

Research Methods in Health Promotion
Dr. Arista Lahiri
Dr. B.C. Roy Multi-Speciality Medical Research Centre,
Indian Institute of Technology Kharagpur
Week 04
Lecture 17: Observational research designs

Welcome back, in this week we were discussing regarding the quantitative methods that we use in health promotion research. In this particular lecture we will be focusing on observational research designs. Now we shall cover the different observational quantitative techniques that we follow in health promotion research and also we will focus a bit more about the survey research design in health promotion. Why? Because survey research design is typically very important in terms of health promotion because often you get a lot of information from different health behavior related surveys. So, what are these different common observational research designs that we use in health promotion research? You know the first design that we call is the cross sectional design. The next one it is called the successive independent samples design longitudinal and cohort designs cohort sequential designs, case control designs and case crossover design.

Now we will discuss all of these all of these designs one by one, but the thing is these designs the longitudinal cohort designs, cohort sequential designs and case crossover designs these are somewhat related to the longitudinal designs and we will discuss them together. And the survey research that we will be discussing it will come under the cross sectional designs. So, what is basically a cross sectional research? For example, cross sectional research means this is your point of observation, this is you, you are observing, you are observing whom? You are observing B and this is your point of observation what is this? This is called time t. So, you are observing B at only a single time point, this is called the cross section of a time.

You may observe different aspects of B we call them say B 1, B 2, B 3 like this that means, that may be diet, that may be physical activity, that may be addiction, these are the different behaviors. You can observe these different behaviors at a particular time point or a particular cross section of time that is called a cross sectional research where because at a particular time point or you know only once you are observing these different behavioral factors ok. So, the cross sectional research it is you know in that way important because it gives you an idea what is actually going on with the individuals ok. So, what is actually going on with the individuals? The difference between the variety of people, the subjects, the phenomena that are being that are there in the population you can identify them through the cross sectional research. And often the cross sectional research is the observational quantitative method that will give rise to certain hypothesis or certain research questions.

So, you can consider cross sectional research as that particular research which is a part of you know the exploration the exploratory part of research. So, what are the different strengths and weaknesses when we discuss about the cross sectional research? As we have seen that in the cross sectional research we are observing a particular individual at only a single time point. So, we can only relate the different factors with that individual, but we cannot conclude any you know causal inference from them. We can relate from a particular cross sectional research that smoking occurrence of smoking and occurrence of lung cancer they are related. You know the prevalence of smoking and through prevalence of lung cancer we may we can have a correlation analysis anything and we can say that they are related, but we cannot say that smoking is causing lung cancer ok.

Why? Because we do not have a separate time point or we do not have the you know the temporal association over here. The temporality means observation which is again done on the same population that is missing. Now, sometimes the cross sectional research it requires you know moderate amount of resources because the sample size may be large for a good inferencing from a cross sectional research, but again you do not have to follow the sample or you do not have to check for these factors again and again at a later date. So, that brings your resource requirement down. Usually the participants there is no question of loss to follow because there is no follow The commitment from the participant is usually typically you know it is it is lower compared to other forms of research and therefore, you may get a better form of response from the participant ok.

So, these are the important strengths that you get from the cross sectional research, but please remember that from cross sectional research you do not get the rate of happening or you know what we call as the incidence you get only the proportion of disease or you may have the prevalence based on what population and how you have sampled. What are the weaknesses? You know these are not good for documenting the rare health outcomes. Why? Because the rare health outcomes as the term says the outcomes are themselves rare. So, it is very difficult to find them if you survey a large number of population. You can utilize a large number of resources, but they not they may not finally, give you the required amount of output that you want.

You may not be able to finally, conclude that this is the actual proportion your analysis it may become flawed. So, it is typically not useful for rare health outcomes or those which have a very short duration that means, very acute type of illness. You will not be having those diseases if you if you go into a community and if you start surveying. For example, you have started the survey today in a particular community and you know that there may be cases of certain cough and cold in that particular area. But on the next date when you go there and you find that there is no case of cough and cold.

So, it is very important you know to design how you are actually going to take the information. Because otherwise the rare diseases or these short duration diseases they will not give you certain amount of information that you actually want from your study. Sometimes it may need extremely large sample size to power the study or to have a good power of inferencing from the study. Because if we have to inform our you know our final causal study from this cross sectional study we have to have sufficient power and we have to have sufficient replication in this cross sectional study that can allow us to you know finally, test an intervention or test a more specific hypothesis in the causal arena. For that you may need an extremely large sample size.

So, in cross sectional study you can do a cross sectional study with a moderate number of sample size when you are typically exploring you do not know what hypothesis you may get. But once you have a hypothesis or you have a hint of the hypothesis you may need an extremely large sample size to get a good power analysis to get you know the your correlations or your association analysis to be powerful enough to bring you certain inferences. There may have been certain selection biases that usually happen because you are again taking it at a particular cross section you usually call it a cohort effect. The cohort effect means the different ages or the generation effect as we call it what happens in a generation effect in a particular cross sectional study only the respondents from a particular generation may be included that is called a cohort or a generation effect. Sometimes you can have a third variable problem why a third variable problem? You omit certain significant variables in your cross sectional study and since you are doing the study in a single point of time you do not have any particular scope of rectifying the study unless you have conducted your pilot study in such a way that the pilot study will be you know have given you certain inputs that you can actually go on with these variables without having a particular third variable problem.

This brings us to the question of survey research. What is the survey research? Survey research it is basically an extension of a cross sectional research which basically gives you know it depends on the method of data collection what you are doing. We often say the agencies they perform different surveys. One example is the online surveys that become so much popular during the COVID-19 pandemic. What do we do? We select certain outcomes and we ask the participants or you know the general public or the target population about how they perform that particular behavior or that particular practice.

So, that is your survey. What is unique to the survey research? It must have a very much well-defined sampling frame because your survey research will ultimately give you certain actionable prevalences. Based on these prevalences you can act on whether to go for any further research or to advocate for any policy action like this. So, your sampling frame must be very much well defined. In other words you have to have you know very well built technical mechanism in your research.

You have to find out the modality of research. How do you want to administer the survey? The interesting development regarding modality during the COVID-19 pandemic has been the self-administered survey that means, the beneficiary is administering the survey by himself or herself, but that is administered through online mode. That means, the surveyor you cannot directly see the surveyor, but you just can see the questionnaire and you can fill up the questionnaire like this. So, we will be discussing the modality modalities and all the other things in the next slide. So, that is one important issue.

Sample size that we have discussed in the previous lecture that is often and very important issue because that will you know the an effective and an appropriate samples will ultimately help you achieve the proper power of inferencing. So, that is very important. The manual operation of procedures the MOP or the standard operating procedures whatever you call it, you have to specify it because the same thing has to be done again and again on the large sample that you have selected. So, that has to be reliable and that has to be valid. You have to create the questionnaire, then you have to pilot test it and then after pilot testing you need to make certain amendments to the questionnaire based on the results of your pilot test.

So, that the problems like you know we mentioned about omitting certain significant variables these kind of problems they do not occur ok. So, these are the essential things to have a good survey and you know survey research that is very important in health promotion because you can find certain causal questions in health promotion through these surveys. For health promotion you need a larger sample and we attain a larger sample through simple surveys often. As we have been discussing the survey modalities, the pros and cons and the appropriate sampling in terms of the modalities for example, if we have discussed three modalities the computer administered, survey modality, the interviewer administered and the self administered. Now the computer administered this can also be the online surveys and the interesting part is it is kind of a mix of self administered and the and the typical computer administered survey because unless the participant is able to understand the questionnaire or is able to is motivated enough to fill up the questionnaire you really cannot complete a computer administered survey.

So, what happens with computer administered survey? You can have you know the IVRS system, the interactive voice system or the chat system that can be there or the simple questionnaire or the online questionnaire that can also be there. The main problem is you know sometimes the questions may not be clear enough to the participants and the participants may just go on and take as per his or her own will without even noting the question because there is no personal contact from the interviewers end with the participant that is a very big problem and also another ethical issue is the trust and confidentiality of responses that they have given in the questionnaire that is also a very important question that the participants are

going to rise I mean that is also another thing that you have to answer to them. Then the important thing in this case is what type of sampling should we follow? See for computer administered random sampling via the internet, random sampling via telephone because you can have you know have the telephone directory and you can sample from them because often it is considered that the telephone directory is basically this is basically the universe it has all the individuals noted within that directory. But again there are problems with that because not all the individuals may be using that telephone or some may even use simply only the mobile phone nowadays. Although there are problems, but these are certain sampling options that we have with the computer administer or the online service.

Sometimes you can even perform online surveys based on the you know utilization of services in different social media. The participants those are registered with different social media you can actually form a frame for your study, but ultimately as with all other surveys that frame will not be an exhaustive one. Your job will be to form as much exhaustive as possible regarding the sampling frame. What happens with the interviewer administered? Suppose you are the interviewer here you are directly going to the participant and you are administering the survey. So, you are asking the questions.

Here random sampling via phone can be an effective option why because you are calling randomly the individuals and you are directly administering the survey. The key thing is in a random I mean in a phone based survey you are not having a personal contact with the individual and all the problems that arise from not having a personal contact with the respondent they will be present in this case also ok. And since you are noting down the responses it may so happen that your certain subconscious bias that may affect the final result that you are noting. And you know advantages is that since you are asking the question the interviewer must not I mean the beneficiary or the participant need not to go through the questionnaire and follow whether there is any skip patterns or something like that that if yes then answer this question if no directly go to this. Since you are only interviewing you can directly follow the commands that is given there in the questionnaire and that save the participant a lot of time.

And therefore, the commitment of the participant remains intact in the end of the questionnaire that is very important as well. Now, what can be self administered questionnaire? Self administered or self administered survey that means, here the beneficiary or the participant himself or herself is filling up the surveys typically with the with the study schedules and everything. Here you email the surveys to the random households now these random household selection may be done through the post office system you can email them with a greetings note or some you know some note regarding recruitment in the study like that. The question is they may be completed by students randomly in school this is an example students in school they can be done randomly because you just give the questionnaire consider it as an exam paper writing and similarly what some of the students they will fill up the questionnaire and they

will give it back to you or you can directly select some of the students and give them the questionnaire and they can just fill it up and give it back to you. So, that is self administered self administered also has certain advantages and it is a disadvantage is the student is able or the participant is able to understand what the question actually is mentioned over there.

And the problem is sometimes this may be too time taking and the participants may you know lose their commitment that they have given to you. Then comes the issue of successive independent samples what happens with successive independent samples it is simple you get multiple samples and those multiple samples are taken at different time points for example, this is oh my god. So, this is you know. Now, what happens with successive independent samples is that you get samples at different points sorry different points means you get samples for example, in 2019 then you get sample in 2000 and say 20 then you get sample say in 2021 year 1, year 2, year 3. The individuals you surveyed in year 1, year 2 and year 3 they may be different or there may be some overlap what you essentially get is a pattern of all the behaviors that you are studying for example, Tobago use or condom use or the fertility pattern in a community.

So, what happened during year 1, year 2 and year 3 with the help of the successive independent samples you can get a trend. So, often this is called a trend study. As we have understood what are the strengths and weaknesses with the successive independent samples it gives you somewhat of a temporal association. So, you can you know understand what happens first and what happens next, but still you cannot properly make a causal inference from this right that is why I mentioned cannot identify the causes, but you can understand which is occurring first and which is occurring later like that. And with the trend studies what happens because each and every sample or each and every survey is done on a single time point and that is essentially a cross sectional survey in itself.

Sometimes we call this as a repeated cross sectional design as well what happens there is a low attrition rate because it is not your onus to select the same population during the next phase of the survey ok. So, that is how this the trend studies are helpful if you are only trying to find a very good evidence to generate your hypothesis or before you know ultimately going for a large scale study the trends will help you. The longitudinal cohort designs longitudinal means it is same as you have you know selected this 1 year, 2 year, 3 year, 4 year only thing is it is not the different individuals that you are studying. You have a same group of individual A is your group and you are studying them throughout these year 1, year 2, year 3, year 4. This is the longitudinal pattern of study longitudinal design means you are studying the same individual on multiple occasions or on multiple time points.

Now a special consideration for this longitudinal design is called a cohort design what happens with a cohort often we use the term cohort in terms of a group of people. For example,

a group of a cohort of all those who were born in the year of 2003 or 2000 like that, but cohort design means it is a longitudinal design where you have 2 groups of people you have A 1 and you have A 2. Who is A 1 and who is A 2? A 1 means if you are interested in terms of a particular exposure whether that is going to cause the disease or not those who are having the disease is A 1 and those who are not having the disease is A 2 disease plus and disease minus and at the end here you are selecting whether the outcome of interest it has actually occurred or not. So, this is a cohort design and in longitudinal design longitudinal design is somewhat similar to what do we do in the cross sectional design. We simply take associations, but we really cannot comment on what is actually going to cause that particular thing.

The strengths and weaknesses that are you know mentioned over here are mostly focusing on the cohort design as the trademark for the longitudinal studies. Here the main important factor is there is no cohort or generation effect that means you are following the participants. So, it is not really about a particular generation they are getting involved in the study you can follow them up for a sustainable duration of time was or you know some 10 years or 20 years. You know the example of Framingham heart study as well long duration of follow ups are there and that will help you omit this cohort or generation effect. Also you know you can study different exposure factors through a single cohort study, but the point is you really cannot expect a cohort study to yield you result in a very short duration of time that is why we have mentioned it is time consuming and since it is time consuming and resource intensive it will become expensive.

It really requires sometimes larger sample size, but the large sample size may not be as large as you require in the case of the cross sectional studies, but still if you have to have a good power in the study you must have a larger sample size. And typically if you are doing a simple observational longitudinal study where you do not have a comparison group inbuilt in your study, then you typically require a larger sample size as compared to a cohort study ok. There is a problem of attrition and missing data and the individuals measured multiple times may result in testing effects because here typically what you have to do here you have to consider the correlation between the data measurements because it is the single person you are measuring again and again. For example, if I am being surveyed today I will try to understand what are the questions and some amount of that question will be stored in my mind. So, if I am again surveyed after 1 year what will happen is now I will you know have a certain degree of bias or I will try to replicate the same questions or I may try to do certain other logical things.

So, that will you know hamper the originality in responses and the responses may get correlated or the clustering of responses this we call the clustering effect due to time because we are taking the samples over and over again. What is a cohort sequential design? So, the sequential means we are taking in individuals 1, 2, 3 like different sequences different time points. So, cohort sequential means it starts with an initial cohort and follows it over for a

period of time and another cohort or cohorts are added in terms of the times or whatever we get. So, that means, if this is a cohort this cohort is being followed for 3 years. Now this is the cohort A sorry cohort A cohort A we can add a cohort B with certain other characteristics at the end of 1 year and you can really follow up with another cohort C which is followed up for only 3 years.

Now you can compare the results among these 3 cohorts and also within these 3 cohorts. So, you have the individual cohort effects and also among the among the different cohort effects. This is a better you know this is an improvement over the typical cohort studies and you can have certain robust estimations from compared to the typical longitudinal observational study as well. So, that is what we have mentioned this weaknesses are same as what we get in cross sectional and typical longitudinal study. Although you can compare, but still if you are not really following you know that particular individual through time that means, what he is doing and after that what is developing you cannot really confidently infer on the causality of the of those issues that you are considering.

So, here also it controls for the generation effect because obviously, the same things as in the in the cohort or longitudinal study you are following up. So, there may not be any generation effect in this study. What is a case control design? It is more or less related to the cross sectional design. You have individuals who have the disease and you have individuals who do not have the disease you simply compare the exposures that you have. So, here also you are measuring both of these in a single time point t_1 is your time point.

So, again it is a modification over a typical cross sectional design here you are observing whether the person is having disease or not and your next you are observing the exposure that the person is has suffered. So, that part you are taking from the history. The strengths and weaknesses typically you cannot study the rare exposures because you already have the outcome if the exposure is rare then it is difficult to correlate with the rare exposure. If there may be recall bias because you are taking the history in the case of case control studies, but it is well suited for rare disease because once you have the disease you have the controls who that is those who do not have the disease and you also have the battery of exposure variables that you have already listed. So, this is a very good design for the rare diseases and remember the cohort studies they are very good design for the rare exposures, but not for the rare diseases because ultimately at the end of the cohort if you do not get the disease then you will not be able to do any scientific inferencing.

The case crossover design this is typically a hybrid design and remember this is not the crossover trial we will be discussing crossover trial when we discuss the experimental designs. Again this is one of the longitudinal designs that we were discussing here the cases they serve as their own control. So, if you if I am considering this individual one over this time point say

this is t_1 and t_2 in time point t_2 suppose this person has developed disease plus and in time point t_1 at this person has not developed the disease this person this disease positive person can serve as the control of this disease negative person although they are the same individuals. And the duration of exposure for example, if you have an exposure over here after which the person has become disease positive is called the hazard period where you have the exposure to that person ok. So, these are certain improvements over typical longitudinal designs here you can easily you know infer certain causal effects and even for very short duration of exposure you can also attribute that exposure to occurrence of certain things.

For example, if there is an acute onset event for example, due to over speeding you have an accident you can attribute through the case crossover design. That is the example that we were discussing texting and riding a cycle. So, is there a causal relationship compare the amount of text that has been sent this is how you measure the effect of texting on the day prior to the accident. I mean you compare the amount of text for example, text you have given the person has sent 100 text and has and that has led to the accident. In the end you compare it on the day of accident with any other day there may be text 10 text and there was no accident.

So, you can compare that once the individual is actually texting and riding the bicycle and then the you can meet with the accident and you can establish a degree of causality. Although this is just an just a rough example you can find out examples on your own as well. So, these are the strengths and weaknesses as we have been discussing for all the research designs here you know inaccurate recall of exposures during the control windows when the event has actually not occurred the person may not recall properly the information that is required. It may be the case that the person is saying during the accident I texted much, but if you consider it may so have happened that in the previous day also the person has you know texted like 100 text the same text. So, that may be the case of inaccurate recall because sometimes you know what happens is a person is influenced from his or her own I mean its own self.

So, that can happen in terms of the case crossover design, but the efficiency in the design is very much because it is self matching. So, it requires comparatively lesser amount of subject although not very few, but still comparatively lesser it eliminates the between person confounding because if you have to include one other control in terms of case control study or cohort study then there may have been certain other confounding factors and there may be certain factors which we do not even know as the confounding factors ok. So, that is excluded because it is the same person who you are studying and you can use multiple control windows because if you can also use multiple hazard periods as well that is what we have mentioned with multiple control windows. So, in conclusion there are different study designs that can be used in health promotion we discussed different observational quantitative designs and we discussed how cross sectional longitudinal designs are different. We discussed the importance of the surveys the surveys often give us with you know they give certain important questions

that we can further study through the different longitudinal designs or the different cohort studies or these case you know the case crossover designs.

Typically these case crossover designs are very much interesting and they are very much important in terms of studying the different health behaviors because they utilize the hazard period and for different behaviors the hazard period is typically more important than counting that total behavior as a perpetual thing. So, these are the different issues that you get when you select the different observational quantitative designs that we have already discussed and typically there are different design where you get the temporality included event is occurring and you know the exposure that has occurred before the event and you can relate those are the typical cause effect relationships and you can also have these cause effect relationships through these models that we have discussed in last couple of slides. So, these are the resources that I urge you to follow That is all for this lecture. Thank you very much.