

# Epidemiologic Study Designs: Cohort Studies

## Desenhos de Estudos Epidemiológicos: Estudos de Coorte

Sara Simões Dias<sup>1,2</sup>\*, Maria João Batalha<sup>1,2</sup>

### \*Corresponding Author/Autor Correspondente

Sara Simões Dias [sara.dias@ipleiria.pt]

Rua de Santo André – 66-68, Campus 5, Politécnico de Leiria, 2410-541 Leiria

ORCID: <https://orcid.org/0000-0001-6782-7481>

<https://doi.org/10.48687/ljsj.216>

### Abstract

This paper is the second one of a series on “study designs” in which a review of cohort study design is presented. A cohort study is a nonexperimental or observational study. In a cohort study, a group of subjects (the cohort) is followed for a period of time and the participants do not have the outcome of interest in the beginning. These studies can be prospective or retrospective and the time between exposure and outcome must be well defined.

**Keywords:** Cohort Studies; Epidemiologic Studies; Longitudinal Studies; Research Design

### Resumo

Este artigo é o segundo de uma série sobre “desenhos de estudo” em que é apresentada uma revisão da literatura sobre estudos de coorte. Um estudo de coorte é um estudo não experimental ou observacional. Os participantes (a coorte) são seguidos ao longo de um período de tempo definido e no início do estudo os participantes não apresentam o variável resultado (**outcome**) de interesse. Estes tipos de estudos podem ser prospetivos ou retrospectivos e o tempo entre a exposição e a variável de interesse tem de ser bem definido.

**Palavras-chave:** Estudos Coorte; Estudos Epidemiológicos; Estudos Longitudinais; Projetos de Investigação

### Introduction

The word cohort is derived from the Latin “*cohors*”, in Roman times, meant a “group of soldiers”. The term “cohort” refers to a group of people with something in common, usually an exposure or an event that is based on the definition decided by the researcher.

This article aims to discuss various issues related to the design of cohort studies, with special focus on the nomenclature of

perspective and retrospective study designs, as well as exploring the outcomes and measures, strengths and limitations of cohort studies.

### Study Design

A cohort study is an observational and longitudinal study. In an observational study, researchers observe and record information about the subjects without manipulating the study

**1.** ciTechCare – Center for Innovative Care and Health Technology, Polytechnic of Leiria, Leiria, Portugal. **2.** School of Health Sciences, Polytechnic of Leiria, Leiria, Portugal.

**Recebido/Received:** 17/06/2024 – **Aceite/Accepted:** 18/06/2024 – **Publicado online/Published online:** 28/06/2024 – **Publicado / Published:** 28/06/2024

© Author(s) (or their employer(s)) and Lusíadas Scientific Journal 2024. Re-use permitted under CC BY-NC 4.0. No commercial re-use.

© Autor (es) (ou seu (s) empregador (es)) e Lusíadas Scientific Journal 2024. Reutilização permitida de acordo com CC BY-NC 4.0. Nenhuma reutilização comercial.

environment or the subjects themselves, while a longitudinal study involves collecting data from the same subjects repeatedly over a period. A key characteristic of a cohort study is that at the starting point, none of the subjects have the outcome of interest. Their exposure to a risk factor is assessed, and they are classified into two groups: exposed and unexposed. The cohort will share a set of characteristics and will be observed over a period of time to analyze their evolution. During the follow-up period, some individuals in the exposed group will develop the outcome, and some in the unexposed group will also develop the outcome of interest. One of the objectives of cohort studies is to compare the outcomes between these two groups.

If a positive association exists between the exposure and the outcome, we expect that the proportion of the exposed group in whom the disease develops (incidence in the exposure group) would be greater than the proportion of the nonexposed group in whom the disease develops (incidence in the non-exposed group).<sup>1</sup> The calculations are in Table 1.

Since exposure is identified before the outcome, cohort studies have a temporal framework that allows for the assessment of causality, offering the potential to provide the strongest scientific evidence.<sup>2,3</sup>

Table 1. Calculation of incidence

**ASSOCIATION MEASURE**

	DEVELOPED THE DISEASE D+	DID NOT DEVELOP THE DISEASE D-	TOTAL	INCIDENCE OF THE DISEASE
EXPOSED E+	a	b	a+b	a/(a+b)
NOT EXPOSED E-	c	d	c+d	c/(c+d)

INCIDENCE ON THOSE EXPOSED =  $a/(a+b)$   
 INCIDENCE IN THOSE NOT EXPOSED =  $c/(c+d)$

**Prospective versus Retrospective Study Designs**

Cohort studies are always longitudinal, although can be prospective or retrospective. In a prospective cohort study, all the data are collected prospectively. Prospective studies are carried out from the present time into the future. The researcher defines the population that will be included in the cohort, classifies the exposed and non-exposed group and follow these participants. At baseline and throughout the follow-up, the researcher also gathers data on other variables crucial to the study, including confounding variables. During the time of follow-up, the outcome of interest is assessed. Some of these outcomes may only occur once (for example: death), and some may occur multiple times (for example diarrhea, wheezing episodes, etc).<sup>4</sup>

An advantage of prospective cohort studies is that all relevant variables can be anticipated in advance, allowing trained study staff to accurately measure and record data related to these variables.<sup>5</sup>

In retrospective cohort study, the researchers start the study at the time follow-up has already been completed. Retrospectively, eligible subjects are identified, a cohort is composed, and exposures are assessed at baseline. Thereafter, the outcome is measured during the follow-up period (Fig. 1).

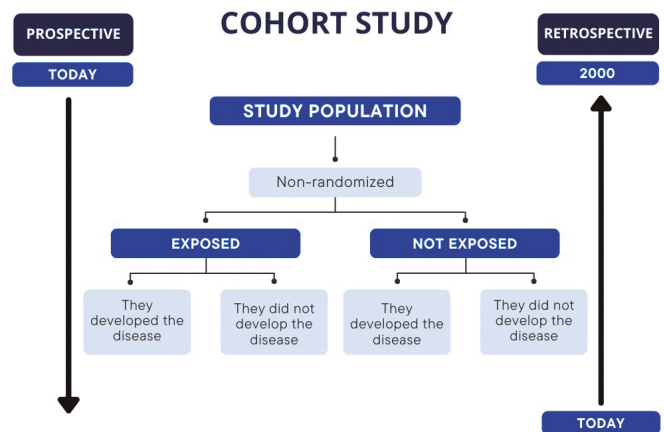


Figure 1. Design of a cohort study

Typically, a prospective design has been ranked higher in the hierarchy of evidence than a retrospective design.<sup>6</sup>

**Strengths and Limitations**

A cohort study has several strengths, with its most significant being the clearly defined temporality between exposure and outcome. Another advantage is that in a cohort study is possible to analyze multiple outcomes in the same exposure. It is only necessary to ensure that the subjects do not have the outcome at baseline. If the exposure is rare, a cohort study is also an efficient method to study the relationship between exposure and outcome.<sup>4</sup> A retrospective cohort study also has the advantage of being completed quickly and being relatively inexpensive compared to a prospective cohort study.

One major limitation of a prospective cohort design is that is time-consuming, because of a long follow-up period, and expensive. In a retrospective cohort study, the exposure and outcome variables are collected before the study begins, which may result in less accurate measurements.<sup>4</sup>

**Follow-up**

In a cohort study the follow-up of the study participants is very important and for this reason is important to minimize loss to follow-up, a source of bias in these types of studies. Loss to follow-up occurs when the researcher loses contact with a

subject, leading to missing data. If too many subjects are lost to follow-up, the internal validity of the study is compromised. A general rule of thumb requires that the loss to follow-up rate not exceed 20% of the sample.<sup>7</sup> Systematic differences in outcome or exposure risk factors between participants who drop out and those who stay in the study should be investigated, if possible, by comparing subjects who remain in the study with those who are lost to follow-up or who drop out.<sup>6</sup>

## Statistical Analysis

Cohort studies allow us to estimate the cumulative incidence and incidence rate. It also allows for other statistical methods, such as regression models: logistic regression, Kaplan-Meier curves, Cox-regression model, Poisson regression, and negative binomial regression, among others. Other statistical models very useful for handling longitudinal data are the fixed and random effects models. These are advanced modeling techniques and should be discussed with a statistician.<sup>4</sup>

## Cohort Studies in Portugal

In Portugal, several cohorts have been developed to monitor the health of the population, investigating the prevalence of chronic diseases, and identifying social, environmental, and genetic determinants that affect individual well-being. Examples include Generation XXI, EPIPorto, EpiDoc, and LiSa.

The Generation XXI cohort is a longitudinal study initiated in 2005, following 8647 children born in the metropolitan area of Porto, Portugal. Developed by the Institute of Public Health of the University of Porto (ISPUP) this study aims to characterize prenatal and postnatal development and identify its determinants to better understand health status during childhood, adolescence, and adulthood. Generation XXI collects comprehensive data in areas such as perinatal and pediatric health, obesity and metabolic health, lifestyles, cardiovascular health, and musculoskeletal health. Additionally, it provides insights into the mechanisms that lead to disease, playing an important role in planning health intervention strategies.<sup>8</sup>

The EPIPorto cohort, also developed by ISPUP, is a population-based epidemiological study initiated in 1999, which follows 2485 adults residing in the city of Porto, Portugal. This study aims to assess the health determinants of the adult population residing in Porto. EPIPorto addresses a set of questions regarding participants' social, demographic, and behavioral characteristics, as well as their medical history.<sup>9</sup>

The Epidemiology of Chronic Diseases (EpiDoC) study was initiated in 2011 and aimed to create a large population-based database for medical and health-related research in Portugal.

EpiDoC constituted one of the first large Portuguese prospective cohort studies, including a representative sample of the Portuguese population, with the main objective of examining the health determinants and outcomes of chronic non-communicable diseases and their impact on health resource consumption. The EpiDoC study was designed by researchers from NOVA Medical School in Lisbon.<sup>10</sup>

Finally, LiSa is a cohort study developed by Center for Innovative Care and Health Technology (ciTechCare), in partnership with the Municipality of Leiria and the Local Health Unit of the Leiria Region. This study aims to assess health literacy among the adult population residing in the municipality of Leiria and to characterize anxiety and depression, metabolic risk, as well as alcohol and tobacco consumption.<sup>11</sup>

## Conclusion

Cohort studies are indispensable tools in epidemiological research and public health studies, providing a longitudinal perspective on health trends and the factors influencing disease patterns across diverse populations. By following individuals over extended periods, cohort studies elucidate how various factors—such as lifestyle, environmental exposures, genetics, and socioeconomic status—contribute to health outcomes over time. These insights are invaluable for identifying risk factors, understanding disease trajectories, and developing evidence-based interventions. The results derived from cohort studies have the potential to profoundly impact public health policies and practices. They provide robust evidence that informs the formulation of more effective and targeted health policies, interventions, and preventive strategies. In conclusion, cohort studies empower decision-makers to allocate resources efficiently, mitigate health disparities, and enhance overall population health.

## Responsabilidades Éticas

**Conflitos de Interesse:** Os autores declaram não possuir conflitos de interesse.

**Suporte Financeiro:** O presente trabalho não foi suportado por nenhum subsídio o bolsa ou bolsa.

**Proveniência e Revisão por Pares:** Não comissionado; revisão externa por pares.

## Ethical Disclosures

**Conflicts of Interest:** The authors have no conflicts of interest to declare.

**Financial Support:** This work has not received any contribution grant or scholarship.

**Provenance and Peer Review:** Not commissioned; externally peer reviewed.

## Declaração de Contribuição

**SD:** Pesquisa bibliográfica, redação e organização do conteúdo.

**MJB:** Auxílio na redação e revisão do conteúdo.

Todos autores aprovaram a versão final a ser publicada

## Contributorship Statement

**SD:** Bibliographical search, writing and organizing content.

**MJB:** Help with writing and reviewing content.

All authors approved the final version to be published

## References

1. Gordis L. Epidemiology. Philadelphia: Saunders Elsevier; 2009.
2. Everitt BS, Palmer CR. Encyclopaedic Companion to Medical Statistics. London: Wiley; 2005.
3. Song JW, Chung KC. Observational Studies: Cohort and Case-Control Studies. *Plast Reconstr Surg.* 2010;126:2234–42. doi: 10.1097/PRS.0b013e-3181f44abc.
4. Setia M. Methodology series module 1: Cohort studies. *Indian J Dermatol.* 2016;61:21. doi: 10.4103/0019-5154.174011.
5. Andrade C. Research Design: Cohort Studies. *Indian J Psychol Med.* 2022;44:189–91. doi: 10.1177/02537176211073764.
6. Vandembroucke JP. Observational research, randomised trials, and two views of medical science. *PLoS Med.* 2008;5:e67. doi: 10.1371/journal.pmed.0050067.
7. Merrill RM, Timmreck TC. Introduction to Epidemiology. London: Jones & Bartlett Learning; 2006.
8. Instituto de Saúde Pública da Universidade do Porto. Geração XXI. Porto: ISPUP; 2020.
9. Instituto de Saúde Pública da Universidade do Porto. EPIPorto: ISPUP; 20xx.
10. Dias SS, Rodrigues AM, Gregório MJ, de Sousa RD, Branco JC, Canhão H. Cohort Profile: The Epidemiology of Chronic Diseases Cohort (EpiDoC). *Int J Epidemiol.* 2018;47:1741–2j. doi: 10.1093/ije/dyy185.
11. Batalha MJ, Gabriel T, Valentim A, Soledade A, Gomes C, Alves B, et al. Health literacy – study protocol for LiSa cohort study. *BMC Public Health.* 2024.