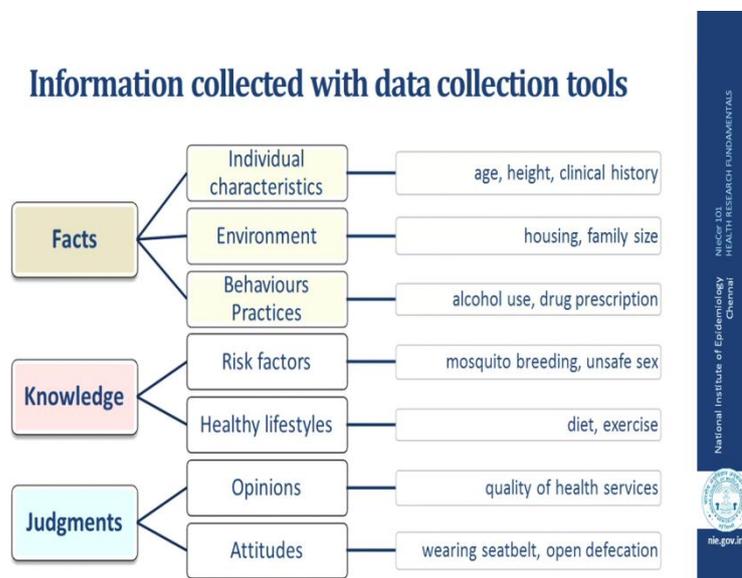


**Health Research Fundamentals**  
**Dr. Tarun Bhatnagar**  
**ICMR – National Institute of Epidemiology, Chennai**

**Lecture - 15**  
**Designing data collection tools**

Hello and welcome to this session of Health Research Fundamentals. Today, we are going to talk about designing data collection tools or the instruments that we use to collect data in health research.

(Refer Slide Time: 00:22)



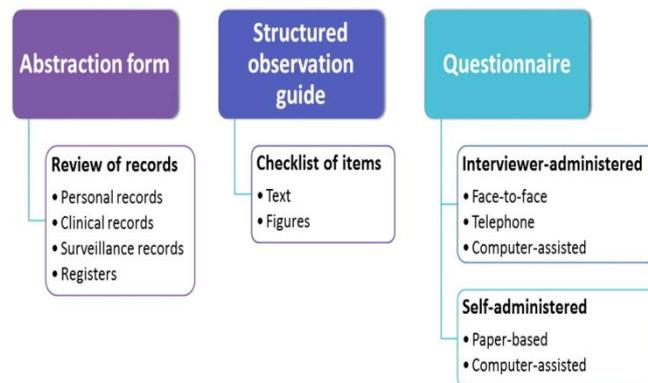
Generally, in health research, the kind of information that we would like to collect can be basically divided into three areas. We want to know facts such as, the characteristics of our study participants, the environment that they live in and their behaviors or practices. Secondly, we might want to know their level of knowledge for things such as, risk factors for getting disease or knowledge about healthy lifestyle, so as to prevent diseases and thirdly and very importantly we might want to collect information on what we call the domain of judgments, basically what are the opinions of the research participants on a certain issue, may be such as quality of health services.

We would also may, we may want to know about the respondent's attitudes towards certain things for example, could be something as wearing seat belts, use of open

defecation and so forth. Now, how do we collect data on all these different kinds of information that we want to collect.

(Refer Slide Time: 01:31)

## Different tools to collect data



For this purpose, we can make use of a variety of tools depending on what kind of data we would like to collect. One, we have what are called abstraction forms, which is basically doing a review of records of the study participants, which could be their personal records, the forms. If you want to get information on their disease conditions, signs and symptoms, a treatment given, then we could look at the clinical records. We could look at data in general, the data that is collected through disease surveillance and we could also look at registers, wherein some information, which may be there and all of this information can be culled out into what we want, in the form of a data abstraction form.

Secondly, another tool that we have for collecting data is a structured observation guide. This is very useful when we would like to document certain processes, whether they are happening, the way they are happening, that they are happening in time or not, are the objectives met and for this purpose we may use a check list of items that we would like to collect data on which could be the textual or figurative and third and most importantly, the tool that is used most commonly is a questionnaire, wherein we would like, we talk to the person and get information.

Now, again questionnaires can be divided into two kinds - it could be interviewer administered, where the data collector actually administers the questions to the respondents, which could be done either face to face, it could be done on a telephone and now even have computer-assisted technologies to do face to face interviews. The questionnaires could also be self administered, if the study participants can read and write and they are knowledgeable enough to understand what the investigators want. These could again be either paper based or now we also have computer assisted self administered questionnaire which can help the respondents to directly put their information into a computer database.

(Refer Slide Time: 03:55)

### Key elements of data collection tools



Aday LA, Corneliu LJ. Designing and conducting health surveys : a comprehensive guide. 3rd ed. CA: John Wiley & Sons, Inc. 2006

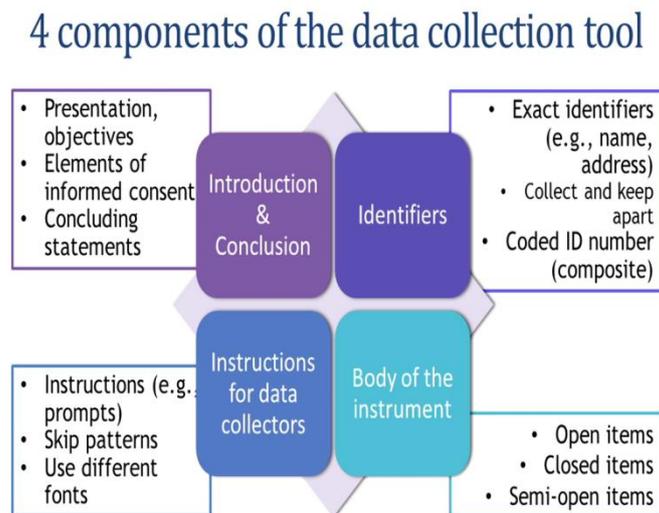


Now, whatever the tool of data collection we use, what we want is that we need a valid response from the study participant. The response should make sense and we should be able to use that information effectively for our health research needs. In order to do that, every data collection tool needs to make sure that there are some key elements that make up this data collection tool. Some of these elements are the clarity of words that you use in a data collection tool, the balance of phrases and sentences, the length of sentences; how long are these questions, the comprehensiveness of the responses in terms, if you are giving categories of responses that you expect from the study participants.

These categories can actually even post constraints in terms of what information could or could not be collected through a data collection tool. Of course the utility of the data

collection tool and specially the utility of the instructions; which I will elaborate in the next slides. The order in which the questions are asked is a very important element which can decide how your respondents answer your questions and of course, the context in which you frame these questions and the tools are used.

(Refer Slide Time: 05:18)



So, if we look at any data collection tool generically it has 4 components, there is the introduction part at the beginning and a conclusion at the end. There are what are called identifiers, then the each question may have linked instructions for the data collectors and of course, the whole body of the instruments which is basically the question items. In terms of the introduction, the introduction is used to present the study to the study participants, state out the objectives and probably get informed consent and so forth. It is also good to always have a concluding statement at the end of your questionnaire and thank the study participant for their time and effort that they have put in to answer your questions.

Now, every data collection tool will also have what we called identifiers, which could be either the actual identifiers, the information such as the name and address of the study participant, which can identify who the person is, it is always a good idea from an ethical and a human subject protection perspective, if you are collecting this data, to collect this in a separate sheet of paper and which can be referred to later on, if need be.

On the other hand, in order to maintain confidentiality investigators also use coded id numbers to give identifiers to the study participants and these id numbers could be composite in such that they could have numbers which denote say, the state which the person belongs to the district the village and then the household and then the individual id. So, it could be a mix of all these numbers, all these codes and then you get a composite code looking at which you can actually identify who the respondent is, but would not be able to actually get an exact identification and which is good from the confidentiality perspective.

In terms of the instructions, there could be general instructions for the data collectors such as prompts in terms of, for example, do not read out all the responses, tick only the one that the study participant mentions. There could be instructions for skip patterns, now the questionnaires may have skip patterns in the sense that there may be some questions which, based on the response to those, the subsequent question may not be relevant and then there is an instruction, which says that you skip this from question 2, you may go to say question 19 and skip rest of the questions because they are not relevant to this study participant.

It is also always a good idea to may be used different fonts, so as to make it clear that what is the actual question and which part of the item is actually an instructions for the data collector and then of course, we have the whole body of the instrument which is basically the question items. Now, these question items could be of various types, we could have what are called open questions, we could have closed ended questions and we could have somewhere, something in the middle.

(Refer Slide Time: 08:39)

## Open questions

- Answers are not suggested
- Respondents must generate an answer
- Advantages
  - Give freedom of response
  - Stimulate memory
  - Can be useful to generate closed responses later
  - Useful at a hypothesis raising stage
- Inconvenient
  - Difficult to code and analyze
  - May be incomplete and / or unfocused

### Examples:

- What disease can you acquire from tobacco?
- What places did you eat at in the week preceding the disease?



Let us see what are these different types of questions. As the name suggests open questions are the ones where the answers are not suggested to the study participants and the respondent has to generate an answer. The good thing about these questions is, it gives total freedom to the respondent to give the answer of what they want. They are not constrained by the categories of answers that already exist. It helps to stimulate the memory of the research participant and gives you a more better answers who as to speak and it is also useful at hypothesis raising stage, wherein we are really not sure what the appropriate answer is and you can generate a lot of responses from the study participants.

Of course, when you generate a lot of responses, open questions, the inconvenience is that it may be difficult to code and analyse. You may have a long list of responses and then to categorize them later may be an issue and sometimes, if it is open, the responses may be unfocused or incomplete and that can pose a challenge in terms of the analysis. Now, to overcome this problem, we can have open questions, but then we can have them with closed ended answers. Although, there is a category of answers given for those that question, but the data collector does not suggest an answer from these categories to the study participant. So, when the answers are freely mentioned by the respondent the interviewer will spontaneously tick those that are specified from the list of categories of responses given in the question here. So, it is expressed as an open question, but you finally analyse this as a closed ended question.

(Refer Slide Time: 10:29)

## Closed questions:

### 1. Dichotomous options

- Suggested answers include 2 options only
  - Yes and No
  - Male and female
- Advantages
  - Forces a clear position
  - May be useful for key, important, well framed issues
- Inconvenient
  - May oversimplify issues

#### Examples:

- Did you eat at restaurant X between 1 and 28 February?
- Have you ever consumed tobacco products?
  - \*A dichotomous question here is likely to over-simplify, unless it is used as an introduction

What are closed questions? Closed questions are the once, where you have a question and you have a set category of answers only which are the once that are acceptable to this investigator. These could be two types, we could have Closed questions with only two options, dichotomous options such as yes-no, male-female and so forth. So, these kind of questions forces a clear position for the respondent to take and it is very useful to get key information specially for important issues and which is very focused. Although sometimes depending on the question it may actually over simplify some of the issues, where a yes-no answer is not something that is going to give you a very good information.

(Refer Slide Time: 11:19)

## Closed questions: 2. Multiple options

- Multiple options of answers are suggested
- One or multiple answers acceptable
- Advantage
  - Larger choice of answer options
- Inconvenient
  - May be difficult to choose only one option

### Example:

- Where do you go to seek treatment for fever?
  - Government Hospital
  - Private clinic
  - Pharmacist
  - Traditional healer
- Do you wear a helmet when riding a bike?
  - Always
  - Sometimes
  - Never



Then we have **Closed** questions with multiple options, so more than two options. Now here again, they can be two kinds of **Closed** questions with multiple options. We could have questions where although there are multiple options, only one of the option is acceptable. So, depending on what the respondent says, one of this option is ticked by the data collector. On the other hand, we could have closed ended questions with multiple options, wherein even multiple responses by the study participants may be acceptable.

The important thing to note is that, while you are designing the questionnaire you need to put a clear instruction for the data collector, whether only one answer is acceptable for this question or more than one answer may be acceptable. So, we have a large choice of answer options; again depending on how the question is framed and what the actual question is, sometimes it may become inconvenient and difficult to choose only one option, if there is a possibility of more than one option, but that possibility is not provided in the questionnaire.

(Refer Slide Time: 12:39)

## Closed questions:

### 3. Quantitative answers

- The respondent must provide a quantified answer
- Advantage
  - Allows creation of continuous variables
- Inconvenient
  - May requires validation:
    - Some “quantified” answers might be limited in the way they can be handled as continuous variables

#### Examples:

- How many time did you visit the clinic in the last 12 months?
  - True continuous variable
  - Four visits is the double of two visits
- How would you describe your pain on a 1-10 scale where 1 would be the minimum and 10 would be the maximum?
  - In fact a qualitative variable with 10 options
  - Requires validation
    - Six may not be the double of three on the scale



So, we have to be mindful of this when you are designing your questionnaires. And thirdly, we could have **Closed** questions which have quantitative answers, where the respondent has to provide a number, such as age, such as may be one example here, if you see; how many times did you visit the clinic in the last 12 months? These kind of questions allows the creation of continuous variables and then measuring and doing the analysis for continuous variables. If you need we can always categorize these variables later on in the time of analysis if needed.

However, sometimes it can become inconvenient to give a quantitative answer because some quantified answers may be limited in the way they can be handled as continuous variables and where the number itself is difficult to **interpret**. So, we need to be careful in how we are framing these kinds of questions.

(Refer Slide Time: 13:33)

## Semi-open questions

- Suggested answers
- Possibility to create another answer
  - Other, specify: \_\_\_\_\_
- Advantage
  - Leaves the door open to unplanned answers
- Inconvenient
  - Difficult to analyze

### Example:

- Did your child have complication following measles?
  - None
  - Pneumonia
  - Diarrhoea
  - Eye problems
  - Other, specify: \_\_\_\_\_

Then there we could have something called, which are called semi-open questions, where you basically have a question with several responses, the answers here are suggested, but there may be one option where, which is kept open and the most common one that we see in data collection tools and questionnaires is others. For example, did your child have complication following measles? There could not have been any complication, that could be pneumonia, diarrhea, eye problems or there could be some other complications, which may not be so common so as to be put in a category, but then you give an option to the respondent to even say things other than what is in the list of categories of responses. So, it leaves the door open for unplanned answers; however, if there are too many of these responses, it may sometimes become difficult to analyze.

(Refer Slide Time: 14:34)

## Formulating questions

- Write short and precise questions
  - Full and complete phrases
  - Avoid ambiguities
- Use simple words of every day language
- Avoid negatives and double negatives
  - ✗ Do you sometimes care for patients without washing hands?
  - ✓ Do you systematically wash hands before caring for each patient?



Now, let us look at, how some of the principles and do's and don'ts of formulating the questions. It is always the good idea to actually **have** short and precise questions, say for example, if you want to know the age of your study participant, just writing age is not a good idea. You should always use full and complete phrases, so, what is your age? So avoiding ambiguities. It is a good idea to use simple words and not use very complicated academic language, use everyday language in terms of questions because again remember that your respondents are lay people.

When you are formulating questions, again it is good idea to avoid negatives; especially double negatives. So, one example that we have here is, do you sometimes care for patients without washing hands. If you see carefully, there is a negative connotation here, and there are sort of two parts of the questions, one is caring for patients and the other one is washing hands or without washing hands. So, a better way to phrase this question could be to ask it directly and more positively, so, do you systematically wash hands before caring for each patient?, which makes it clear and unambiguous.

(Refer Slide Time: 15:57)

## Formulating questions

- Ask only one question at the time
  - ✗ Did you refuse treatment because you feared side effects?
  - ✓ Did you refuse treatment?
  - ✓ If yes, was this because you feared side effects?
- Be specific
  - ✗ Are you aware of the modes of transmission of HIV?
  - ✓ Among these practices, can you tell me those that could expose you to HIV?
- Use neutral tone to avoid influence
  - ✗ Have you been promiscuous in the last six months?
  - ✓ How many partners have you had in the last six months?



Again when you ask a question, it should be only one question at a time; say one example here that we have here, so, did you refuse treatment because you feared side effects?. Now, here actually we have two questions; one is asking: did you refuse treatment? And the other is trying to find out the reason of why, if the person refuse treatment, why did they do so. It may be that the respondent may not have refused treatment then how does that respondent answer this question. So, it is a good idea to actually split such questions into two questions, where in one could be first; the first part say could be, did you refuse treatment? Depending on yes or no, if the answer is yes, the following question could be; was this because you feared side effects? It makes things very clear.

Again, the questions need to be specific and not vague, so an example that we have here is say, you want to know from the people about how HIV is transmitted and the question is: Are you aware of the modes of transmission of HIV? which is sort of an open ended, it leaves space for people to answer whatever they may want to answer. But if you really want to know whether HIV is transmitted through sexual route, through heterosexual route, homosexual route, blood transfusion, drug use, etcetera, then it is better to actually put these as categories of responses and then phrase the question as, among these practices can you tell me those that could expose you to HIV.

So, you know that you are going to get the proper answer to the question in which you wanted to have. It is also a good idea to always use a neutral tone and avoid judgmental tones, which can influence the response of the study participants. Remember that you are there as a data collector to just collect data and not be judgmental of what the respondent is telling you. One example that we have here, so if you want to know about say the sexual practices of people, instead of asking them have you been promiscuous in the last 6 months? Again, the words promiscuous has a negative connotation , so instead of that it could be more neutral and more academic kind of a question, wherein you just be direct and specific and ask about: How many partners have you had in the last 6 months? without being too judgmental.

(Refer Slide Time: 18:36)

## Sorting the order of questions

---

From simple to complicated

---

From general to specific

---

From casual to intimate

---

Group together questions related to the same topic

---

Identification questions at the beginning or at the end

---

In chronological order, if questions related to sequence of events

---

Introduce simple questions as a break if the questionnaire is complex

---

Triangulate through multiple questions on the same topic if the subject is important



Now, the next thing that you would want to do, when you are designing your questionnaire is, to actually sort out what would be the order of the questions that you have. Remember that, the way in which you ask questions should be such as if your having a conversation with a study participant and it should be a smooth flow of questions one after the other linked to one another.

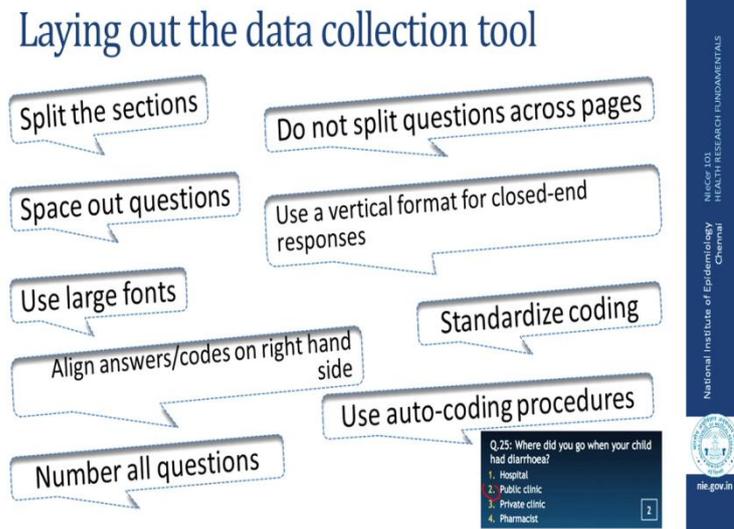
Some general principles to keep in mind is that always ask simple questions first and keep the complicated questions for the later part. You can ask more general questions, socio demographic characteristics, like that, and then go on to the specific questions of what your study is all about. Ask more casual questions in the beginning, which the

respondent will be easy to answer; more of factual questions and then more intimate questions, sensitive issues, questions about attitudes and opinions could be something that could come later on.

It is always again a good idea, remember, to actually group together all the questions which are related to the same topic of inquiry instead of having them spread across different parts of the questionnaire, which can actually confuse the respondent. In terms of asking the identification questions either they can all be asked at the end in terms of the name, age, gender, say the address and so forth or they could even be asked at the end so as to get to the specifics right at the beginning.

If you are trying to collect information about a sequence of events, then your questions should be in that proper chronological order of how things would have happened in real time, which will actually help the respondent to recall the responses in a better way and also in a more logical way. Again, if your questionnaire is complex, there are lots of questions, it is always good to give a break in the middle and may be have some simple questions and then come back to your complex questions. Many a times, we may ask the same kind of question in different ways in the same questionnaire and then that is used, if that is the subject matter that is really important for the study and you really want to know what the respondent is telling you make sense is valid is true and then. So, the multiple questions on the same topic could be asked at different places in the questionnaire and then when you are trying to analyze it, you can triangulate these responses to get to what information you would like to extract from this questions.

(Refer Slide Time: 21:27)



Once you have sorted out the order, now what is needed is to actually lay out all these questions in the questionnaire. Again remember, laying out the format, the structure of the questionnaire is again critical because the way the respondents look at the questionnaire, the design and look of the questionnaire influences or can influence the response of the participants. So, if you have different sections in the questionnaire, it is a good idea to split the sections, have one section, may be have line and then have the next section.

Do not try to cramp questions all together, have spaces between the questions, so that it is readable clearly. Try to use large fonts, not too small fonts so may be font size of 11 or 12 would be ideal. Again, do not split questions across pages. If you have a question and then if half of it goes to the next page, it becomes difficult for the data collector to actually read the question, you will have to turn the page and so forth. So, if that is happening, make sure that you sort of bring the whole question on one page. In terms of formatting and aligning, alignments are again, gives you a nice, neat look for the questionnaire. So, it is a good idea to actually align your questions on the left hand side and your answers and codes on the right hand side, which gives a neat two column kind of a look to your questionnaire, which makes it more appealing.

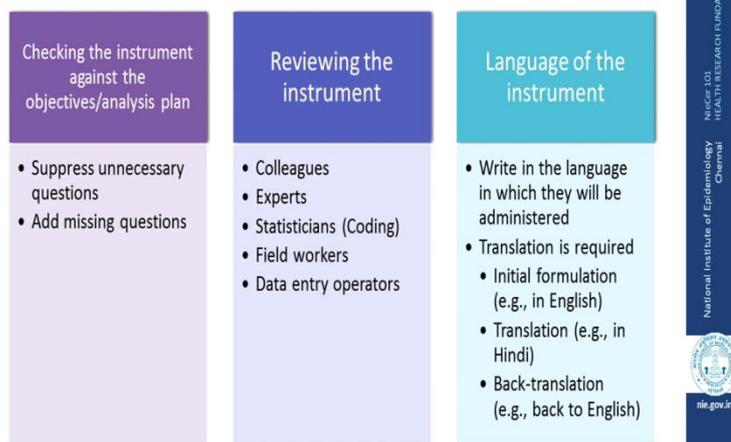
Do not forget to number your questions starting from 1, 2 and whatever that you have. Again, one very important thing to keep in mind is coding. Remember that ultimately

what you are going to do is use this data and to enter it into software, give it codes to analyze it. So, it is always a good idea, to actually have a coding system inbuilt in the questionnaire itself. So, you need to standardize your coding. So wherever say, the simplest example I could give you is, a yes-no. So, you could have a code of 1 for yes and 0 for no, for example. Make sure that every question, every response where you have yes, no; you have coded it as 1, 0. Say you have male, female. So, you always code the male may be as one and the female as two or something like that.

Another way to actually simplify this coding is to use what we call as auto-coding. So, the numbers that you give to the categories, say you have four response categories. So, you would number them 1, 2, 3, 4 and if say the response is number 2, then you use the same number 2, as a code for this question item. So, these are some of the ways in which you make sure that the layout of your questionnaire is neat and it is presentable and it something that helps both the data collector as well as the data entry operator and the person who is going to enter and analyze the data at the later stage.

(Refer Slide Time: 24:44)

## Finalizing the data collection tool



When you are finalizing your data collection tool, make sure that the questions that you have are something that is relevant to the study that you are doing. So, as may have been mentioned in earlier sessions, the investigator needs to be a slave of the study objectives and the analysis that is already pre-planned for the study. So, make sure that the questions that you have are relevant to answer those study objectives. Do not put

unnecessary questions just because you are going out in the field and doing a study does not mean that you can ask anything and everything and if you feel that there are certain missing questions make sure to add them.

Once you have done all this, it is also a good practice to actually review your instrument before you take it to the field. The reviews could be done by your colleagues and experts in the field, you could also give it to the statisticians, to actually review look at the coding; whether that is going to be something that is going to be useful for them and then you could even the field workers or the data collectors, who are going to actually collect the data, they can be your key informants to actually go through the questionnaire and tell you whether the flow is appropriate, whether the questions make sense, is there any ambiguity or is there something that is not understood and so forth.

Keep in mind that the language of the instrument; the questionnaire or the data abstraction form or so forth has to be in the language in which you are going to interview the study participants. So, if your study participants speak Tamil, then the questionnaire should be in Tamil, if it is Hindi, it should be in Hindi. Generally, as investigators, English is the common language, so may be your initial formulation of the questionnaire would be in English. Then what you need is a translation, you need to translate it into the local language and then very importantly have somebody else do a back translation into English, so as to make sure that the translated version in the local language makes the same sense as you wanted it to be and when you frame those questions in English.

(Refer Slide Time: 27:05)

## Pilot testing the data collection tool

- Check that the instrument is:
  - Clear
  - Understandable
  - Acceptable
- Check flow and skip pattern
- Check pertinence of coding
- Estimate the time needed to ask all the questions
- Pilot test with a few volunteers
  - Persons similar to the study population
  - Persons who are not to be included in the study

Before going to the field it is always important to pilot test your tools. You need to make sure that your study instrument is clear, the questions that you have asked are understandable to the people and they are acceptable, people are not wary of answering those questions. You need to check the flow and the skip patterns. Make sure that the coding works and it also gives you a sense of how much time it is going to take for you to actually finish the questionnaire. All this can be done by doing a pilot testing by actually administering this questionnaire to a few volunteers who are similar to the study population that you are going to do, but remember that these people on which you pilot test your questionnaire, should not be included later on in the main study.

(Refer Slide Time: 28:57)



So, when we are designing the health research tools, we need to keep certain principles in mind. You need to first make sure, what is it that you want to measure, remember epidemiology is all about measuring. Then you need to relate these concepts to your study designs and the study objectives. You need to match the scales, how you are going to measure these to and then how you are going to do the analysis. Make sure that the scales, the questions, the questionnaires, the data collection tools that you are using are reliable and valid for the population that you are going to apply them to. Taking all these things in mind, choose the most appropriate method of data collection; whether it is a data abstraction form or a structured observation guide or a questionnaire and the type of questions that you are going to put in these data collection instruments.

Keep in mind your study participants, in terms of the language of the questionnaire and also the way in which you are trying to measure the concepts that you are doing and then decide finally, how best you are going to ask the actual question in the study questionnaire. Remember, a study questionnaire can make or break the study; this is something, once you have collected data you may not have the opportunity to go back. So, it is essential that the data that you collect is valid and reliable and in order to do that, it is key that the data collection instruments that you develop are totally valid and appropriate to the study that you are trying to conduct. That is it for today.

Thank you.