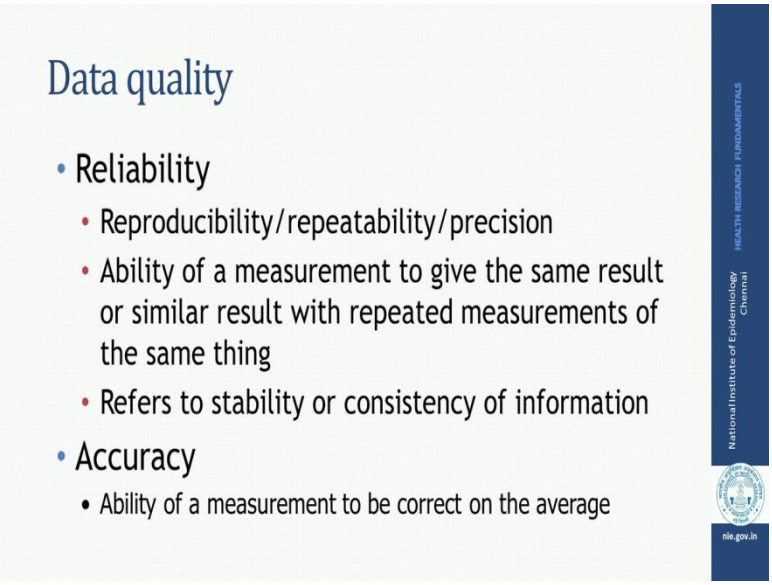
Health Research Fundamentals
Dr. Prabhdeep Kaur

ICMR-National Institute of Epidemiology, Chennai

Lecture - 16
Principles of data collection

Welcome to this session of Health Research Fundamentals. Today, we will be discussing about Principles of data collection. As you all might have gone through various parts of this course, wherein we had been teaching you about, how you develop your research proposal? How you write your protocol? And now you have gone through all that and you are all set to collect your data. Let me tell you that, this is one of the most important components of your research study because this is what is going to determine, what you get out of your data.

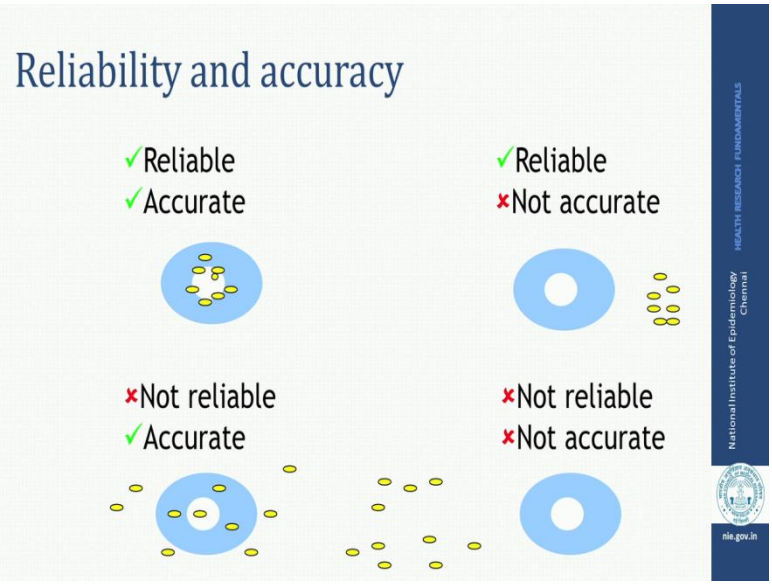
(Refer Slide Time: 00:40)



So, today as we go through this lecture let me discuss about two main principles of data quality, which is reliability and accuracy. So, reliability, what does it really mean? It means that your study should have repeatability and precision. What it means? If this same study is repeated by different investigators or the same measurement is done b at different time points, you should still be able to get more or less similar results. It also refers to stability and consistency of the information. Now, reliability does not ensure

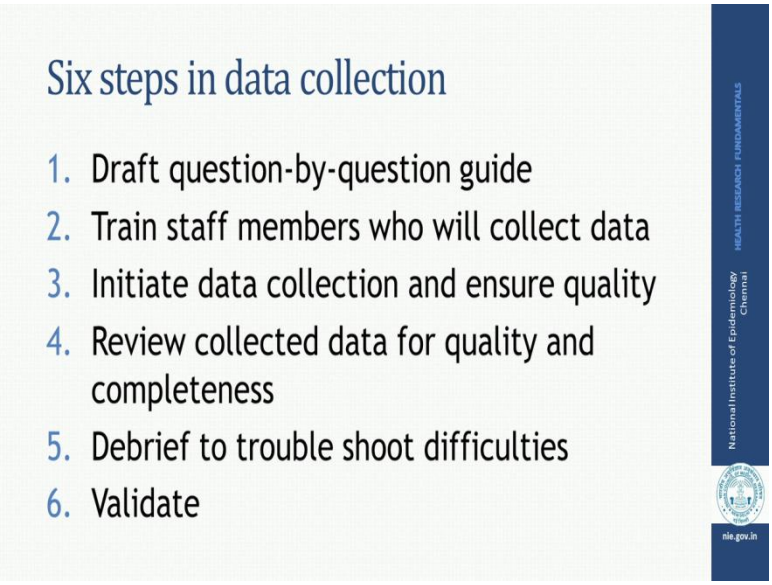
that your data is accurate. Accuracy refers to ability of a measurement to be correct. It could happen that both the attributes may not happen at the same time or they may happen. So, let me give you an example of what would be the various scenarios in terms of data quality.

(Refer Slide Time: 01:42)



As you can see in the slide, your study might give you reliable and accurate result, which means your data is giving the measurement that you really want, the measurement is accurate and repeatable. This is an ideal case scenario. The worst case scenario is that your measurement may not be reliable at all and may not be accurate also. Now, there can be other scenarios, where your measurement is reliable, that is, if you repeat this measurement again and again, you get the same measurement, same kind of result repeatedly. However, this could still not be accurate and there can be another scenario that you may get the accurate result, but when somebody tries to repeat this experiment you may not be able to repeat it at all. So, an ideal study should ensure that whatever outcome you are measuring is repeatable as well as it is accurate, now to ensure that, we can follow certain principles of data collection.

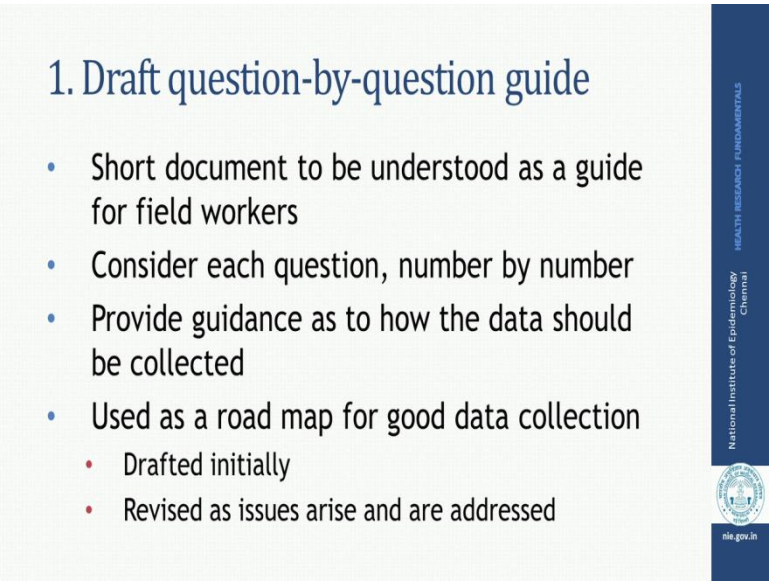
(Refer Slide Time: 02:44)



In today's lecture, we will be going through six broad principles which are listed here. I will be walking you through each of these principles. So, the first is, as you have your questionnaire ready and you are all set to start your data collection. One more step you need to do is, you need to develop a guide, guide wherein for each of the questions you can have a small explanation as to how this data actually need to be collected. The second component that we will talk about is how you could do a good training of your staff members to ensure that the data quality is good. The third component is how do you ensure data quality, when the work has already been initiated. It could be in a clinic setting or it could be in the field and the next component is how you need to do periodic reviews to ensure good quality of your data.

During this process, it may happen that your staff will come across various difficulties in data collection and you need to have periodic debriefing for that and finally and very, very important component is how you validate that your study results are actually correct. So, let us go over each of these components one by one.

(Refer Slide Time: 03:58)

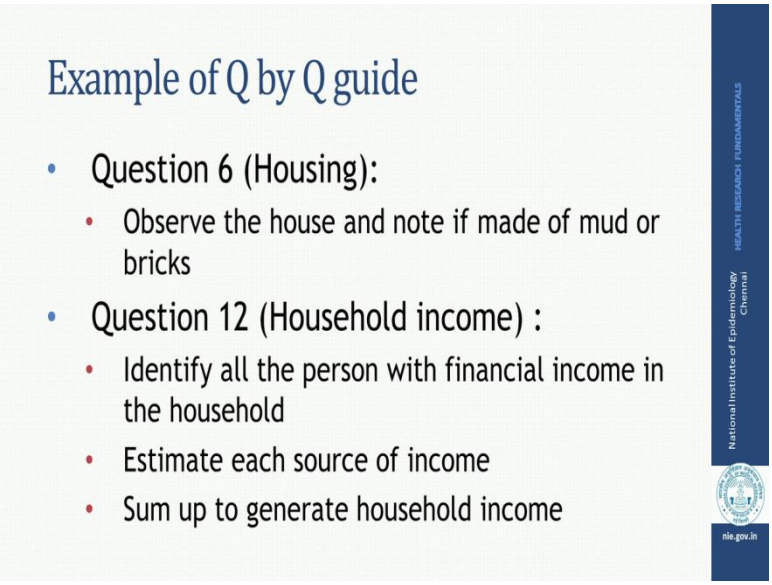


So, what do you really mean by a question by question guide? It is nothing but a short, simple document, which is a guide for your field worker or your clinic research staff or whoever is involved in the data collection. Now, why this is an important component? Everybody may not understand the question in the same way. Different individuals or your staff members may have their own interpretations as to how this questions need to be asked and how it need to be explained to the respondent. The q-by-q guide ensures that all the investigators uniformly ask the question and they all understand the question in the same way. What you do here? You just take your questionnaire and write a short explanation under each of your question, as to how this question should be asked and what could be the possible responses and in certain situations you may have some probes. You would like the investigator to probe little more on a particular question. If you want them to do so, you would like to make a mention of that.

Similarly, it allows you to explain, where they should skip? Where they should give more emphasis? All this guidance can be given through this small document. This not only provides a guide as to how data should be collected, but it is your road map and at any point of time, whenever there is a doubt, whenever there is a lack of consistency with in investigator you can always ask them to go back to this document to ensure that the data collection is being done in a uniform way. Now, it could happen that as you start collecting data, you may come across certain difficulties in the field and you may have to

revise this guide, which is all right and what is important is whatever changes are done, they should be documented in this q-by-q guide at that point of time.

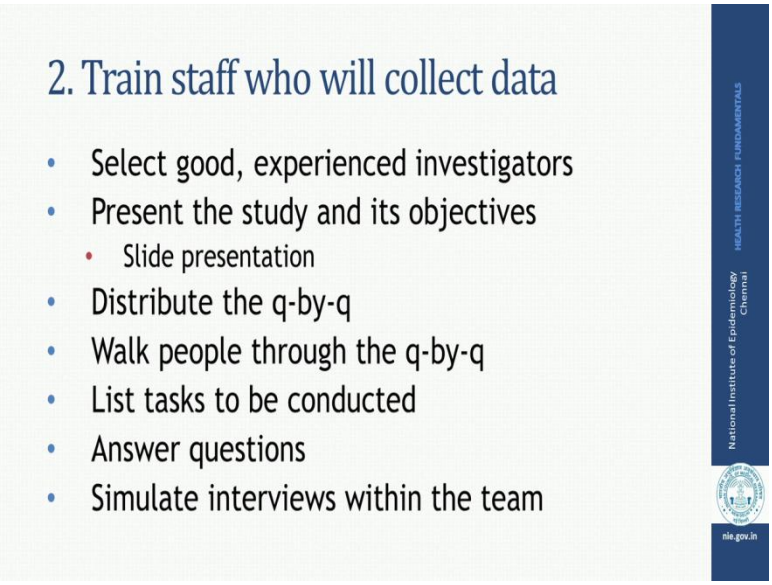
(Refer Slide Time: 05:52)



This is just to show you an example. For example, you have a question on, what is a type of house like we could give several options, this is a kutcha house, it is semi pakka house or a pakka house. Everybody may not understand, what is a kutcha house? So, you may want to give an explanation. Observe the house and note, if it is made of mud or bricks and then you could write an explanation. If you find that the house is made of mud then mark it as a kutcha house.

Similarly, another example when you are asking a question regarding household income? Now, it could happen that in a particular house hold there is more than one member, who are earning members of the family. And in this situation, you would like your investigators to first of all inquire, how many earning members are there in the family? Then they need to find out each one of them, how much they earn? And what is their monthly income? They need to add up all, that amount and then write the answer as what is a total household income. So, similarly for each question you can have the detailed explanation.

(Refer Slide Time: 06:58)



Now, once your q-by-q guide is ready, you are all set to start your data collection. Before you get on the ground, whether it is a clinic, whether it is a field base study, it is a household study, whatever will be the type of study, training, good training is essential to ensure that the quality of data is good and in that, the first and foremost step is choosing right kind of people. You need to understand, what your study requirements are, and depending on the study requirements, you need to recruit or select right kind of investigators. Now, this could be depending on, suppose you are doing a study, which is on clinical research. Then you need to make sure that your investigators know how to ask the clinical questions. Are they knowing, how to use the clinical terms, when they are interacting with the patient? If it is a field based study, you need to make sure that your investigators are familiar with the local language. They know how to culturally interact with the people at the community level. So, selecting the right kind of investigator, who will be suitable for your data collection, is the most important step.

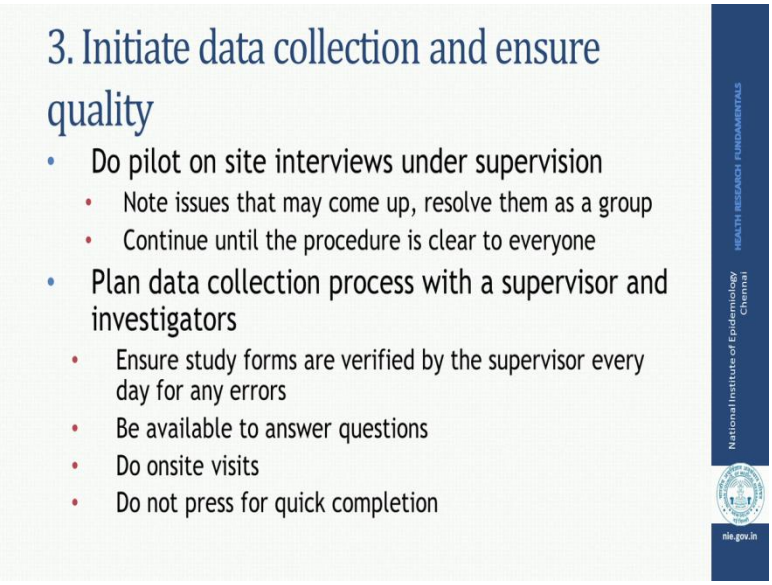
Once you have done that, you need to have a classroom session with them. This is nothing, but walking them through your, first of all you introduce them to what your study is all about, why are you doing this study? What are your objectives? And what is that you would like them to know the basics of, what are the various definition you are using in the study? What are the various component of the study? What is the kind of data that needs to be collected? For example, your study may have different component, there could be a questionnaire, there could be different measurements and maybe you

want them to measure something like weight, height. So, you need to familiarize them, what are the various kinds of a data collection that they need to do?

Once you have done that and given them the overview then you need to share this q-by-q guide, question by question guide. Walk them through each question, allow them to ask questions and let them go through it and see whether they are able to interpret those questions, whether they are able to understand those questions. Once, they have done that, you need to tell them what exactly they need to do in the field? How they need to ask those questions? And what should be the explanation that they should understand before they ask the questions?

Now, as a first step having gone through this exercise, you would like them to probably do this interviews, first of all in the classroom, wherein one of the investigator could act as the interviewer and the another investigator could be the respondent and they should simulate this in front of you, which allows you to guide them, to tell them, how the question need to be asked. Whether the question is being understood by the respondent, whether question is being asked in the appropriate language that is given in the questionnaire. Having done that you are all set to start your data collection, but there is one more step.

(Refer Slide Time: 09:50)



Now, having done something in the class room does not mean you will be able to do it on the ground because the real situation may be very different. Some of your investigators

may be working in that kind of a situation for the first time. For example, you may hire a person who is trained in clinical research, but they have never been in a clinical setting. Similarly, you may hire a field investigator, who is a graduate in social work or any other background. However, they may never have been in the community setting. So, once they get on to the ground, they may not feel the same level of comfort. So, before you get on to the main data collection, you need to have pilot. A pilot, where they do the interviews exactly in a similar setting, what your study setting is, but it may not be the same place, it may not be the same clinic or it may not be the same village.

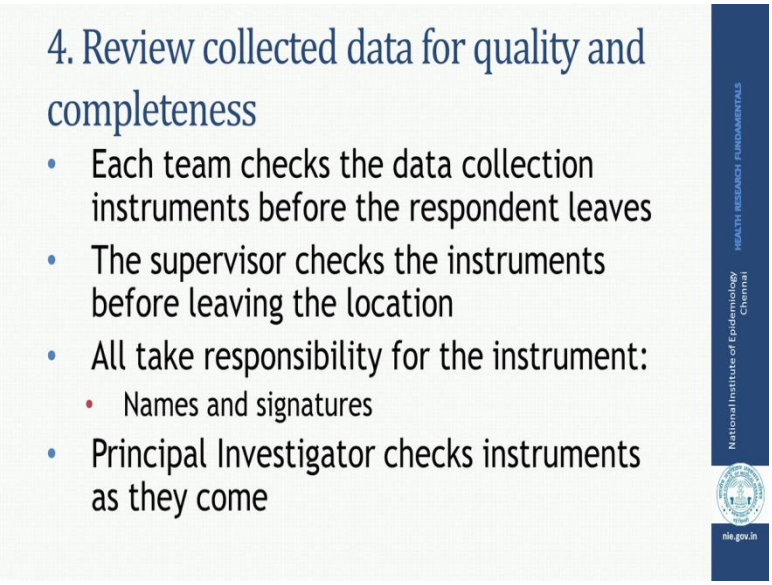
In this situation you will see that, they may come up with different kind of questions. They may not have anticipated, how the respondent will respond. They may not have anticipated what difficulties might be there in making the respondent understand the questions. So, as you through this pilot, you explain them the process, clarify their doubts and this procedure you could continue unless everybody is very, very clear about what the questions are and how the questions need to be asked.

Now, having gone through this, you are all set now to do your final data collection. Here, what is very important is that, you need to plan your team structure. So, there has to be a supervisor, who could be present all the time or who could be present periodically along with the investigators. Ensure that the study forms are verified by supervisor every day. Now, why is this important because once you have left the clinic or once you have left the village or the place where you are doing data collection, some of the errors cannot be corrected, during, as the data collection starts they may come up with different queries. So, you need to be available to answer those questions. This could be over the phone; this could be over messages or whatever way you can use.

Now, once the data collection is ongoing, the next step that as a principle investigator you need to do is few on site visits, to ensure that the data collection is done as per the protocol and the questionnaire is being actually used in the field the way you had perceived it. I think one of the important thing you need to keep in mind is very often what happens is there are time pressures. You want to finish your study quickly, you want to be done with it, you may pressurize them, you may say, I would like you to do ten questionnaire every day. However, you need to make sure that the quality is not diluted because you are pressing them to work under very tight time lines. Having done

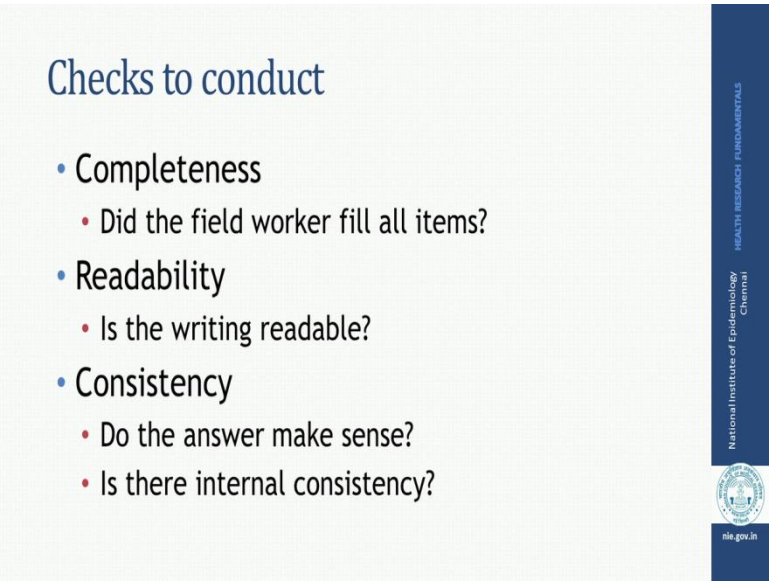
this, having your data collection, as data collection goes on at various time points you may want to review.

(Refer Slide Time: 12:52)



Suppose, let us say you have already collected now, 50 forms. At this stage, you may want to make sure that the data collected is of good quality and the data is complete. How can you do that? The first step can be done in the field itself. As your supervisor finishes the day, he can collect all the forms from various team members and check them on that day itself. The next step is the forms will reach you as a principle investigator and you need to go through them to make sure that you are satisfied with the quality of data.

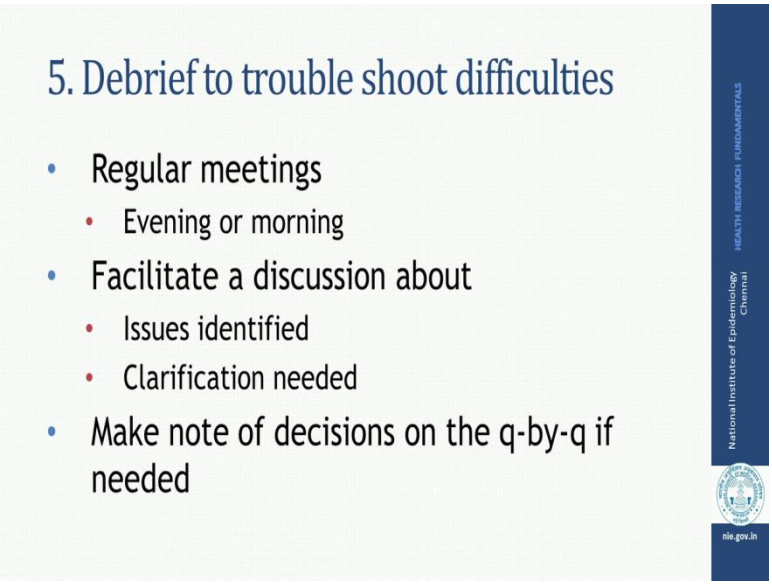
(Refer Slide Time: 13:35)



Now, what kind of checks you can do, when you have got the forms in your hand? I think the first and foremost and the easiest one, but the most important one is ensure there are no blanks. It could happen that when they are asking multiple question, since your questionnaire may have about 20 questions or 25 questions, they may just forgot to ask a few questions or they may have asked the question, but they did not mark it in the questionnaire, but in either case you will see them as a blank. So, you need to ensure completeness of the data.

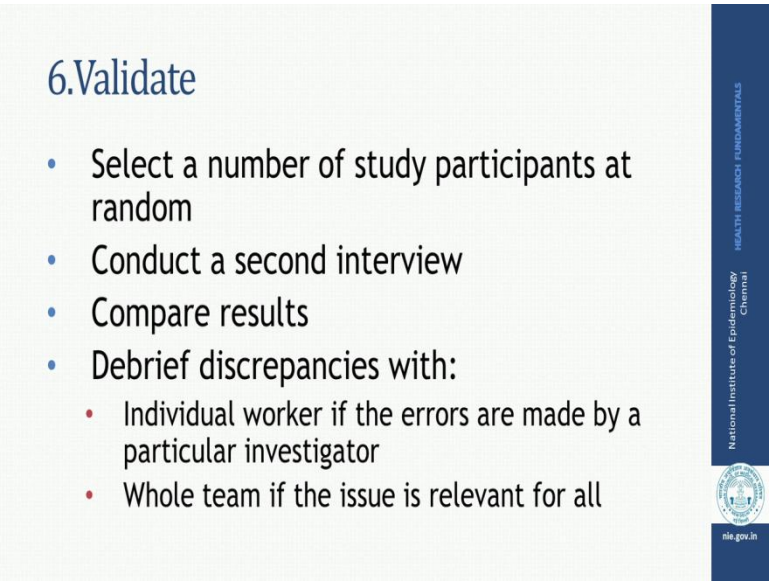
The second is readability, sometimes the way it is marked it may be hard for you to even understand and some of the answers could have some actual explanation to be written. Some kind of clinical symptoms or some kind of narratives have to be written, ensure that those things are readable. The next is the consistency, does the answers make sense? Do you feel that this could be the way the people would have answered or you got this answer because may be your respondents were not understanding the questions.

(Refer Slide Time: 14:41)



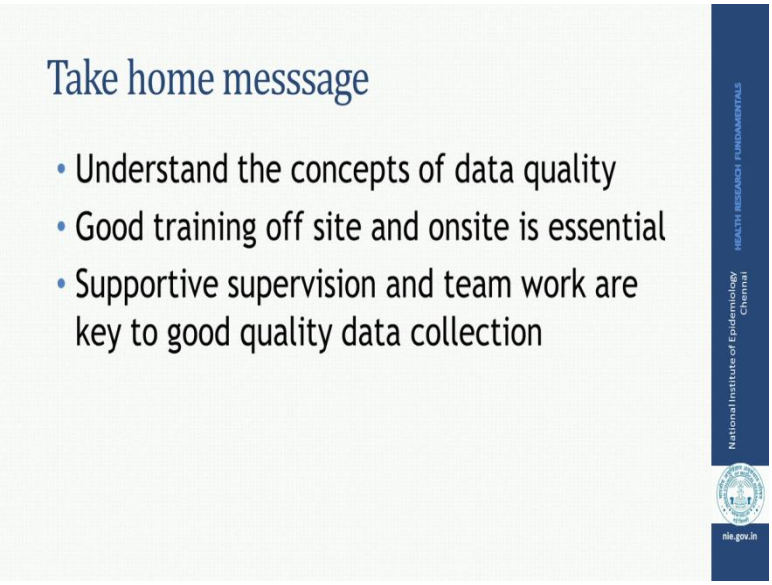
Periodic reviews are essential, in case of any study, whether it is a clinical study; it is a field study or any type of research study. Now, these regular meetings, one it could be done by the supervisor, which could be in the field itself or after coming back or you, can have a periodic review once in a week, once in a month, depending on what is duration of your study. What is that these review meetings will be useful for? First of all to clarify, if there are any queries they have about the questionnaires, if they understand the questions well, if they came across something which they had not anticipated. Now, it may happen that during these meetings you may end up doing some changes in the way your answers are drafted or in the way questions are asked. If you are doing that ensure, that these are well documented and they are added it in your q-by-q guide.

(Refer Slide Time: 15:36)



Now, coming to the last, but not the least very, very important component of data collection, which is validation of your data. How can this be done? So, for any study a small sub sample, it could be as small as even 5 percent should be selected and an independent second interview should be done. This is to ensure that the data collected by your investigator is actually valid. By comparing the results, you will be able to find out, if there are any discrepancies, if somebody has made any major errors or if it could be that a particular investigator might have been repeating certain errors or it may happen certain errors are repetitive across the team. So, depending on whatever the problem is you may want to discuss this with the individual team member or with all the study team.

(Refer Slide Time: 16:27)



So, just to sum up, what we have gone through today. Before you get on to your data collection, understand the concepts of data quality, good training both in the classroom as well as on site, in a similar setting where you are going to conduct the study is essential for a good research study and supportive supervision, team work are key to good quality data collection.

Thank you very much.