

# Study designs: Part 3 - Analytical observational studies

Priya Ranganathan, Rakesh Aggarwal<sup>1</sup>

Department of Anaesthesiology, Tata Memorial Centre, Mumbai, Maharashtra, <sup>1</sup>Director, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India

**Abstract** In analytical observational studies, researchers try to establish an association between exposure(s) and outcome(s). Depending on the direction of enquiry, these studies can be directed forwards (cohort studies) or backwards (case–control studies). In this article, we examine the key features of these two types of studies.

**Keywords:** Case–control study, cohort study, epidemiologic methods

**Address for correspondence:** Dr. Priya Ranganathan, Department of Anaesthesiology, Tata Memorial Centre, Ernest Borges Road, Parel, Mumbai - 400 012, Maharashtra, India.  
E-mail: drpriyaranganathan@gmail.com

## INTRODUCTION

In a previous article<sup>[1]</sup> in this series, we looked at descriptive observational studies, namely case reports, case series, cross-sectional studies, and ecological studies. As compared to descriptive studies which merely describe one or more variables in a sample (or occasionally population), analytical studies attempt to quantify a relationship or association between two variables – an exposure and an outcome. As discussed previously, in observational analytical studies, the exposure is naturally determined as opposed to experimental studies where an investigator assigns each subject to receive or not receive a particular exposure.

## COHORT STUDIES

A cohort is defined as a “group of people with a shared characteristic.” In cohort studies, different groups of people with varying levels of exposure are followed over time to evaluate the occurrence of an outcome. These participants have to be free of the outcome at baseline. The presence or absence of the risk factor (exposure) in each subject is recorded. The subjects are then followed up over time (longitudinally) to determine the occurrence of

the outcome. Thus, cohort studies are forward-direction studies (moving from exposure to outcome) and are typically prospective studies (the outcome has not occurred at the start of the study).

An example of cohort study design is a study by Viljakainen *et al.*, which investigated the relation between maternal vitamin D levels during pregnancy and the bone health in their newborns.<sup>[2]</sup> Maternal blood vitamin D levels were estimated during pregnancy. Children born to these mothers were then followed up until 14 months of age, and bone parameters were evaluated. Based on the maternal serum 25-hydroxy vitamin D levels during pregnancy, children were divided into two groups – those born to mothers with normal blood vitamin D and those born to mothers with low blood vitamin D. The authors found that children born to mothers with low vitamin D levels had persistent bone abnormalities.

## Advantages of cohort studies

- For an exposure to be causative, it must precede the outcome. In a cohort study, one starts with subjects who are known to have or not have the exposure and are free of the outcome at the start of the

Access this article online	
Quick Response Code:	Website: www.picronline.org
	DOI: 10.4103/picr.PICR_35_19

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

**How to cite this article:** Ranganathan P, Aggarwal R. Study designs: Part 3 - Analytical observational studies. *Perspect Clin Res* 2019;10:91-4.

study, and the outcome develops later. Hence, one is certain that the exposure preceded the outcome, and temporality (and therefore probable causality) can be established. In the above example, one can be certain that the maternal vitamin D deficiency preceded the bone abnormalities.

- For a given exposure, more than one outcome can be studied. In the above example, the authors compared not only bone growth but also the age at which the babies born to low and high vitamin D mothers started walking independently.
- In cohort studies, often several exposures can be studied simultaneously. For this, the investigators begin by assessing several 'exposures', for example, age, sex, smoking status, diabetes, and obesity/overweight status in every member of a population. The entire population is then followed for the outcome of interest, for example, coronary artery disease. At the end of the follow-up, the data can then be analyzed for several contrasting cohorts defined by levels of each "exposure" – old/young, male/female, smoker/nonsmoker, diabetic/nondiabetic, and underweight/ideal body weight/overweight/obese, etc.

#### Limitations of cohort studies

- Cohort studies often require a long duration of follow-up to determine whether outcome will occur or not. This duration depends on the exposure-outcome pair. In the above example, a follow-up of at least 14 months was used. An even longer follow-up over several years or decades may be necessary – for instance, in the above example, if the investigators wanted to study whether maternal vitamin D levels influence the final height of a person, they would have needed to follow the babies till adolescence. During such follow-up, losses to follow-up, and logistic and cost issues pose major challenges.
- It is not uncommon for one or more unknown confounding factors to affect the occurrence of outcome. For example, in a cohort study looking at coffee drinking as a risk factor for pancreatic cancer, people who drink a large amount of coffee may also be consuming alcohol. In such cases, the finding that coffee drinkers have an increased occurrence of pancreatic cancer may lead the investigator to incorrectly conclude that drinking coffee increases the risk of pancreatic cancer, whereas it is the consumption of alcohol which is the true risk factor. Similarly, in the above study, the mothers with low and high vitamin D levels could have been different in another factor, e.g. overall nutrition or socioeconomic status, and that could be the real reason for the differences in the babies' bone health.

#### Uses of cohort studies

- Since cohort study design closely resembles the experimental design with the only difference being lack of random assignment to exposure, it is considered as having a greater validity compared to the other observational study designs.
- Since one starts with subjects known to have or not have exposure, one can determine the risk of outcome among exposed persons and unexposed persons, as also the relative risk.
- In situations where experimental studies are not feasible (e.g., when it is either unethical to randomize participants to a potentially harmful intervention, such as smoking, or impractical to create an exposure, such as diabetes or hypertension), cohort studies are a reasonable and arguably the best alternative.

#### Variations of cohort studies

Sometimes, a researcher may look back at data which have already been collected. For example, let us think of a hospital that records every patient's smoking status at the time of the first visit. A researcher may use these records from 10 years ago, and then contact the persons today to check if any of them have already been diagnosed or currently have features of lung cancer. This is still a forward-direction study (exposure traced forward among exposed and unexposed to outcome) but is retrospective (since the outcome may have already occurred). Such studies are known as 'retrospective cohort studies'.

Large cohort studies, such as the Framingham Heart Study or the Nurses' Health Study, have yielded extremely useful information about risk factors for several chronic diseases.

#### CASE-CONTROL STUDIES

In case-control studies, the researcher first enrolls cases (participants with the outcome) and controls (participants without the outcome) and then tries to elicit a history of exposure in each group. Thus, these are backward-direction studies (looking from outcome to exposure) and are always retrospective (the outcome must have occurred when the study starts). Typically, cases are identified from hospital records, death certificates or disease registries. This is followed by the identification and enrolment of controls.

Identification of appropriate controls is a key element of the case-control study design and can influence the estimate of association between exposure and outcome (selection bias). The controls should resemble cases in all respects, except for the absence of disease. Thus, they should be representative of the population from which the cases were drawn. For

instance, if cases are drawn from a community clinic, an outpatient clinic or an inpatient setting, the controls should also ideally be from the same setting.

Sometimes, controls are individually matched with cases for factors (except for the one which is the exposure of interest) which are considered important to the development of the outcome. For example, in a study on relation of smoking with lung cancer, for each case of lung cancer enrolled, one control with similar age and sex is enrolled. This would reduce the risk of confounding by age and sex – the factors used for matching. Sometimes, the number of controls per case may be larger (e.g. two, three, or more).

Furthermore, to minimize assessment bias, it is important that the person assessing the history of exposure (e.g., smoking in this case) is unaware of (blinded to) whether the participant being interviewed is a case or a control.

For example, Anderson *et al.* conducted a case-control study to look at risk factors for childhood fractures.<sup>[3]</sup> They recruited cases from a hospital fracture clinic and individually matched controls (children without fractures) from a primary care research network. The cases and controls were matched on age, sex, height, and season. They found that the history of previous use of vitamin D supplements was significantly higher in the children without fractures, suggesting an inverse association between vitamin D supplementation and incidence of fractures.

#### Advantages of case-control studies

- Case-control studies are often cheap, and less time-consuming than cohort studies.
- Once cases and controls are identified and enrolled, it is often easy to study the relationship of outcome with not one but several exposures.

#### Limitations of case-control studies

- In case-control studies, temporality (whether the outcome or exposure occurred first) is often difficult to establish.
- There may be a bias in selecting cases or controls. For instance, if the cases studied differ from the entire pool of cases of a disease in an important characteristic, then the results of the study may apply only to the selected type of cases and not to the entire population of cases. In the above example,<sup>[3]</sup> the cases and controls were derived from different sources, and it is possible that the children that attended the hospital fracture clinic had different socioeconomic backgrounds to those attending the primary care facility from where controls were enrolled.
- Confounding factors, as discussed in cohort studies, also apply to case-control studies. For instance, the children with fractures and controls could have had different

overall food intake, milk intake, and outdoor play time. These factors could influence both the likelihood of prior use of vitamin D supplements (exposure) and the risk of fracture (outcome), affecting the measurement of their association.

- The determination of exposure relies on existing records or history taking. Either can be problematic. The records may not contain information on exposure or contain erroneous data (e.g., those collected perfunctorily). This is particularly challenging if the missing or unreliable data are more likely to be present in one of the two groups being compared – cases or controls (misinformation bias). During history taking, cases may be more likely to recall exposure than controls (recall bias), for example, the mother of a child with a congenital anomaly is more likely to recall drugs ingested during pregnancy than a mother with a normal child. In the study by Anderson *et al.*,<sup>[3]</sup> the mothers of children with fractures could have underestimated the amount of vitamin D their children have received, believing that this was the reason for the occurrence of fracture.
- Finally, since case-control studies are backward-directed, there is no “at risk” group at the start of the study; therefore, the determination of “risk” (and relative risk or risk ratio) is not possible, and one can only estimate “odds” (and odds ratio). For a detailed discussion on this, please refer to a previous article.<sup>[4]</sup>

#### Uses of case-control studies

- Case-control studies are ideal for rare diseases, where identifying cases is easier than following up large numbers of exposed persons to determine outcome.
- Case-control studies, because of their simplicity and need for fewer resources, are often the initial study design used to assess the relationship of a particular exposure and an outcome. If this study is positive, then a study with more complex and robust study design (cohort or interventional) can be undertaken.

#### A special variation of case-control study design

Nested case-control design is a special type of case-control study design which is built into a cohort study. From the main cohorts, participants who develop the outcome (irrespective of whether exposed or unexposed) are chosen as cases. From among the remaining study participants who have not developed the outcome, a subset of matched controls are selected. The cases and controls are then compared with respect to exposure. This is still a backward-direction (since the enquiry begins with outcome and then proceeds toward exposure) and retrospective study (since outcomes have already occurred when the

study starts). The main advantage is that since one knows that the outcome had not occurred when the cohorts were established, temporal relation of exposure and outcome is ensured.

#### **Financial support and sponsorship**

Nil.

#### **Conflicts of interest**

There are no conflicts of interest.

#### **REFERENCES**

1. Ranganathan P, Aggarwal R. Study designs: Part 1 – An overview and classification. *Perspect Clin Res* 2018;9:184-6.
2. Viljakainen HT, Korhonen T, Hytinantti T, Laitinen EK, Andersson S, Mäkitie O, *et al*. Maternal vitamin D status affects bone growth in early childhood – A prospective cohort study. *Osteoporos Int* 2011;22:883-91.
3. Anderson LN, Heong SW, Chen Y, Thorpe KE, Adeli K, Howard A, *et al*. Vitamin D and fracture risk in early childhood: A case-control study. *Am J Epidemiol* 2017;185:1255-62.
4. Ranganathan P, Aggarwal R, Pramesh CS. Common pitfalls in statistical analysis: Odds versus risk. *Perspect Clin Res* 2015;6:222-4.