Lecture Notes on Epidemiological Study Designs

Epidemiological study designs refer to the principles and methods of conducting systematic investigations where a specific question is answered using data collected in the field. For example, if we aim to study the prevalence of asthma among children under five years of age in Canterbury, we can conduct a cross-sectional survey of all children under five years of age who live in Canterbury, and then conduct a door to door survey of children in the households and search for the presence of children with asthma. In order to do so, we will need to identify a population, a sample, define a health condition, have a plan of searching for the outcome, and conduct the investigation.

Two different classes of study questions are explored in study designs: whether we want to generate a hypothesis or whether, after generating the hypothesis, we want to test such hypothesis. Accordingly, the study designs are selected.

- When the idea is to generate a research hypothesis, you can use case series or case study design
- In case study design, a single case is described in details to generate a set of hypothesis
- In case series study design, you can describe in details the features of a series of cases.
- A form of case series is establishment of Disease Surveillance programme, where from different sources in the community, health services experts obtain data about cases that have occurred, or reported, for example, reports of specific diseases from clinics, nursing units, or schools and indeed any place from where illnesses or health states can be determined.
- Cancer surveillance programmes depend on cancer registry which also follow similar pathways
- When the idea is to test a research or study hypothesis and you want to obtain individual level data, you can conduct a nuber of different studies

Study Designs to test hypothesis

For individual level studies, you can conduct a cross-sectional survey, a case control study design, or a cohort study design

Cross-sectional surveys are those where a section of the population are studied to ascertain or find out the prevalence of a specific health related state. As part of the study design, researchers can also study other exposures and outcomes. For example, in a study on finding the prevalence of people with high blood pressure in Canterbury, researchers identified households and they randomly travelled to the households, and measured blood pressure of all the members of that household. After a fixed period of time, the researchers reported the prevalence of high blood pressure on the basis of their findings, and also because they administered quesionnaires to understand the possible reasons why some people were suffering from high blood pressure not others, they were able to provide additional analyses.

Advantages of Cross-sectional studies

- Relatively simple study design that can be rapidly conducted with little investment of resources
- This study design is not time consuming
- You can study multiple exposures and outcomes all at the same time

Limitation of Cross-sectional studies

- Because information on exposure and outcome are collected at the same time, a definite causal association cannot be established
- These studies are open to selection bias. For example, if the researchers were conduct a study at a University during the examinations the prevalence of anxiety among students, they would overestimate the prevalence of anxiety.
- Similarly, some cross-sectional studies are open to recall bias or response bias. For example, in a factory, if researchers were to study the prevalence of severe back pain, they would under-estimate the prevalence of low back pain as many people who would have otherwise present would be absent

because of low back pain. Such people would be omitted and the resulting study would suffier from "healthy worker effect" and under-estimation of the true effect

Case Control Studies

Case control studies refer to a study design where the researchers identify people or select people on the basis of their disease outcomes. For example, in a case control study of dementia and prior covid infection, if researchers were to identify people with and without dementia, and then investigated their prior clinical records to search for evidence of their being diagnosed with Covid-19 infection, then such a study would be referred to as a case control study design.

The following table explains the gist of case control study design:

Exposure	Cases	Controls
Exposed	A	B
Not Exposed	C	D
Total	A+C	B+D

Here's how the study works:

- We compare the risk of exposure for Cases to risk of exposure to Controls
- We calculate the Odds of exposure to Cases and Odds of exposure to Controls
- We then compare the two Odds and express the risk measure as Odds Ratio

Using this concept:

- The Odds of exposure to Cases is A over C or A / C
- Odds of Exopsure to Control is B / D
- Hence the Odds Ratio is calculated as (A / C) / (B / D) or A * D / B * C
- Here A, B, C, D are numbers of people
- The Odds Ratio represents the risk of exposure to the disease

Example:
Suppose we are studying the risk of exposure to Covid infection and Dementia

Exposure to Covid-19 infection	Dementia	No Dementia
Exposed	40	50
Not Exposed	60	50
Total	100	100

This is a fictitious example:

- You can see that Odds of Exposure to Covid-19 among those who had Dementia = 40/60 =0.667 0.66
- Odds of exposure to Covid-19 among those who did not have Dementia = 50/50 = 1
- Hence Odds Ratio = 0.67
- Ths Odds Ratio suggests that Exposure to Covid-19 had a "protective effect" on Dementia

How to conduct Case Control Studies

- Start with defining your cases and controls
- Determine your desired effect size, and prevalence of the outcome among the control population, your alpha and beta error
- The estimate on this basis your sample size
- The Cases and controls need to be as similar to each other as possible, so that if a control were to develop the disease under study that control would now be a case
- Conduct the study and estimate the Odds Ratio using the data; a number of data analytical strategies are used, for example logistic regression, or stratified analysis.

Advantages of Case Control Study Design

- You can study rare diseases
- You can study multiple exposure variables for the disease
- The study design itself is straightforward and relatively easy to conduct

- It saves on time and resources

Limitations of Case Control Study Design

- The study design is by nature retrospective: this means we assess exposure and outcome all in the past, the outcome needs to have happened already
- As a result, a temporal sequence of cause and effect cannot be established in this study design
- Open to recall bias and information bias
- Open to selection bias as cases and controls can be differently selected

Cohort Study

A cohort study design is one where people are sampled on the basis of their exposure status. The principles are as follows:

- Sample people who are exposed to an exposure variable of interest
- Sample other people who are NOT exposed to the exposure variable of interest
- Neither the exposed NOR the non-exposed must have the outcome of interest
- Then follow the exposed and non-exposed cohort for a pre-specified period of time
- Determine the incidece of the outcome in both cohorts
- Compare the incidence rates in both cohorts to determine rate ratios or relative risk estimates

The following table provides a schematic concept of cohort studies

Exposure Status	Count	Number years of follow up T	Outcomes	Rate
Exposed	N1	T	A	A/N1*T
Not Exposed	N2	T	B	B/N2*T

These numbers represent numbers of people or length of time.
N1*T and N2*T are referred to as person-times
From the above table we derive:

- Rate among exposed = A/N1*T
- Rate among non-exposed = B/N2*T
- Therefore Rate Ratio (Relative Risk) = (A/N1*T)/(B/N2*T)
- You can also calculate the risk difference = (A/N1*T) - (B/N2*T)
- Risk Difference is also referred to as Absolute Risk Reduction

Example of a Cohort Study (fictitious example)

Suppose you want to study the risk of Dementia following Covid-19 infection
You can start with all people who did and who did not have history of covid-19 within the past five years (excluding all people who have current covid-19 as well), then decide to follow up these people for the next five years to ascertain the rate of dementia among these people. You can set up a table as follows:

Exposure to Covid-19	Count	Number years of follow up T	Had Dementia	Rate
Exposed	500	5	16	0.0064
Not Exposed	500	5	14	0.0056

If we consider the above figures, we see that the rates are not that different between those who were exposed and not exposed to covid-19 when it came to dementia.

- the Rate Ratio of Dementia with Covid-19 = 1.14
- We can also estimate the Rate Difference as = 0.0008 (which is a very small difference between the two)
- This can be further explained as how many people would need to be infected with covid-19 in this case so that we could have ONE EXTRA case of dementia arising from covid-19 infection = 1/0.0008 =1,250 or about 1250 people would need to be infected with covid-19 to have one extra case of dementia assuming that covid-19 is a cause for dementia. You can on this basis may claim that if this

study is true (this is a fictitiious study) then you will need many people to be infected to claim that one in a thousand might suffer from dementia in five years' time.

Advantages of Cohort Study

- It is a strong study design to ascertain if the Exposure might be a Cause of the Outcome
- The study design is prospective, such that you have the exposure first and outcomes occur later
- You can study multiple outcomes for single exposures

Limitations of a Cohort Study Design

- The study design is expensive in terms of time and money
- The study design is not suitable for disease that have long latency
- Latency means that the length of time from beginning of the disease process in the body to its manifestation. For example cancer (any cancer) takes a long time to develop; such diseases are best studied using case control study design

Randomised Controlled Trial

A randomised controlled trial (or RCT) is a study design that is used to test the efficacy of an intervention. Efficacy refers to the effect of an intervention when it is administered in a controlled environment. This is indeed the case with randomised controlled trials where such trials are conducted like experiments in the community. RCTs are conducted within industry or drug companies or indeed in cases or situations where the reseachers are interested to test the effectiveness of an intervention or try out an experimental approach to an intervention.

How to conduct an RCT

- Estimate the desired sample size of an RCT using desired effect size, the alpha error, the power you want, and the expected frequency of the outcome among people who will receive the alternative medication or intervention or placebo.
- Start with a random numbers table
- Using the random numbers table, decide who will receive the intervention and who will receive the alternative intervention.
- Usually, the alternative intervention is another active or traditionally administered intervention. In case of say vaccine studies, you may think of one vaccine versus another. Sometimes, researchers test an active drug against a placebo. A placebo is one which does not work in any real sense of the term but the person who is administered the medication or the treatment believes it to be an active agent. Hence the derivation of the word "placebo" which means I please
- Recruit individuals in both arms of the study and administer the interventions.
- Follow individuals in both arms and observe the incidence of the outcome in both arms
- Calculate the effect size of your study as the difference in the rate of outcomes in the two interventions or as rate ratio in the two interventions

An Example of Randomised Controlled Trial

(Fictitious Example)
Suppose we want to conduct a randomised trial to test the efficacy of a certain vaccine V1 for Covid-19 when compared with another vaccine, say V2, on their abiliity to prevent first onset of Covid-19 infection. We assume that all individuals whom we will include in our study will NOT have had covid-19 infection at all. Then we will obtain 500 such individuals for our study and will divide them into 250 each to receive V1 and V2. Then we will follow them for 18 months and observe how many people are diagnosed with Covid-19 and make comparisons. The conditions are as follows:

Intervention	Number	Number developed Covid-19	Rate
V1 vaccine	250 for 18 months	7	0.0015
V2 vaccine	250 for 18 months	15	0.003

From here, we can calculate the rate difference as 0.0017 (this is also referred to as Absolute Risk Reduction or ARR) where the rate of infection for people receiving V2 vaccine had higher rates of infection.
We can also calculate the rate ratio as 0.46 where which tells us that the risk of covid-19 among those who receive V1 vaccine is LESS than that for those who receive the V2 vaccine, so we may say that on the basis of this study that V1 vaccine is more protective of COVID-19 than V2 vaccine.
We can further explore to calculate the numbers needed to treat (NNT) as 1/ARR and that is in this case is 562. This tells us that we need to treat about 560 people with V1 before it can prevent one MORE case of Covid-19 than V2 vaccine. This is important because as we see that the relative incidence of COVID-19 is very low and this is why this difference is clinically useful information for us to make a statement that any vaccine is useful in this situation.

Advantages of RCTs

- RCTs satisfy almost all of the causal criteria, you can test the temporatliy and experimental evidence aspect of the causal criteria
- RCTs are therefore very highly rated for evidence
- In RCTs, randomisation procedure allows for control of both known and unknown confounding variables; while other study designs allow us for control of known confounders this is the only study design that allows for the control of unknown confoundng variables
- RCTs when they are blinded allow for elimination of selection and response biases. Blinding means that neither the experimenter nor the patient or particiipants would know who was allocated which arm of the trial.

Limitations of RCTs

- RCTs are very expensive study designs both in terms of time and resources
- RCTs suffer from limitations that the study results are not generally extended to the population outside of who were studied as the conditions of selection of participants can be very specific
- RCTs also suffer from limitations when participants cross over from the intervention to the control arm; this happens when the participants may have found out which arm they were on and have switched arms without informing the investigator. When that happens or when participants leave the study arm or the control arm, then it is a dilemma.
- Therefore RCTs are analysed using a strategy referred to as Intention to Treat (ITT) analysis. In ITT analysis, the analyst analysed the results as per the original allocation of the participants in the intervention as opposed to the alternative arm.

Systematic Reviews and Meta Analyses

Systematic reviews and meta analyses refer to the process of pooling of study results. Studies that we have considered so far are referred to as primary study designs because here, we obtain data directly from the participants in our studies. In systematic reviews and meta analyses, we do as follows:

- We begin our study with a well-specified study question using the PICO question format
- Then we search lliterature databases to find out as many studies as possible from this question based on our specified inclusion and exclusion criteria
- Then we use the reference lists of these studies to identify further studies
- Then we abstract data from these studies that we prepare for analyses
- If our primary studies are RCTs or other study designs from where we can obtain numerical estimates, then we can conduct a meta analysis. For conducting a meta analysis, we first test if the study results would be homogeneous. Homogeneity means that the study results suggest that they are similar in direction and magnitiude around a central study estimate. If that is the case, we proceed towards statistical pooling of the study estimates. On the other hand, if we find gross heterogeneity in the study results, that is, the study results suggest that the study results are widely different form each other, we may either conduct a different type of pooling of results or we do not statistically pool the study findings, instead we conduct a thematic analysis.
- A thematic analysis is suitable for systematic review and a statistical pooling of results is a meta-analysis.

Evidence Pyramids

The term "Evidence Pyramid" suggest that the study designs can be arranged in a pyramid shaped manner to indicate the level of confidence we can have with respect to the various risks of bias (selection bias, response bias, confounding, and play of chance). The Evidence pyramid has as its base case studies and case series, and at the apex, it has meta analysis of RCTs. This indicates evidence pooled from well-conducted RCTs are given high priority as good evidence and evidence pooled from case series and case studies are given lower priorities in terms of robustness of evidence.

What to look for when we conduct critical appraisal of evidence

- Start with the study design. If it is a set of RCTs or if you have an RCT to appraise, usually thhis is a good study design
- Next, even within an RCT, check if the randomisation was properly done as for an RCT a proper randomisation is a marker of high quality
- Check if the study even it is an RCT if the investigators and the participants were "blinded" with respect to the allocation of intervention and comparison. This is referred to as "allcoation concealment" and refers to the fact that the investigator and the other staff associated with the study did not know who was allocated what arm of the intervention or the control group
- Read how data were obtained from the participants in the study. This is important and try to assess if there wre scopes of selection or response bias might be possible. Just because a study is a randomised controlled trial does not imply there was no response or selection bias although you may argue that confounding variables were controlled for by design
- Check what analyses were conducted. In case of case control or cohort study, check if the investigators conducted a multivariable analyses; if so, this would indicate that they were able to control for known confoundng variables. Note if the investigators conducted a matched analysis in case of case control study: if true, it would indicate that they were able to control for most known confounding variables (at least the ones on which they conducted the matching)
- In case of RCTs, check if the investigators were able to conduct an Intention to Treat analysis.

Summary

This was a tour of the study designs relevant for evidence based health. In all cases, this knowledge helps you to understand the importance of conducting an evidence appraisal as is built within a study design. Most study designs that are aimed at understanding the causal relationship between an exposure or an intervention and an outcome aims to capture some aspects of causality. Hence randomised controlled trials are viewed as pinnacles of study designs that aim to capture the effectivness of an intervention; case studies and case series cannot capture causal association between an exposure and an outcome but they are useful for understanding the phenomenon of interest and hence useful for generation of hypothesis that can be tested with other study designs. Overall, study designs help you to critically appraise available evidence.