Health Research Fundamentals
Dr. P. Ganeshkumar

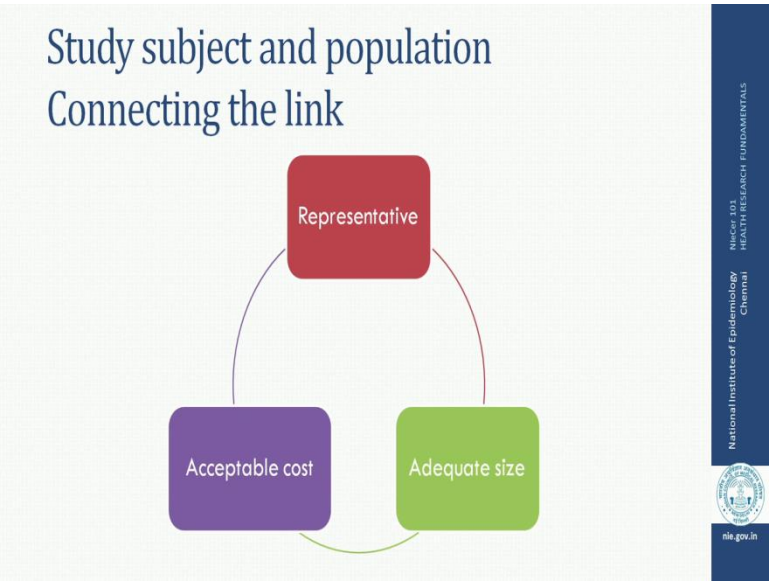
ICMR-National Institute of Epidemiology, Chennai

Lecture - 13
Selection of Study Population

Hi! I am Dr. Ganeshkumar from ICMR School of Public Health, National Institute of Epidemiology. Today, we are going to have a lecture on Selection of Study Population.

A good choice of study subjects serves the vital purpose of ensuring that the finding in the study accurately represents the population of interest. So, that is an important part, how the selection of a study population gathers the right information about the health research over the population of interest.

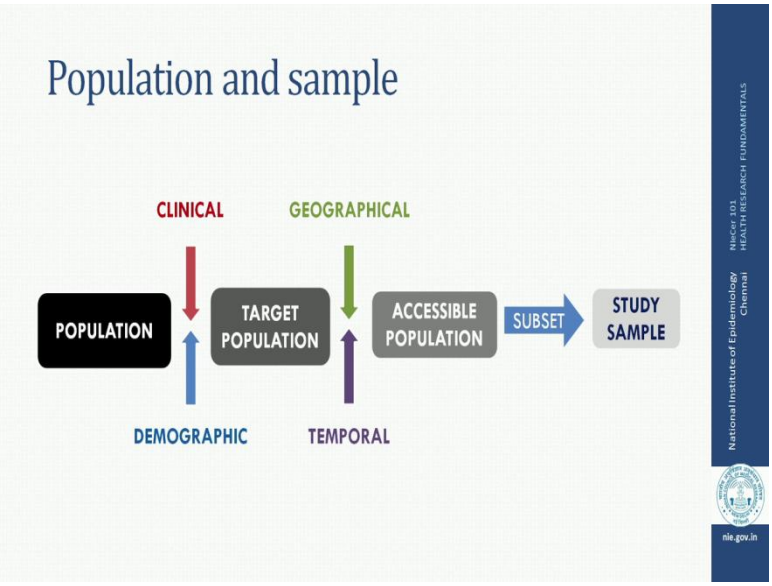
(Refer Slide Time: 00:43)



See, when we are going to a select a study population or a selection of subjects from the population of interest. There are 3 important things which we have to remember particularly; first number one is that, we have to select the study subjects at an acceptable cost in terms of time and then money. Second, we need to have an adequate size of the study population so that it controls the random error, the adequate size is most important. Number three, your study population; the study subject should be representative enough to the population of interest so that your findings can be generalizable to the population of interest.

That is how in this video lecture we will be seeing, how to select a study population with a good representativeness? And we will be also discussing about what are all the issues and what are all the solutions of selecting a study population? And what are the recruitment strategies in a clinical research? Whereas, how to achieve your adequate sample size and the process of sampling in a research will be dealt in the separate lecture. So, this lecture will not be covering about what is adequate size? and As well as, what is the techniques or sampling techniques in selectioning the study population?

(Refer Slide Time: 02:14)



Ok, In a health research, in a clinical research or a public health research, when you are going to select a study population, these are all the terminologies which we need to understand, which ideally happens over the process of doing a research. For example, in a layman term, a population is a larger area for example, population of India or the population of Tamilnadu, a larger geographical area. From that population by setting up or for defining certain clinical under demographic characteristics, what you are trying to derive is called a target population. So, your target population is determined by certain clinical and demographic strategies, characteristics.

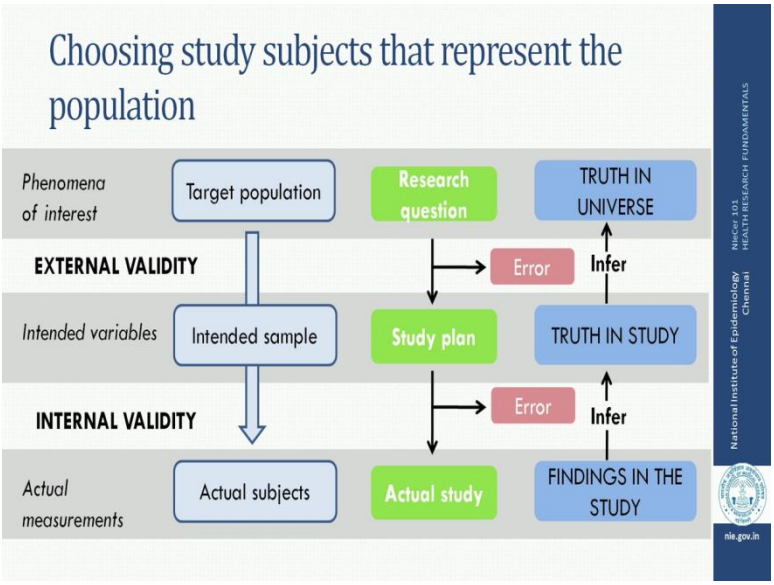
By applying certain geographical and as well as temporal characteristics over there temporal target population, what you are trying to derive is a called accessible population. So, if from these accessible population by means of certain subset of this accessible population, what we are trying to derive is called Study sample. So, this is

how a process of deriving a study sample from the population. So, from this step you can able to understand that two important terminologies; one is called Target population which is defined by clinical and demographic characteristics, second is a Accessible population which is defined by geographical under temporal characteristics of the subset of the target population and from the subset of the accessible population what we derive is a study sample.

I will give one example for these; for example, our target population which based on the research question, if the research question is like I want to study about what is a low dose of metformin to reduce the dysmenorrhea among PCOS females in reproductive age group. So in that here, when you specifically see that the clinical and demographic characteristics are those PCOS females in the reproductive age group and they are sufferings from polycystic ovarian disease and they should be having a clinical feature of dysmenorrhea. And, the geographical and temporal characteristics are defined by the accessibility of the population; this is the subset of the target population.

For example, when I am going to conduct in a city and those patients who are attending my OPD are my accessible population and I am recruiting them for this study. So, this defines the geographical and temporal is that from January 1 to December 31st of a year, a specified year. So, all the patients who are attending my OPD's with this clinical and demographic characteristics in this period will be recruited for my study. And from the, I estimated sample size, I will be recruiting those subset of accessible population called study samples.

(Refer Slide Time: 05:10)

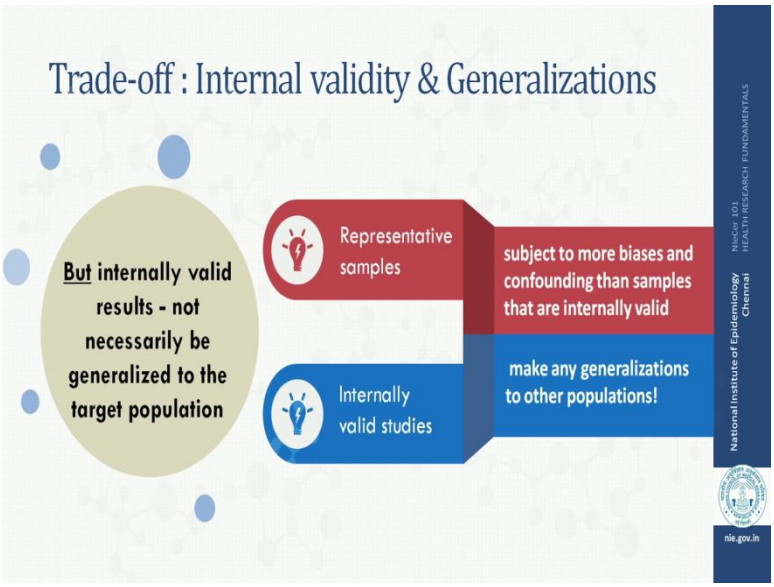


So, let us see how the journey of choosing the study subjects that represent the population happens. So, I am just throwing a kind of an algorithm here which will explains you certain two important terminologies called External validity and Internal validity. For example, by means of a phenomena of interest; that means, by your research question your setting up certain clinical and demographic characteristics what to derive is a target population, and from that you have an intended sample and that intended sample you are measuring certain intended variables and what you derive is a actual subjects, that is study subjects, and where you do your actual measurements. So, same how you do is that, you do your research, you choose your research question of the phenomena of the interest and you have a study plan of conducting the study and measuring it with an actual measurements and that is your actual study here.

So, how it happens in reverse? Say, from the findings of your study, from those actual subjects what you infer is called a truth in the study and this truth in the study is what you are trying to infer over the truth in the universe, so this goes in a reverse. So, you are trying to generalize your study findings over the population of interest and that is where you have depicted as truth in the study and truth in the universe. So, here you can clearly note that there are certain errors, which may happen when you are deriving the subjects from the target population to the actual subjects. And previously, I explained that these kinds of random errors can be handled by an adequate size of the sample population.

Now, you can clearly see here that, where this external validity and internal validity lies. So, the internal validity is a kind of a degree that how far the dependent variable influenced by the independent variables and this is very much consistent within the study subjects. If it is very much consistent and that has a good internal validity and same, when this derive, mean this findings or we are going to generalize to the population the generalizability is called that external validity, where the findings of the study are applied over the population of interest. So, internal validity is very, very essential. So, that is a most important part when we are trying to generalize the results.

(Refer Slide Time: 07:50)

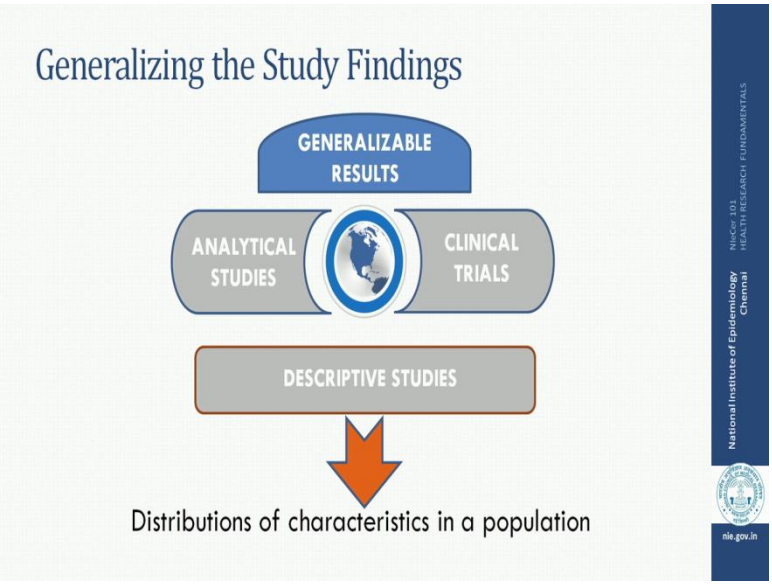


So, now you could able to understand that there is some kind of trade-off between internal validity and generalizations. So, what happens when we are going to select representative samples from the study population? When we are choosing representative samples from the population then they are subject to more biases and confounding than the samples that are internally valid. Then what happens when those studies with good internal validity? So, it makes any generalizations to other populations, but there is a hinge in that, what it is?

The internal valid results, are not necessary be generalized to target population. So, that is how and I am trying to explain here that there is always a trade-off between internal validity and generalizations. So, generalizability hardly a categorical answer of yes or no, it is very rare. It has many trade-off, it is kind of a mix where scientific unpractical

decisions has been taken when we are choosing the study subjects and when we are going to infer the study findings and applicable to the population of interest. So, it depends upon the design, it depends upon the method and also it depends upon the specific research question and where we are applying to derive the findings.

(Refer Slide Time: 09:18)

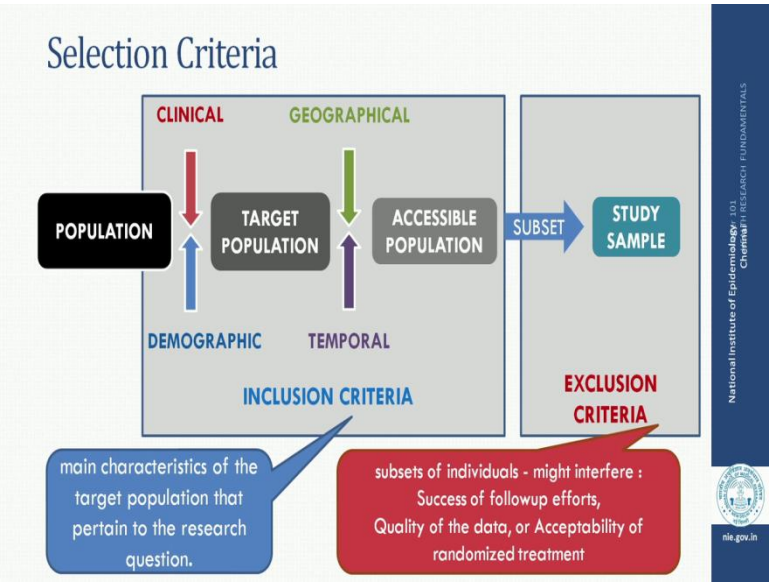


In general, generalizing the study findings can be very valid can be easily happen in clinical studies and clinical trials and analytical studies. So, generalizable of these study findings is very common. So, very popular example is say Framingham Cohort Study. So, in that Framingham cohort study what they have selected is from your individuals from your county called Framingham. From that, when they are selecting it whether the findings, whichever the findings of the study can be generalizable to the whole population of United States of America? No, it depends upon the study findings.

What they have identified in Framingham is that, the association between those cardio-vascular risk factors or the association, the strength of association of hypertension has a cardiovascular risk factor is consistent not only to the population but also to other kind of genetically different ethnical group called Americans, Africo-Americans and all those things. Whereas, when we are trying to generalize certain other findings the prevalence of hypertension, which is identified in this Framingham study may not be generalizable to other ethnicity.

So, that is how descriptive studies always have an issue that, because there the study results cannot be generalizable as such. It is specific to the population which it has been studied and because descriptive studies mainly study the distribution of the characteristics of the population which may not be generalizable as like that when we are doing in analytical studies or a clinical trails. So, now, you can understand that generalizing the study findings is based on even the study design and the type of method and how the findings depends upon the findings, whether it is an analytical study findings or it is a descriptive study findings, it is very important.

(Refer Slide Time: 11:31)

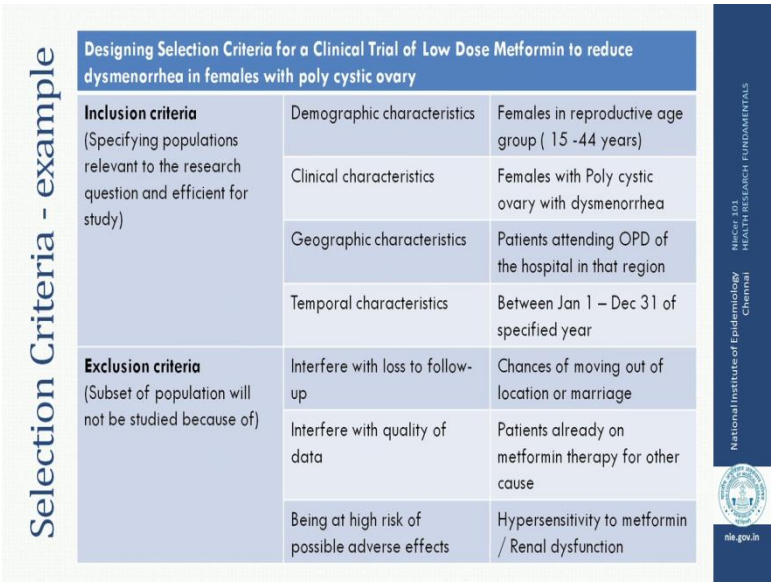


Let us see that, the other important part of selecting the study population with health research is called selection criteria. So, what is that selection criteria? So, this flow chat which you have understood already, which we have discussed already is about that how we derive a target population by a set of clinical and demographic characteristics and how we derive the accessible population, which is the subset of the target population by means of geographical and temporal characteristics. So, the terminology called inclusion criteria and exclusion criteria is based on this, how? See, inclusion criteria is the main characteristics of a target population that pertain to the research question. So, that main characteristics means here the clinical characteristics, demographic characteristics and often the selection criteria in terms of geographical and temporal characteristics are defined in the inclusion criteria, who has to be included in my study. So, it depends upon these characteristics.

Now, what is that external exclusion criteria. So, in exclusion criteria we are deciding in our study that who are those subjects, which we are not going to include because they might interfere in the success of the follow-up efforts or they might interfere with the quality of the data and they may interfere with the acceptability of the study or even there will be interfering with the ethical concerns. So, those study subjects or those groups of people with those characteristics, we do not want to include in our study is defined as that exclusion criteria.

So the step as such is that; first, we need to define a specific inclusion criteria based on the specific inclusion criteria, what you derive is a accessible population and by means of an exclusion criteria you are neglecting or you are avoiding to include certain individuals, so that what we are trying to derive is a subset of your accessible population called study samples, with an estimated sample size in it.

(Refer Slide Time: 13:52)



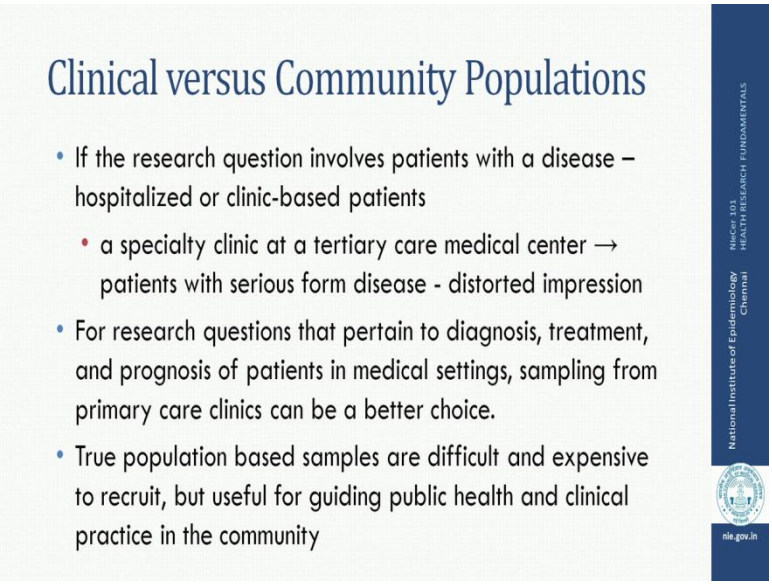
So, more in that inclusion criteria and exclusion criteria, here I am giving an example what we have discussed already. See, the study, designing a selection criteria say for a clinical trial of low dose metformin to reduce dysmenorrhea in females with the polycystic ovary. Here, let us see what how we can fix up our inclusion and exclusion criteria? So, the inclusion criteria, as we discussed already that the main characteristics of the population those who have to be included in my study, which is relevant to the research question and which will be very efficient to the study. So here demographic

characteristics, I will include females in the reproductive age group of 15 to 44 years; yes, it defines that particularly and the clinical characteristics are those reproductive age group females with polycystic ovary and suffering from dysmenorrhea.

And the geographical characteristics, patients attending the OPD of the hospital where the study is going on in the particular region and temporal characteristics is that, between the period January 1 to December 31 of the specific year. So, temporal is that time. With this set of inclusion criteria, now in this study what are the exclusion criteria I am fixing up, a subset of the population which should not be included, which cannot be included in my study. See, when I am concerning with those subjects who will interfere with the loss to follow up or chances of moving out form the study area may have a higher chance of getting drop out, loss to follow up or some amount by means of marriage, those who are potential, in this particular study period the loss to follow up.

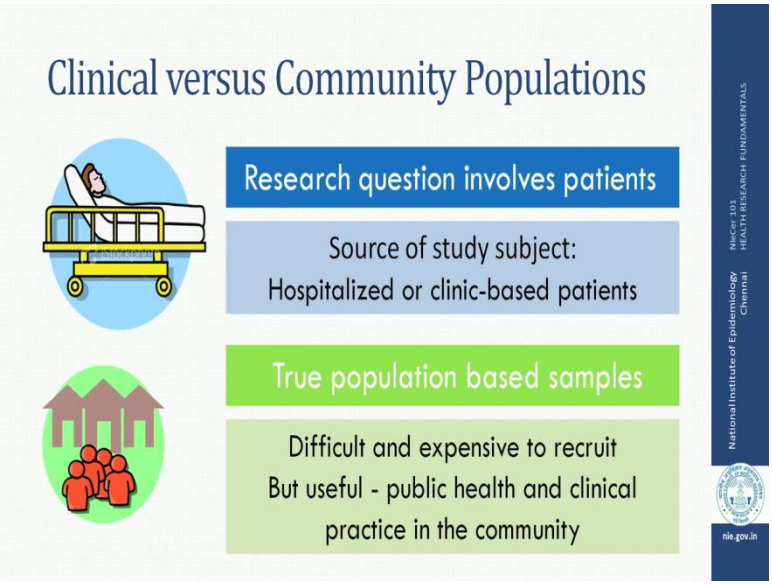
Number two, interfering with the quality of the data, how? For example, here in this example patient who are already in metformin therapy for some other cause may be due to diabetes. So, they may not be included in my study and what is in terms of being on high risk of possible adverse effects with this kind of a design feature, I will exclude those individuals who are hypersensitive to metformin therapy and contraindication for metformin is called renal dysfunction. So, those individuals with the renal dysfunction will also be excluded from this metformin therapy. So, that is how this gives you an idea about, in terms of selection criteria, how to fix our inclusion criteria and how to fix our exclusion criteria in the selection of study subjects as selection criteria.

(Refer Slide Time: 16:32)



With this understanding, what will be your source of the population, Clinical versus Community Population?

(Refer Slide Time: 16:39).



Say like, clinical population where we will include those studies where we have a research question involving the patients, but the previous example those are all the PCOS reproductive age group, females with PCOS with dysmenorrhea. If it is so, then obviously, thus source of study population obviously will be from hospitals or from the clinic based patients. So, it is preferred that best when we need to have a good internal

validity and as well as, so that to be generalize these findings to the population of interest it is preferred with good quality of data the source of clinical population is from the primary health care clinics.

For example, when we are studying a true population; true population based sample which are of a huge geographical area, a community based populations and which will be a very good source for healthy subjects, when we are trying to study over a healthy subjects, over a healthy individuals. For example, Vaccine - efficacy studies obviously, it has to be from the community based studies.

So, in this term, what is a problem here in terms of true population based study is that, it is very difficult. House to house of enumeration have a more difficulty of including them and getting the study subjects within the specific time period is another difficulty and it is very expensive to recruit also. But, in terms of deriving a good public health and clinical decisions among the community then obviously, population based sample has to be done, which I have given the example called vaccine efficacy studies. So, what will be your source of population? It depends purely upon your research question.

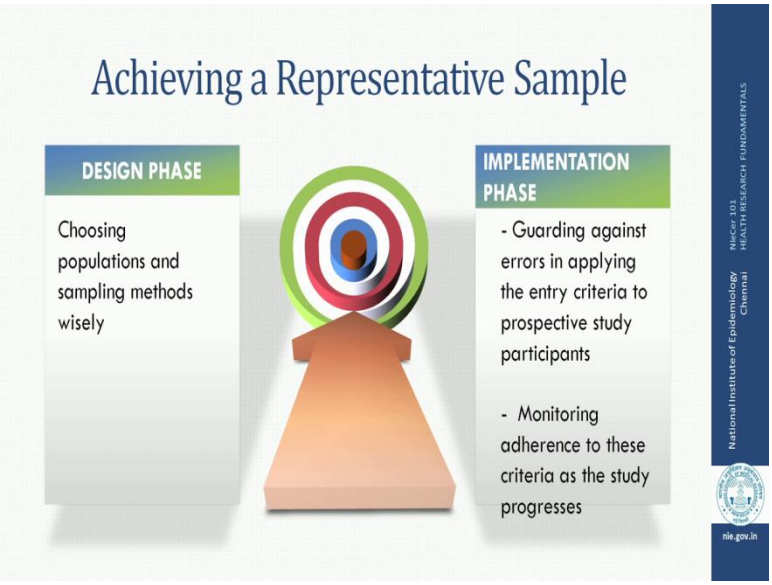
(Refer Slide Time: 18:38)

So, let us see, what are the recruitment strategies? And what has to be considered when we are doing the recruitment of study subjects in your study? So, what is that recruitment basically? Most important factor whenever we are going to recruit the study subject is that feasibility, because feasibility is an important factor when we are to consider to

choose those accessible population and when we are trying to do a sampling procedure. Feasibility decides both, what is our sampling procedure? And how we are going to choose this accessible population from the target population? Or how we are going to choose this subset of samples from this accessible population when the feasibility is an important factor?

And, there are two important goals in term of recruitment, what are the goals of the recruitment? Number one goal is that the subject should be adequately representing the target population, because you are trying to select a subset of a target population by means of geographical and temporal characteristics with predefined clinical and demographic characteristics. So, these subjects should adequately represent it, whatever the kind of findings you are trying to infer and apply is over this target population. And second important goal is that there should be enough subject to meet the sample size requirement because previously we discussed that it should have an adequate size to counter the errors the random error generated in the study so that adequate size should be there, that are the 2 important goals.

(Refer Slide Time: 20:18)



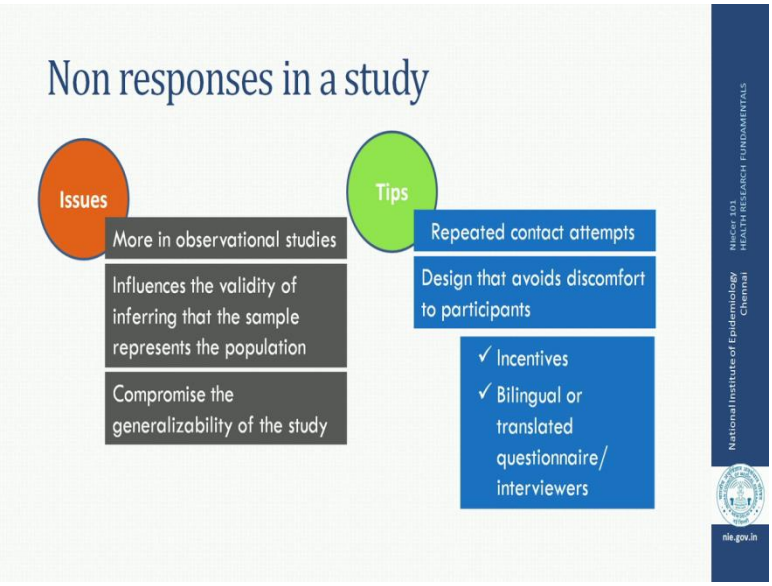
Let us see that one by one, how to achieve a representative sample? Achieving a representative sample, it happens at the beginning of the design phase and ends at the implementation phase. How it happens in the design phase? And the design phase itself, we will decide, what is the choosing the population accessible population from the target

population and choosing those sampling methods wisely from the target population can happen at the designing phase itself, Whereas, in implementation phase how it ends? Number one is that, when you are trying to guard the errors by applying certain criteria over a period of time; that means you have these selection criteria.

By means of selection criteria what you can do in the implementation phase is that, you can achieve this representative sample and you can exclude those individuals, who cannot give a consistent or representative finding to the study. And second is that, unit monitor this entire study; monitor out this study subjects are adhere to this entire period, there should not be any loss to follow up and if there is a loss to follow up how it has to be guarded? And, that is what we are trying to discuss in the next slide.

So, as an overall what you can see is that, when you are achieving a representative sample which is begin at the design phase and it ends in the implementation phase.

(Refer Slide Time: 21:48)



So, as I told that something called Non responses in the study that is called loss to follow up and non responsive rate. This non responses in the study is basically happen more in observational studies and very least in clinical trials and in this, it influences the validity of your findings, which represents the study population and most important thing, third is that it compromise your generalizability. So, two things are affected very important; number one is that, internal validity will be affected when there is a non-responsive rate there is a loss to follow which reduces your sample size requirement and second is that,

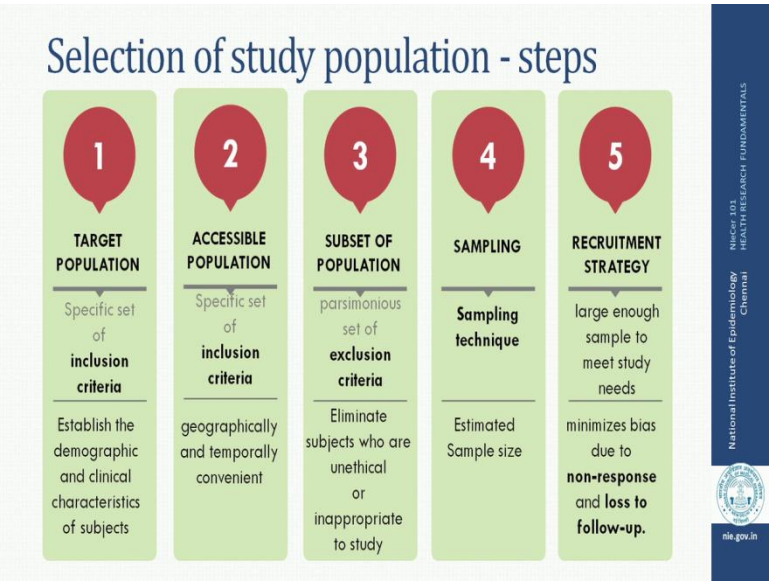
the representativeness and generalizability will also be affected. So, these are the two important things because of this non-response rate.

But, how we can able to tackle these non-responses in the study? Then say, at the initial instance where you could able to get the adequate size and the person is not responding or not available, then repeated contact attempts may give you or may give you an access to the recruit them again into your study. And second is that, design that avoids discomfort to the participants by after they got recruited in between, if they have a loss to follow-up how to tackle this, that.

You can play some kind incentive mechanism, you can take away certain discomforts like, there should not be too sensitive an instrument should be, which should not be invasive. So, you can try to not disturbing in your other technique in the study or methodology in the study, you can still think of designing certain study instruments which has to be in the local language of them, which it has to be understandable and those interviewers should talk in their local language to the participant. So, the bilingual staff with the translated questionnaire may reduce the discomfort of the study subjects.

So, by these mind of a mechanisms, it is where we could able to tackle of recruiting this adequate samples at the beginning itself and as well as where we could able to tackle to stop them loss to follow up while the study is going on. And finally, I am trying to give a summary of how this selection of study population in a step by manner can happen.

(Refer Slide Time: 24:11)



So, in the first step what happens is that, we define a target population by a specific set of inclusion criteria by means of a clinical demography, geographical and temporal characteristics. Second is that, this accessible population is again by specific set of inclusion criteria as said by geographical and temporal characteristics. Third step is that, what we are deriving is a subset of this accessible population, which we are excluding by setting up some kind of exclusion criteria because we are eliminating the subjects which are unethical and also inappropriate to study.

Fourth step is that, we are doing a sampling procedure here, so by defining the sampling technique and by then estimated sample size which is large enough to control the random error will be done in the fourth step. And finally in the fifth step, is a recruitment strategy where we are trying to recruit those adequate subjects and as well as those subjects by those recruitment strategies to reduce the non response rate and as well as last to follow-up. So, that is how this is where the important considerations has to be think of about selecting the study population for your health research.

Thank you.