



Special Article

Scientific Methodology Vascular Surgery

Analytical observational studies

Estudios observacionales analíticos

Antonio Martín Conejero¹, Marcos Alonso García²

¹Department of Angiology, Vascular Surgery, and Endovascular Surgery. Hospital Clínico San Carlos. Madrid, Spain. ²General Directorate of Public Health. Ministry of Health. Community of Madrid. Madrid, Spain

Abstract

Observational and analytical epidemiological studies are those that allow the study of causal relationships and associations between variables, without the researcher assigning or introducing the study factor into the sample. That is, the researcher limits himself to observing how an exposure gives rise to an outcome. There are different types of observational studies, some prospective and others retrospective. In this article we will see some of the most common designs, such as the cohort study and the case-control study.

Keywords:

Analytical.
Observational. Cases-
controls. Cohorts.

Resumen

Los estudios epidemiológicos observacionales y analíticos son aquellos que permiten estudiar relaciones causales y asociaciones entre variables sin que el investigador asigne o introduzca en la muestra el factor de estudio. Es decir, el investigador se limita a observar cómo una exposición da lugar a un desenlace. Existen diferentes tipos de estudios observacionales, algunos prospectivos y otros retrospectivos. En el presente artículo veremos algunos de los diseños más habituales, como el estudio de cohortes y el estudio de casos y controles.

Palabras clave:

Analítico.
Observacional. Casos y
controles. Cohortes.

Received: 22/06/2023 • Accepted: 22/06/2023

Conflict of interest: the authors declare no conflict of interest.

Martín Conejero A, Alonso García M. Analytical observational studies. *Angiología* 2023;75(6):385-390

DOI: <http://dx.doi.org/10.20960/angiologia.00544>

Correspondence:

Antonio Martín Conejero. Department of Angiology, Vascular Surgery, and Endovascular Surgery. Hospital Clínico San Carlos. C/ del Prof. Martín Lagos, s/n. 28040 Madrid, Spain
e-mail: amartinconejero.cto@gmail.com

INTRODUCTION

Like we've already been discussing in other articles published (1), analytical designs are those that allow us to draw conclusions after comparing groups of patients. They are considered methodologically superior to descriptive studies precisely because of this inter-group comparison design. Let us not forget that in descriptive studies, there is only 1 group or cohort, which limits the possibilities of analyzing relationships.

We've also seen (1) that analytical studies can be experimental (which is something we'll jump into in further articles that will be published) and observational; these are characterized by the fact that the researcher does not introduce the study factor or assign that factor to 1 particular group of participants.

In this presentation, we will discuss the well-known cohort and case-control studies, each of them with their most common variants.

COHORT DESIGN

Prospective cohorts

Cohort designs are those that start with a risk factor (RF) and keep a follow up of the disease (D) or final outcome that we will be measuring, that is, we understand that a cohort design often has a prospective nature as it moves forward from the RF to the D (Fig. 1). Therefore, none of the study participants should have the final outcome at the beginning of the study or time 0.

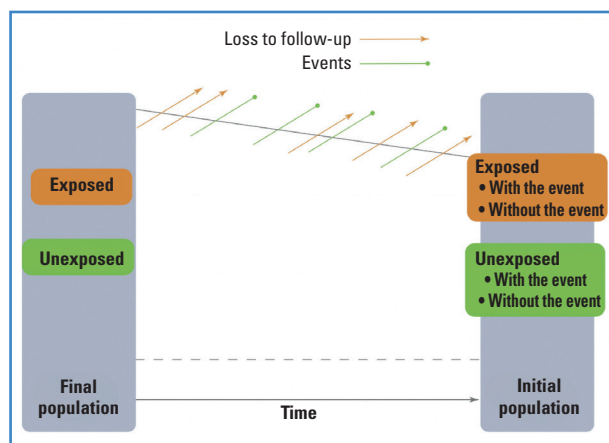


Figure 1. Cohort study design.

So, what should we do with those subjects we have selected from exposed and unexposed groups? Well, we must follow them over time and see which participants experience the event and which do not. Therefore, if we are going to define the new observed events (since at time 0, no one experiences it) over time, the frequency measures we can obtain with this type of studies are the incidence rates (both in the exposed and unexposed groups), and consequently, the measure of association will be the relative risk (RR).

We can quantify this association by building a ratio between the incidence rate of the phenomenon in the exposed group to the variable (I_e) and the incidence rate of the phenomenon in the unexposed group (I_o) (Table I):

$$RR = \frac{\text{Incidence rate in the exposed group}}{\text{Incidence rate in the unexposed group}}$$

$$\frac{I_e}{I_o} = \frac{a / a + b}{c / c + d}$$

Table I. Estimate of the RR

	Sick	Healthy	Total
RF	a	b	a + b
No RF	c	d	c + d
Total	a + c	b + d	a + b + c + d

RF, risk factor.

Additionally, if we know the exact time when events occur, we can also estimate individual timeframes at risk, as well as the incidence densities.

As figure 2 illustrates, events will occur over time (and we will also lose study participants for whatever reason), and at the end of the study, we will be able to see in each group (exposed and unexposed) who experienced the event under study and who did not.

Depending on how we recruit subjects for our study, cohorts can be of 2 types:

1. *Fixed or closed cohort.* Recruitment of all patients at the same time.
2. *Dynamic cohort.* Recruitment of participants at different timeframes. This one is the most common of the two.

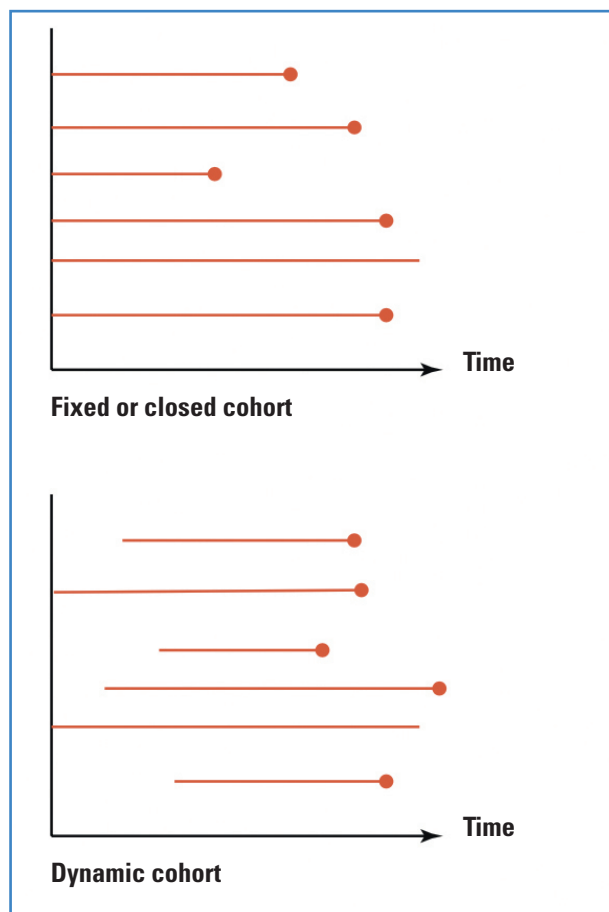


Figure 2. Types of cohorts based on recruitment.

In both cases, the analyses and what we'll be able to measure in each of the 2 cohorts are identical.

As it happens with all studies, cohorts have pros and cons.

Pros:

- The primary endpoint of this type of study is to analyze a common outcome found in a population of patients.
- Rare exposures among the population. For example, if we want to study the association between the use of a rare drug and the development of eczema (very common).
- Several events and reactions of the same exposure can be evaluated (if we have participants exposed and unexposed to coal in a mine, we can assess lung cancer, silicosis, COPD, etc.). In other words, multi-effectiveness.
- There is a clear temporal sequence between exposure and final outcomes.
- Exposure is known before the outcome.

- The frequency measure is the incidence rate, which allows us to estimate association measures such as the RR.

Cons:

- They are not useful to study rare diseases (thousands of people would be needed to assess a disease with very low prevalence).
- They are expensive and time-consuming (patients need to be followed for months or even years).
- They are unsuitable for diseases that will develop long after exposure.
- Losses can occur at the follow-up. This can be a problem when losses to follow-up are induced by exposure.
- They can have diagnostic biases. Disease criteria can change substantially over time, or diagnostic techniques can be more accurate.

RETROSPECTIVE COHORTS

The design and concept of a historical or retrospective cohort are both similar to what we studied in the previous case. Similarly, groups compared based on whether subjects are exposed at the beginning of the study should be identified. The difference is in the fact that by the time the assessment is being made, someone or everybody may have already experienced the event of interest.

As defined in this design, it is essential to have proper registries and records available (in health sciences, the common thing to do is to use the health records at our disposal).

The aim of the study is to go back in time and identify the cohorts (exposed and unexposed), and same as it happens in a prospective cohort, see what new cases appear in each cohort (Fig. 3).

Pros compared to a prospective cohort study:

- More economical and efficient (no follow-up required).
- Shorter duration (many cases of the event have already occurred).
- Very useful in diseases with a long period of time elapsed between the RF and the D.

Cons compared to a prospective cohort study:

- There is not a clear temporal sequence as in the case of a prospective cohort.

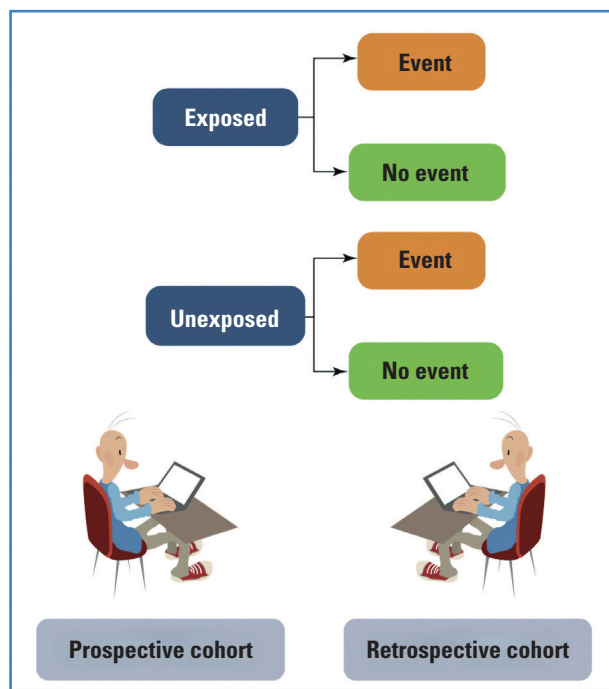


Figure 3. Investigator's relationship with exposure.

- The investigator does not have control over the nature and quality of the measurements taken.
- Important information on the topic of discussion may be missing.

It is plain to see that it's a retrospective cohort. Our participants are divided into exposed or unexposed from the very moment they are diagnosed. We would need the patients' health history to create these groups. At the time of the study, some patients may have already experienced the study event (death), while others don't.

As figure 3 illustrates, the study design is the same; the only thing that changes is the investigator's perspective. Therefore, the RR can be applied the same way in both retrospective and prospective cohorts.

DESIGN OF CASE-CONTROL TRIALS

In this case, unlike cohort studies where we started with exposed and unexposed groups, we start by identifying a group of cases (individuals with a particular disease or condition) and a control group (individuals without the disease or condition). The objective is to assess the presence or absence of risk factors (exposure) in the past to see whether the prevalence of exposure is different between cases and controls.

While it is true that historically case-control trials have always been very useful, they have always had epidemiological limitations regarding their interpretation because, unlike cohort studies, they do not follow the natural order of events: they start from the disease and then they look for exposure (Fig. 4).

The primary endpoint of a case-control trial is to study diseases with low prevalence or very long latency periods (disadvantages of cohort studies).

These are the pros and cons of case-control trials:
Pros:

- They are inexpensive and faster to conduct compared to a cohort study.
- They are easier from a logistical point of view, and do not require follow-up since, at time 0, we already have individuals with the condition or disease being studied.
- There are not follow-up losses.
- They are the best option for studying rare diseases with a long latency period.
- They allow us to study a wide variety of exposures in a "simple" way.

Cons:

- Sometimes it is challenging to determine whether exposure has caused the disease or whether the disease has changed exposure.

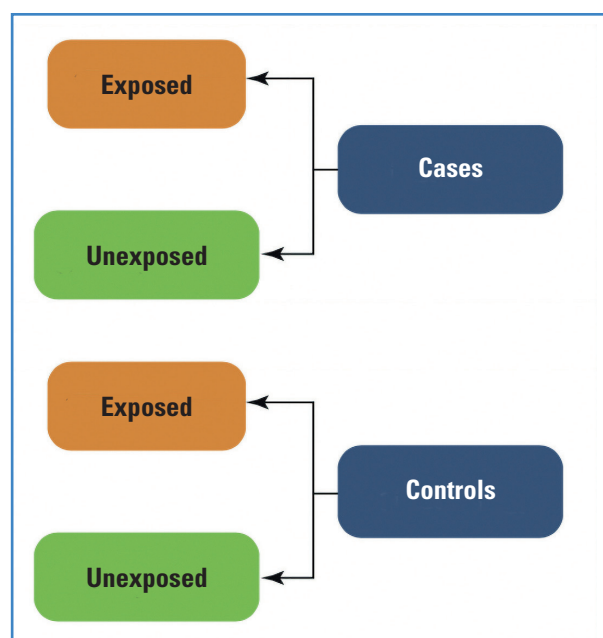


Figure 4. Time sequence of case-control trial designs.

- There is a higher risk of categorization bias and selection bias. One of the biggest challenges of case-control trials is selecting the proper control group.
- They are unsuitable for the study of rare exposures.
- Only 1 disease can be assessed at a time.
- Incidence rates, and therefore RR, cannot be estimated as measures of frequency.
- The measurement of association that can be obtained is the odds ratio (OR).

$$OR = \frac{\text{Odds ratio of case exposure}}{\text{Odds ratio of control exposure}} = \frac{\frac{\text{Cases exposed}}{\text{Cases unexposed}}}{\frac{\text{Controls exposed}}{\text{Controls unexposed}}}$$

A case-control trial is appropriate when the groups compared are similar in a series of variables that could bias the comparisons.

One of the challenges associated with this type of trial is that the information on exposure to the risk factor needs to be obtained by the investigator. This information is often collected through personal interviews or surveys, which is prone to significant bias—typical of case-control trials—known as recall bias, which would affect the controls in our study since they may not accurately remember exposure to a factor that happened a while ago. One way to control this type of bias is using the same method to obtain information for both groups, or if possible, use records that we know are error-free.

We will now discuss 2 variants of case-control trial designs used in research.

Prospective cases and controls

A prospective case-control trial is a type of observational research study conducted with data “that will happen” in the future, unlike retrospective case-control trials (the common type) that are based on historical series.

In a prospective case-control trial, we start by selecting a group of individuals who have not yet developed the disease or condition of interest but are at risk of doing so. These individuals are followed

over a specified period of time to see whether they develop the disease.

Demographic data, risk factors, exposures, and other relevant data are gathered for each study participant at the follow-up. At the end of the follow-up period, the participants who developed the disease or condition are categorized as *cases*, while those who have not as *controls*.

For example, let's say we want to evaluate various risk factors for type II endoleak after EVAR at the 1-year follow-up. Patients who have undergone EVAR are followed during this time. Those who develop the endoleak will be categorized as cases, while those who complete this time without experiencing it will be categorized as controls.

After identifying cases and controls (endoleaks and no endoleaks), retrospective information on exposures or risk factors that each individual may have experienced before developing the disease or condition is collected. This information is then compared between cases and controls to determine whether there is an association between the risk factors and the disease or condition being studied.

Unlike retrospective case-control trials, where cases have already developed the outcome, and exposure to the risk factor is sought retrospectively, in prospective case-control trials, investigators follow the participants over time and collect information on exposure and the development of the outcomes of interest as they occur.

The advantage of this design is that it allows a more accurate and detailed collection of information on exposure to the risk factor prior to developing the outcome, thus reducing recall bias and retrospective information distortion. However, it can require prolonged participant follow-up, be more expensive and labor-intensive compared to other research designs.

In conclusion, a prospective case-control trial is a research design that compares cases that have developed an outcome of interest with prospective controls who are still at risk of developing it, while collecting information on exposure to the risk factor as the events occur.

Nested case-control trial

A nested case-control trial, or nested case-control within a cohort is essentially a case-control trial

conducted within a cohort (a group of subjects) with specific characteristics.

For example, if we want to investigate whether AAAs are associated with AHT, but we want to eliminate all potential confounding factors such as gender or smoking, we may select a cohort of male smokers alone.

We would follow this cohort over time, some individuals would end up developing AAA (cases), while others don't (controls). Within each of these 2 groups, we would be able to see the prevalence of AHT. Therefore, the resulting association cannot be explained by gender or smoking, but only by the differential factor between the 2 groups of patients (which would be AHT).

Pros:

- Similar to a standard case-control trial.
- One of the most significant pros is that it allows us to examine unknown or previously unconsidered risk factors when the cohort was initiated.

REFERENCES

1. Martín Conejero A. Estudios epidemiológicos o cómo tenemos que diseñar nuestra investigación (primera parte). *Angiología* 2023;75(5):321-5. DOI: 10.20960/angiologia.00543
2. Martín-Conejero A. Metodología básica de la investigación clínica. Madrid: CTO Editorial; 2019.