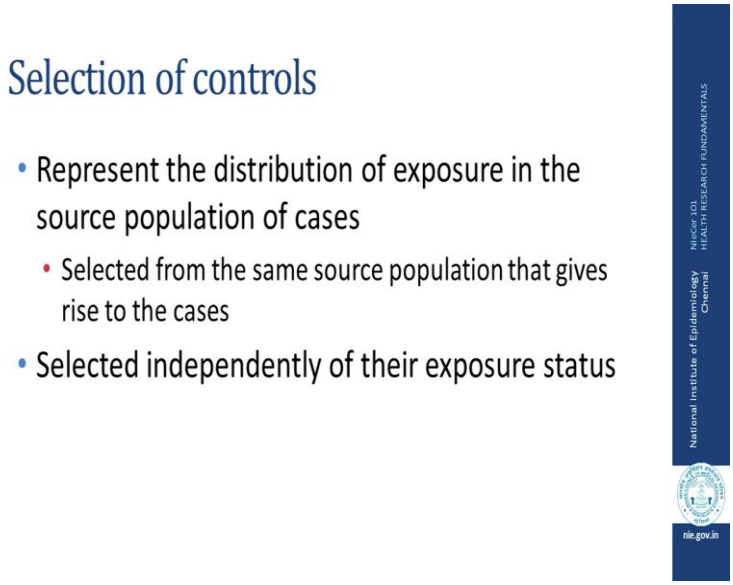
Prevalent cases mean those cases which already occurred in the past, whereas incident cases are the newly occurring cases. So, if you take a prevalent case they are readily available and by including them we can save our time and money. But, in spite of this obvious advantage it is generally recommended to include incident cases, mainly because prevalent cases maybe related more to the survival with the disease than the development of disease.

(Refer Slide Time: 16:13)



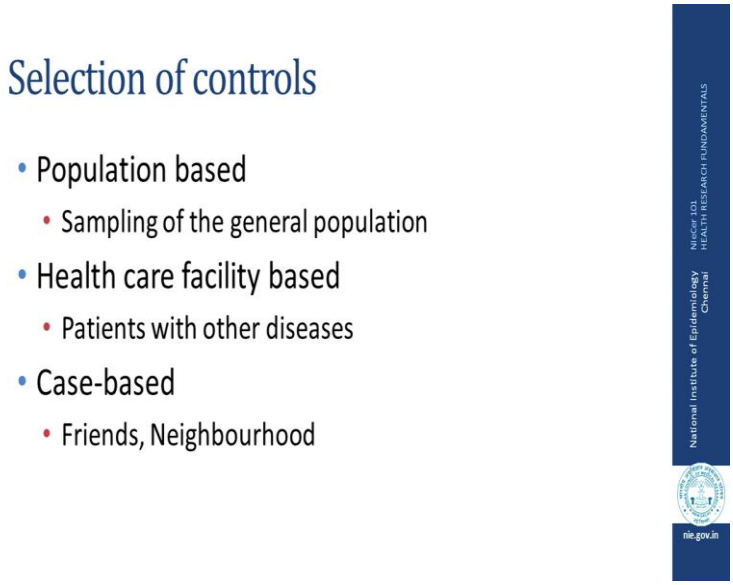
Where from I can select these cases? Again, there could be two important sources; one is from hospitals or clinics, it is easier to find cases in this hospitals and clinics; however, it is quite possible that cases which are admitted are more severe cases and may not represent the cases in community. The other approach could be a population based selection of cases and one such example could be cancer registry and these cases are more likely to represent the source population, primarily because they are not biased by factors drawing patient to a particular hospital.

(Refer Slide Time: 16:52)



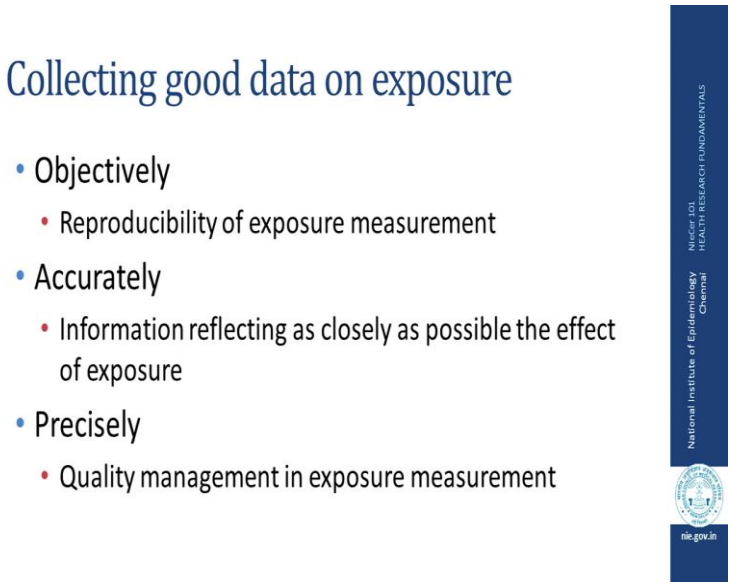
Where from I can select the controls. As I mentioned earlier, control is the one who does not have the disease under investigation. Why do we need control? Controls essentially represent the distribution of exposure in source population. So, they generally tell you the background rate of exposure in the population from which cases have come. Like cases, they also need to be selected independent of their exposure status.

(Refer Slide Time: 17:30)



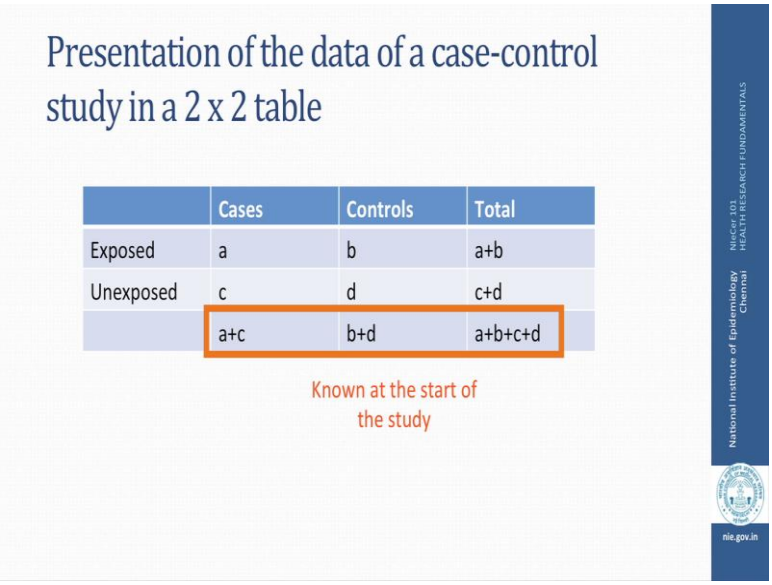
There again could be 3 sources of controls, first is the population based controls and you could sample from general population. Second is you could select controls from health facility and in case of Doll and Hill study they selected control from the health facility, but we could select patient with other diseases. And the third source of controls could be case-based controls that are from friends or neighborhood.

(Refer Slide Time: 17:59)



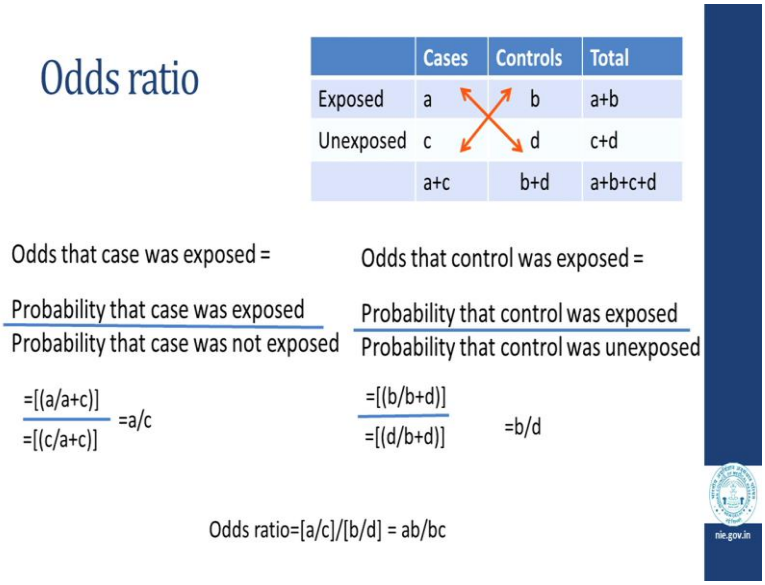
Once you select cases and control, the next important step is collecting data about past exposures. And again, there are 3 important principles collecting the data on exposures objectively so that your measurements are reproducible, accurately and precisely.

(Refer Slide Time: 18:19)



Once you collect the data on exposures, this is how your 2 by 2 table will look like. This is exactly the same table which we saw for the cohort study. However, in case control studies, when the study started we knew who were case and we knew who were controls. So, a + c, were cases to start with and b + d were the controls. And we found out that of a + c cases, a were exposed and c were unexposed and same way b + d controls b were exposed and d were unexposed. In case control study we cannot calculate incidence of disease, like what we could calculate in case of cohort studies.

(Refer Slide Time: 19:06)



So, what do we do? Then what we do is, essentially we calculate a measure of association called Odds ratio. This odds ratio is the odds that cases was exposed is given by this formula:

probability that case was exposed

---

probability that case was not exposed.

And we know that probability a case was exposed is: a / a + c

and probability that the case was not exposed is: c / a + c

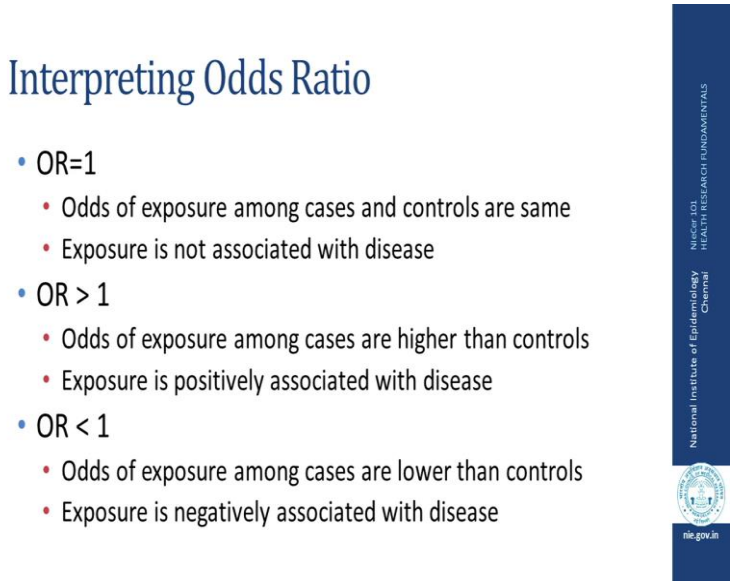
and the ration of these 2 probabilities is: a / c.

Same way we also calculate the odds that control was exposed, which comes out to be: b / d

and the ratio of these 2 odds becomes odds ratio which is ab / bc, which is nothing but a

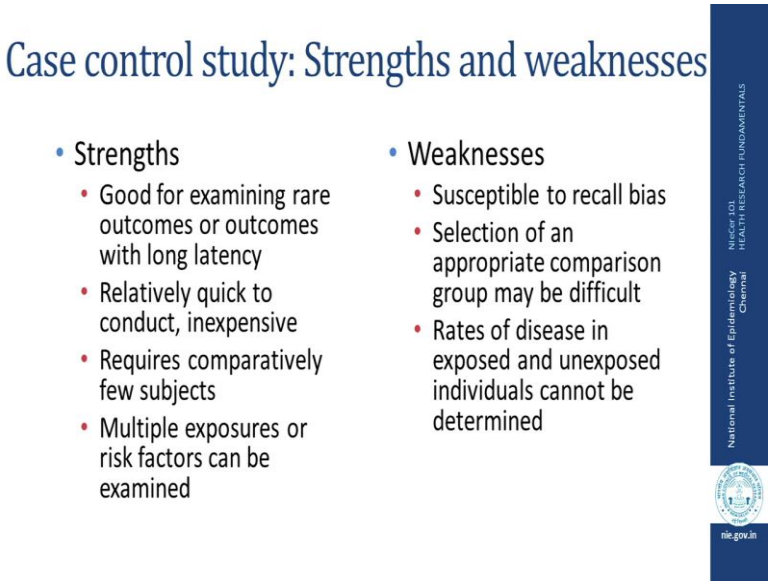
gross product ratio.

(Refer Slide Time: 20:03)



How to interpret this odds ratio? Again like relative risk there could be 3 scenarios, one is odds ratio equal to 1. If odds ratio is equal is to 1, it means that odds of exposure among cases and controls are same and we can conclude that exposure is not associated with disease in such situation. Odds ratio could be more than 1, it means that your odds of exposure among cases are higher than that of controls and we can conclude that exposure is positively associated with the disease. If odds ratio is less than 1, it means that odds of exposure are among cases are lower than that of controls and we can conclude that exposure is negatively associated with the disease.

(Refer Slide Time: 20:51)



Case control study also has certain strengths and weaknesses. These studies are especially good, if the outcome is rare or the diseases have the long latency period. They are fairly easy or quick to conduct and hence inexpensive. Requires relatively less subjects than that of cohort studies and multiple exposures or risk factors can be examined at the same time. The weaknesses of cohort study include that they are susceptible to several biases, the recall bias one of the most important bias. Sometime selection of control could be a problem, selection of an appropriate comparison group may be difficult and we cannot calculate incidence or disease in these studies.

Thank you.