# What type of research study should you do?

You must choose your type of research study based on practical as well as scientific aspects.

**Practical aspects:**

Your study must be do-able (feasible) in the available time that you have. This usually means that you must either use already available data, or that you must link-up to an ongoing study with a "platform" onto which you can easily add a few additional questions. It also for example is not a good idea to start following up a group of children collecting data on them and then run out of time and resources. Now is a good time to look at the timelines in Introduction 3 "How much time do you need?" The worst thing that can happen is for you to start collecting new data and then to realise that you cannot get data on enough children or for a long enough time to make the study scientifically sound.

**Scientific aspects:**

Studies are broadly divided into quantitative and qualitative science:

1. Qualitative studies focus on the analysis of social aspects of diseases and do not involve numbers or statistical methods.
2. Quantitative studies focus on numerical measurements and the data obtained are analysed using statistical methods.

This manual focuses exclusively on quantitative studies. If you want to do a qualitative study, you must get expert advice from and collaborate with social scientists.

Most quantitative studies have two components namely descriptive and analytical. However, to keep it simple, we classify quantitative studies into two main types of studies namely:

1. A descriptive study provides summary information on the data collected. This might be in the form of descriptive statistics such as frequencies, means, minimum or maximum values or graphs such as a bar chart. Case reports and case series are special types of descriptive studies and fall into this category. By far the most descriptive studies and case reports will not be sufficient for you to register at the HPCSA and therefore should not be considered.

2. Analytical studies in which variable are compared to an outcome and comparisons made between those who have a specific variable and those who do not have the varaible. There are two types of analytical studies:

a. Observational (association or comparative) studies e.g. cross-sectional, case-control and cohort observational studies where a specific intervention is not implemented, but where variables (e.g. HIV status) are collected and the outcome (e.g. death) is measured and comparisons made between those with and those without the variable (HIV+ or HIV-)
b. Experimental (interventional) studies e.g. randomized control trial where a specific intervention (e.g. a new drug) is implemented and the outcome measured. These are usually prosepctive cohort studies and not something that junior researchers should do as their first study.

## Descriptive studies

An example of a descriptive study is a case report of a child with a specific condition but without comparing it to what is known in the literature and without stating why this child's presentation is different from what is known. Another example will be the description of how many children were admitted to Tygerberg Children's Hospital and how many of these were HIV infected and merely stating that 100 children were admitted of whom 15 were HIV infected. In this example there is no comparison with data from other hospitals, or with data from previous years or comparing the clinical outcome of the infected and uninfected children etc. The art of research is to develop a critical and inquiring scientific mind and to ask a question that will allow you to do a comparison and to calculate risks, ratios etc.

For a fun example and a real quick read just to see how a simple observation that teaspoons disappear in the tearoom, could be changed into a comprehensive and proper analytical research study. Look at this article published in the BMJ[1]

## Analytical studies:

An analytical study is a study where a deduction is made using a statistical method. An example of this will be to use the data of the 100 children admitted to Tygerberg Children's Hospital (as above) but to calculate and compare the proportion of children with TB in the HIV infected and HIV uninfected children. In this example HIV status is a single variable (the key determinant or independent variable) and TB status is the primary outcome (dependent variable). The two proportions say 2/85 = 0.024 (or 2.4%) in the HIV uninfected group vs 2/15 =

1 Megan S C Lim, Margaret E Hellard and Campbell K Aitken. The case of the disappearing teaspoons: longitudinal cohort study of the displacement of teaspoons in an Australian research institute. BMJ 2005;331;1498-1500 or on *http://www.biostat.jhsph.edu/courses/bio622/misc/Disappearing_teaspons.pdf*

0.133 (13%) in the HIV infected group are then compared with a formal statistical test. The deduction or inference is then made on whether this difference is significant or not.

Fortunately there are only a limited number of analytical study designs and in this step we will concentrate on observational studies where you as a young researcher will not implement a specific intervention, but will collect data and do comparisons/measure associations between a variable and an outcome.

The study design will determine the statistical methods used to analyse your research question. The study design provides a framework for carrying out the research in a systematic way and addresses the two essential elements identified when you developed your research question (Step 4) namely:
1. the key determinant and
2. the primary outcome

*There are four standard types of study design:*
- Cross-sectional design
- Case control design
- Cohort design
- Randomised control trial (not discussed in this guidance as this is not the design a young researcher should select for her/his first study).

There can be a great deal of confusion and discussion around exactly which design will be the best to answer a specific scientific question, even among highly qualified experts. You need to know the basics about study design and an easy way is to develop/draw a simple two-by-two table with 4 blocks ( see also Appendix 3), which will form the basis for your study design, sample size or power calculation and the most basic analysis.

| | | Know Outcome | |
|---|---|---|---|
| | | Pneumonia | No pneumonia |
| Know Determinant | (HIV+) | a | b |
| | (HIV-) | c | d |

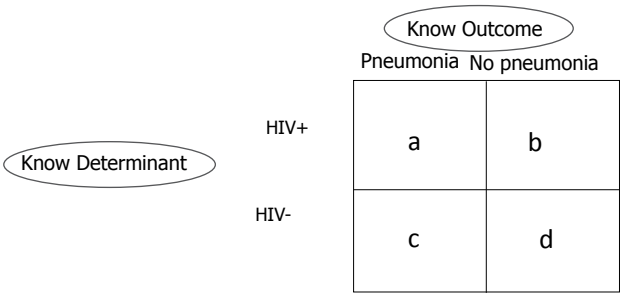
*The four categories found in the two-by-two table are:*

- Those with the determinant and the outcome of interest (a);
- Those with the determinant and without the outcome (b);
- Those without the determinant and with the outcome (c);
- Those without the determinant and without the outcome (d).

For example, if your question is whether children who are HIV+ more often have pneumonia than children who are HIV-, you can use any of the three basic study designs all with specific advantages and disadvantages.

**Cross-sectional study design:**

Relevant when the target population is studied at a certain time or during a specific period of time and when you know the determinant and the outcome at the same time.




You will for example make a list of all the children in the hospital on a specific day/for a specific period of time and from this list you will know the HIV status and whether each child had pneumonia or not and you will classify each child at the same time according to the determinant (HIV+ or HIV-) and the outcome (pneumonia or not pneumonia). There is no longitudinal component. A cross-sectional study differs from a case-control study in that it provides data on the entire population under study (all children in the hospital on a specific day or during a specific period of time), whereas case-control studies include only individuals with a specific outcome (the cases or those with pneumonia), with a selection, often a tiny minority, of the rest of the population (the controls or those without pneumonia).

Cross-sectional studies are important and often the first evidence of associations. For example, the first reports of the association between smoking and lung cancer and also the association between phototherapy and the improvement in neonatal jaundice, were from cross sectional studies.

*Advantages of cross-sectional study design:*

- Simple and easy to collect the data.
- Not expensive.

*Disadvantages of cross-sectional study design:*

- Chronological sequence of events cannot be observed - can only calculate prevalence of pneumonia and HIV status and infer whether there is an association between the two. One cannot determine causality.
- Dependent upon information already recorded with all its existing flaws if one uses already collected data.

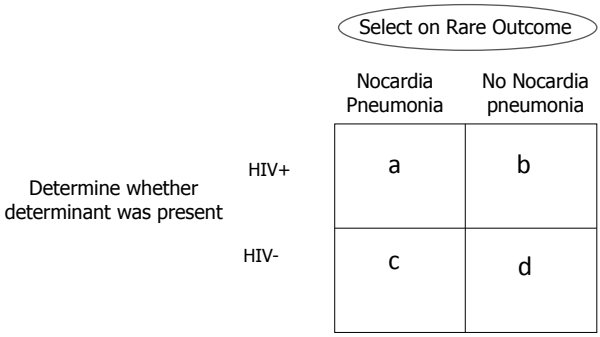
*Statistics for cross-sectional study:*

- Calculate prevalence or proportion and also report precision in the form of confidence intervals.
- Should adjust for confounders that could explain part of the association found between the key determinant and the outcome.

**Case-control study design:**

TIP: This is the only way to study a rare condition.

The case population (e.g. children) is selected according to the rare outcome you want to study. For the example we will use a really rare disease, so let's study Nocardia pneumonia.




You can get access to the information of the children with this rare outcome that has been collected over a known period of time. You include all the cases (those with Nocardia pneumonia) and select the controls (those without Nocardia pneumonia) usually on a 1:3 ratio and collect the same information from the controls as well. The interest in a case-control study is usually a variable that represents exposure in some way. In our example this exposure could be HIV infection. Therefore for each child (cases with Nocoardia pneumonia and controls

without Nocardia pneumonia) determine the exposure status (HIV+ or HIV-). If a larger proportion of the cases (Nocardia pneumonia) have the determinant (HIV+), then there is an association between the determinant (HIV+) and the outcome (Nocardia pneumonia) and the hypothesis is valid. The variables can again be cross tabulated in a 2x2 table.

Because you have to obtain cases and select relevant controls for a case-control study, you must think about the selection criteria very carefully and state these very clearly. The controls must be exactly the same as the cases, except for the primary outcome of the study. For example, you can select children who have Nocardia pneumonia and are in the hospital as cases and compare them to children without Nocardia pneumonia who are also admitted to hospital over the same period as the cases. The assumption is that children admitted to the hospital come from the same communities served by the hospital. Because there will be more children without the rare disease you can select more of them. A ratio of 3 controls for every 1 case is a good option. For example you would select 50 children in hospital who had Nocardia pneumonia and 150 children in hospital who did not have Nocardia pneumonia and look into their hospital files and determine which children were HIV+ and which were HIV-.

*Advantages of case-control study design:*
- Only practical study design to use for studying rare condition.
- Quick and relatively cheap.
- Sample size is economical.
- Easy to identify cases.
- Incident case will allow good planning.

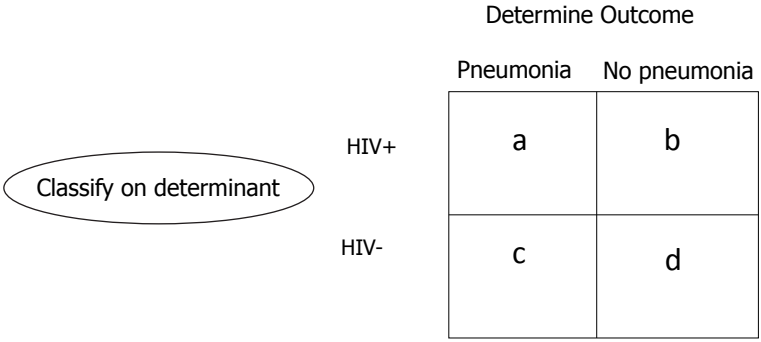
*Disadvantages of case-control study design:*
- Cannot study the sequence of events and therefore cannot conclude whether or not a determinant is a cause.
- Cannot measure incidence.
- Difficult to ensure representative controls.
- Non-standardised methods of measurement if using already collected data.
- Data on exposure of interest not complete or not available on controls.

*Statistics for case control study:*
- Calculate prevalence or proportion of the exposure in the cases and the controls
- Calculate the odds ratio of disease in the cases and controls with 95% confidence interval


When the sequence of events cannot be studied, it is not possible to determine causality. For fun go to *http://tylervigen.com/* to see how careful one needs to be about associations.

**Cohort study design:**

The population (e.g. children) is classified according to the presence of the key determinant of the study at baseline into two groups (for example, HIV+ or HIV-) at the time the study is started, and then all the children are followed over time to determine the outcome (pneumonia or not pneumonia). The timeframe will depend on the expected incidence (of the outcome) in the two groups.




You can do a prospective cohort study by starting at a specific time and collecting information of all children admitted to the hospital and classify them as they are admitted into two groups as having the determinant or not (HIV+ or HIV-) and then follow them over time to see who will develop the outcome (pneumonia). The risk of doing a prospective cohort is that you may end up spending a lot of time and in the end not have enough children (not enough power due to a small sample size) to reach a conclusion or, you can classify a number of children but not be able to follow them up to determine if they develop the outcome (they move, they die, cannot be traced).

Another way of answering your question using a cohort design will be to do a retrospective cohort. For this design you use already collected data of all children admitted say in 2008 and classify them into those with and without the determinant (HIV+ and HIV-) and use the already collected data to determine the outcome (those who developed pneumonia or not pneumonia from 2008 until the time when you do your study). If more children who were HIV+ developed pneumonia than those who were HIV-, there is an association between the determinant (HIV+) and the outcome (pneumonia) and the hypothesis is valid.

*Advantages of cohort studies:*

- Sequence of events can be observed.
- Incidence (Rate) can be calculated.
- Several determinants can be studied simultaneously.
- Standardised methods can be used to measure.

*Disadvantages of cohort studies:*

- Large population required especially when the outcome is uncommon or when several determinants are considered.
- Long time-scale.
- Expensive in resources.
- Drop-outs from cohort over time may bias results.
- Standard methods and criteria may drift.

*Statistics for cohort study:*

- Calculate incidence and precision.
- Calculate risk measures such as risk difference, risk ratios or hazard ratios if time to event is known.

## International Standards for reporting studies:

To ensure that reported studies comply with international standards there have been numerous statements and checklists developed. These guidelines have been developed for various study designs. There is even a guideline for reporting case reports. It is useful to look through the checklists of these guidelines to ensure that your study design complies with the requirements so that you can later easily publish your research.

We found the STROBE statement the most useful as this gives guidance for case-control, cohort and cross sectional studies.

*The following are just a few of the published guidelines:*

| | |
|---|---|
| 1. STROBE. | Observational studies in epidemiology<br>*www.strobe-statement.org* |
| 2. CONSORT : | randomised case control studies<br>*www.consort-statement.org* |
| 3. CARE: | reporting case reports<br>*www.care-statement.org* |
| 4. PRISMA: | systematic reviews and meta-analysis<br>*www.prisma-statement.org* |
| 5. STARD: | diagnostic tests<br>*www.stard-statement.org* |
| 6. SQUIRE: | effectiveness of interventions to improve care<br>*www.squire-statement.org* |

There are many other available guidelines for genetic, economic studies etc.

Gie, R., & Beyers, N. (2014). Getting started in clinical research: Guidance for junior researchers. Cape Town: Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University.